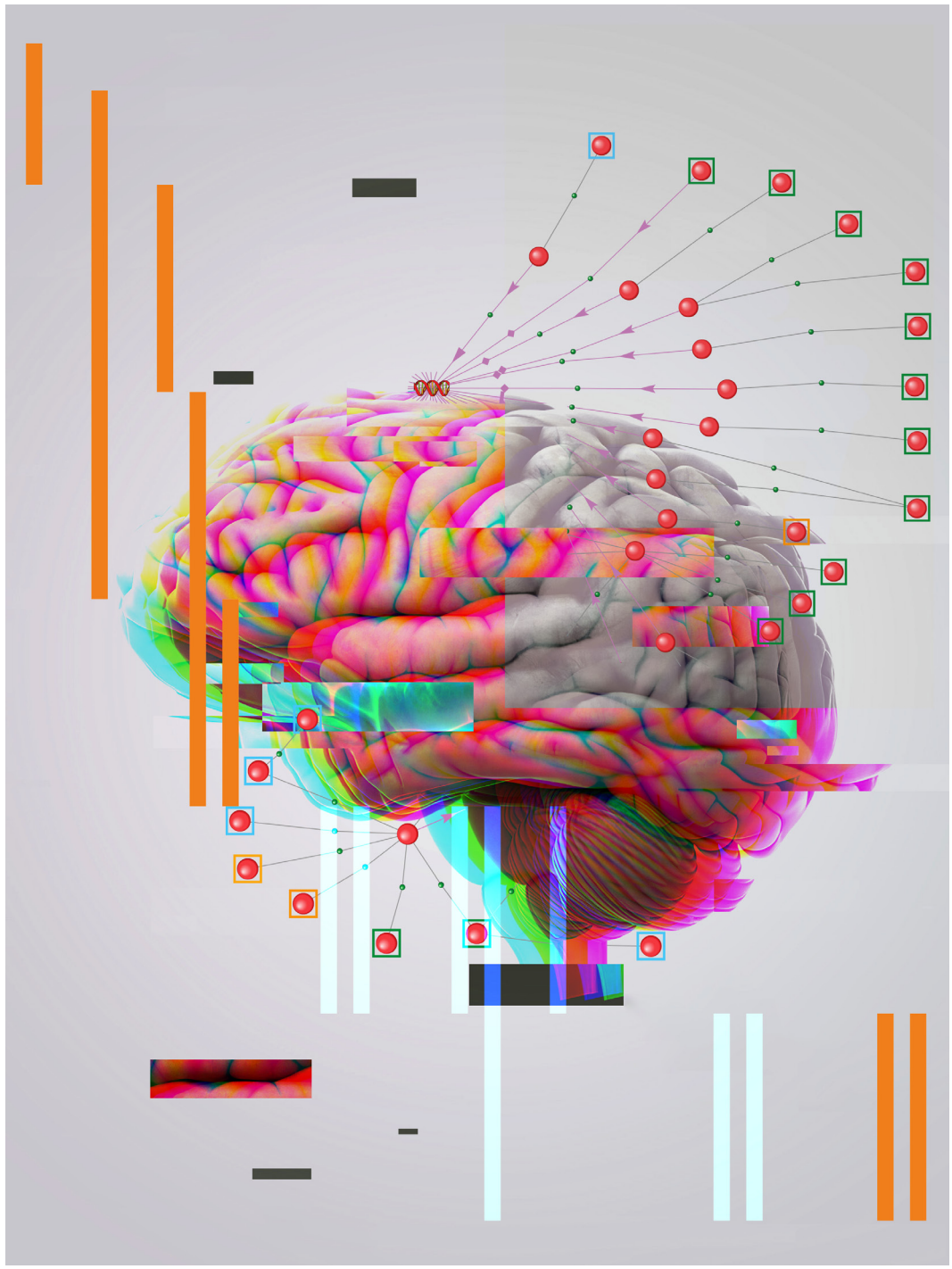
Online ISSN 2500-3259

ВАВИЛОВСКИЙ
ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ
VAVILOV JOURNAL OF GENETICS AND BREEDING

2024 • 28 • 8



Биоинформатика • Системная компьютерная биология • Геномика • Транскриптомика • Протеомика • Эволюционная биология • Нейроинформатика • Биомедицина

vavilov.elpub.ru
vavilovj-icg.ru
vavilov_journal@bionet.nsc.ru

CELL TECHNOLOGIES
REGENERATIVE MEDICINE
INTELLIGENT DATA SCIENCE
SYNTHETIC BIOLOGY
POSTGENOME
RESEARCH & DEVELOPMENT

YEREVAN ARMENIA 2025

КОНГРЕСС CRISPR-2025

5 – 10 октября
2025 года

Глубокоуважаемые коллеги!
Мы рады сообщить, что с 5 по 10 октября 2025 года состоится
Третий международный конгресс “CRISPR-2025”!

На этот раз организаторы конгресса встретят участников в гостеприимном солнечном Ереване (Армения).

Это событие станет площадкой для экспертного обсуждения последних достижений, вызовов и этических аспектов, связанных с использованием широчайшего спектра технологических подходов, направленных на редактирование геномов.

Организаторы:

- Институт цитологии и генетики СО РАН (Новосибирск, Россия),
- Институт химической биологии и фундаментальной медицины (Новосибирск, Россия),
- Российско-армянский университет (Ереван, Армения),
- Институт молекулярной биологии НАН РА (Ереван, Армения)
- Сургутский государственный университет (Сургут, Россия).

Партнеры: BIOCAD, Фонд научно-технологического развития Югры.

Предыдущие конгрессы, проведённые в 2018-м и 2023-м годах в Новосибирске, продемонстрировали востребованность и актуальность формата мероприятия.

Стремительно совершенствующиеся методы геномного редактирования, в том числе в области клинического применения CRISPR-систем, требуют всестороннего осмысления с учётом различных областей знаний.

Именно поэтому по сложившейся традиции мы предлагаем широкий спектр направлений, которые будут обсуждаться на конгрессе “CRISPR-2025”:

- **C**ell Technologies
- **R**egenerative Medicine
- **I**ntelligent Data Science
- **S**ynthetic Biology
- **P**ostgenome
- **R**esearch & Development.

Традиционно участники конгресса представят передовые разработки и фундаментальные исследования в области CRISPR, их применение в медицине, сельском хозяйстве, эволюционной и палеобиологии.

Предстоящий конгресс станет пространством для взаимодействия учёных, врачей и представителей индустрии, обмена идеями и эффективного сотрудничества.

В программе

- пленарные и секционные выступления,
- конкурс молодых учёных,
- конкурс постерных докладов,
- панельная дискуссия «От предсказания лекарств с помощью искусственного интеллекта, точного редактирования генома до клеточной и генной терапии».

Для всех желающих познакомиться с армянскими традициями оргкомитет готовит интересную культурную программу.

Приглашаем постоянных и новых участников: учёных, врачей, студентов и представителей бизнеса.

Информационный сайт конгресса: https://crispr2025.rau.am/

ИЦГ СурГУ BIOCAD Biotechnology Company

2024 • 28 • 8

Online ISSN 2500-3259

Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

VAVILOV JOURNAL OF GENETICS AND BREEDING

Основан в 1997 г.
Периодичность 8 выпусков в год
doi 10.18699/vjgb-24-88

Учредители

Сибирское отделение Российской академии наук

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

Главный редактор

А.В. Кочетов – академик РАН, д-р биол. наук (Россия)

Заместители главного редактора

Н.А. Колчанов – академик РАН, д-р биол. наук, профессор (Россия)
И.Н. Леонова – д-р биол. наук (Россия)
Н.Б. Рубцов – д-р биол. наук, профессор (Россия)
В.К. Шумный – академик РАН, д-р биол. наук, профессор (Россия)

Ответственный секретарь

Г.В. Орлова – канд. биол. наук (Россия)

Редакционная коллегия

Е.Е. Андронов – канд. биол. наук (Россия)
Ю.С. Аульченко – д-р биол. наук (Россия)
О.С. Афанасенко – академик РАН, д-р биол. наук (Россия)
Д.А. Афонников – д-р биол. наук, доцент (Россия)
Л.И. Афтанас – академик РАН, д-р мед. наук (Россия)
Л.А. Беспалова – академик РАН, д-р с.-х. наук (Россия)
А. Бёрнер – д-р наук (Германия)
Н.П. Бондарь – канд. биол. наук (Россия)
С.А. Боринская – д-р биол. наук (Россия)
П.М. Бородин – д-р биол. наук, проф. (Россия)
А.В. Васильев – чл.-кор. РАН, д-р биол. наук (Россия)
М.И. Воевода – академик РАН, д-р мед. наук (Россия)
Т.А. Гавриленко – д-р биол. наук (Россия)
И. Гроссе – д-р наук, проф. (Германия)
Н.Е. Грунтенко – д-р биол. наук (Россия)
С.А. Демаков – д-р биол. наук (Россия)
И.К. Захаров – д-р биол. наук, проф. (Россия)
И.А. Захаров-Гезехус – чл.-кор. РАН, д-р биол. наук (Россия)
С.Г. Инге-Вечтомов – академик РАН, д-р биол. наук (Россия)
А.В. Кильчевский – чл.-кор. НАНБ, д-р биол. наук (Беларусь)
С.В. Костров – чл.-кор. РАН, д-р хим. наук (Россия)
А.М. Кудрявцев – чл.-кор. РАН, д-р биол. наук (Россия)
И.Н. Лаврик – д-р биол. наук (Германия)
Д.М. Ларкин – канд. биол. наук (Великобритания)
Ж. Ле Гуи – д-р наук (Франция)
И.Н. Лебедев – д-р биол. наук, проф. (Россия)
Л.А. Лутова – д-р биол. наук, проф. (Россия)
Б. Люгтенберг – д-р наук, проф. (Нидерланды)
В.Ю. Макеев – чл.-кор. РАН, д-р физ.-мат. наук (Россия)
В.И. Молодин – академик РАН, д-р ист. наук (Россия)
М.П. Мошкин – д-р биол. наук, проф. (Россия)
С.Р. Мурсалимов – канд. биол. наук (Россия)
Л.Ю. Новикова – д-р с.-х. наук (Россия)
Е.К. Потокина – д-р биол. наук (Россия)
В.П. Пузырев – академик РАН, д-р мед. наук (Россия)
Д.В. Пышный – чл.-кор. РАН, д-р хим. наук (Россия)
И.Б. Рогозин – канд. биол. наук (США)
А.О. Рувинский – д-р биол. наук, проф. (Австралия)
Е.Ю. Рыкова – д-р биол. наук (Россия)
Е.А. Салина – д-р биол. наук, проф. (Россия)
В.А. Степанов – академик РАН, д-р биол. наук (Россия)
И.А. Тихонович – академик РАН, д-р биол. наук (Россия)
Е.К. Хлесткина – д-р биол. наук, проф. РАН (Россия)
Э.К. Хуснутдинова – д-р биол. наук, проф. (Россия)
М. Чен – д-р биол. наук (Китайская Народная Республика)
Ю.Н. Шаврукова – д-р биол. наук (Австралия)
Р.И. Шейко – чл.-кор. НАНБ, д-р с.-х. наук (Беларусь)
С.В. Шестаков – академик РАН, д-р биол. наук (Россия)
Н.К. Янковский – академик РАН, д-р биол. наук (Россия)

2024 • 28 • 8

Online ISSN 2500-3259

Online edition

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

VAVILOV JOURNAL OF GENETICS AND BREEDING

Founded in 1997
Published 8 times annually
doi 10.18699/vjgb-24-88

Founders

Siberian Branch of the Russian Academy of Sciences

Federal Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences

The Vavilov Society of Geneticists and Breeders

Editor-in-Chief

A.V. Kochetov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

Deputy Editor-in-Chief

N.A. Kolchanov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia
I.N. Leonova, Dr. Sci. (Biology), Russia
N.B. Rubtsov, Professor, Dr. Sci. (Biology), Russia
V.K. Shumny, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

Executive Secretary

G.V. Orlova, Cand. Sci. (Biology), Russia

Editorial board

O.S. Afanasenko, Full Member of the RAS, Dr. Sci. (Biology), Russia
D.A. Afonnikov, Associate Professor, Dr. Sci. (Biology), Russia
L.I. Aftanas, Full Member of the RAS, Dr. Sci. (Medicine), Russia
E.E. Andronov, Cand. Sci. (Biology), Russia
Yu.S. Aulchenko, Dr. Sci. (Biology), Russia
L.A. Bespalova, Full Member of the RAS, Dr. Sci. (Agricul.), Russia
N.P. Bondar, Cand. Sci. (Biology), Russia
S.A. Borinskaya, Dr. Sci. (Biology), Russia
P.M. Borodin, Professor, Dr. Sci. (Biology), Russia
A. Börner, Dr. Sci., Germany
M. Chen, Dr. Sci. (Biology), People's Republic of China
S.A. Demakov, Dr. Sci. (Biology), Russia
T.A. Gavrilenko, Dr. Sci. (Biology), Russia
I. Grosse, Professor, Dr. Sci., Germany
N.E. Gruntenko, Dr. Sci. (Biology), Russia
S.G. Inge-Vechtomov, Full Member of the RAS, Dr. Sci. (Biology), Russia
E.K. Khlestkina, Professor of the RAS, Dr. Sci. (Biology), Russia
E.K. Khusnutdinova, Professor, Dr. Sci. (Biology), Russia
A.V. Kilchevsky, Corr. Member of the NAS of Belarus, Dr. Sci. (Biology), Belarus
S.V. Kostrov, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia
A.M. Kudryavtsev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia
D.M. Larkin, Cand. Sci. (Biology), Great Britain
I.N. Lavrik, Dr. Sci. (Biology), Germany
J. Le Gouis, Dr. Sci., France
I.N. Lebedev, Professor, Dr. Sci. (Biology), Russia
B. Lugtenberg, Professor, Dr. Sci., Netherlands
L.A. Lutova, Professor, Dr. Sci. (Biology), Russia
V.Yu. Makeev, Corr. Member of the RAS, Dr. Sci. (Physics and Mathem.), Russia
V.I. Molodin, Full Member of the RAS, Dr. Sci. (History), Russia
M.P. Moshkin, Professor, Dr. Sci. (Biology), Russia
S.R. Mursalimov, Cand. Sci. (Biology), Russia
L.Yu. Novikova, Dr. Sci. (Agricul.), Russia
E.K. Potokina, Dr. Sci. (Biology), Russia
V.P. Puzyrev, Full Member of the RAS, Dr. Sci. (Medicine), Russia
D.V. Pyshnyi, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia
I.B. Rogozin, Cand. Sci. (Biology), United States
A.O. Ruvinsky, Professor, Dr. Sci. (Biology), Australia
E.Y. Rykova, Dr. Sci. (Biology), Russia
E.A. Salina, Professor, Dr. Sci. (Biology), Russia
Y.N. Shavrukov, Dr. Sci. (Biology), Australia
R.I. Sheiko, Corr. Member of the NAS of Belarus, Dr. Sci. (Agricul.), Belarus
S.V. Shestakov, Full Member of the RAS, Dr. Sci. (Biology), Russia
V.A. Stepanov, Full Member of the RAS, Dr. Sci. (Biology), Russia
I.A. Tikhonovich, Full Member of the RAS, Dr. Sci. (Biology), Russia
A.V. Vasiliev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia
M.I. Voevoda, Full Member of the RAS, Dr. Sci. (Medicine), Russia
N.K. Yankovsky, Full Member of the RAS, Dr. Sci. (Biology), Russia
I.K. Zakharov, Professor, Dr. Sci. (Biology), Russia
I.A. Zakharov-Gezekhus, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

СОДЕРЖАНИЕ • 2024 • 28 • 8

807 ОТ РЕДАКТОРА
Н.А. Колчанов

Геномика и транскриптомика

808 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Кандидатные SNP-маркеры изменения экспрессии гена *SCN9A* человека в качестве интегратора генерации, чувства, ответа на боль и анестезии.
П.А. Доценко, К.А. Золотарева, Р.А. Иванов, И.В. Чадаева, Н.Л. Подколодный, В.А. Иванисенко, П.С. Деменков, С.А. Лашин, М.П. Пономаренко

822 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Программный комплекс MetArea для анализа взаимоисключающей встречаемости в парах мотивов сайтов связывания транскрипционных факторов по данным ChIP-seq.
В.Г. Левицкий, А.В. Цуканов, Т.И. Меркулова

834 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Компьютерный анализ показывает отличия митохондриальных микроРНК от остальных микроРНК.
П.С. Ворожейкин, И.И. Титов

Эволюционная биология

843 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Новый подход к анализу эволюции SARS-CoV-2, основанный на визуализации и кластеризации больших объемов генетических данных, компактно представленных в оперативной памяти.
А.Ю. Пальянов, Н.В. Пальянова

854 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Поиск и функциональная аннотация многодоменных белков семейства ФА2 у плоских червей. *М.Е. Бочарникова, И.И. Турнаев, Д.А. Афонников*

864 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Реконструкция и компьютерный анализ структурно-функциональной организации генной сети регуляции биосинтеза холестерина у человека и эволюционная характеристика участвующих в ней генов.
А.Д. Михайлова, С.А. Лашин, В.А. Иванисенко, П.С. Деменков, Е.В. Игнатьева

874 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Orthoweb: программный комплекс для эволюционного анализа генных сетей.
Р.А. Иванов, А.М. Мухин, Ф.В. Казанцев, З.С. Мустафин, Д.А. Афонников, Ю.Г. Матушкин, С.А. Лашин

Системная компьютерная биология

882 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Изучение особенностей метаболизма тканей глиобластомы и перитуморального пространства при использовании таргетированного метаболомного скрининга методом ВЭЖХ-МС/МС и генных сетей. *Н.В. Басов, А.В. Адамовская, А.Д. Рогачев, Е.В. Гайслер, П.С. Деменков, Т.В. Иванисенко, А.С. Вензель, С.В. Мишинов, В.В. Ступак, С.В. Чересиз, О.С. Олешко, Е.А. Бутикова, А.Е. Осечкова, Ю.С. Сотникова, Ю.В. Патрушев, А.С. Поздняков, И.Н. Лаврик, В.А. Иванисенко, А.Г. Покровский*

897 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Программный модуль для оценки метаболического потенциала мутантных штаммов бактерии *Corynebacterium glutamicum*. *Ф.В. Казанцев, М.Ф. Трофимова, Т.М. Хлебодарова, Ю.Г. Матушкин, С.А. Лашин*

904 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Реконструкция и компьютерный анализ генной сети, отражающей роль микроРНК в регуляции ответа пшеницы на засуху.
М.А. Клещев, А.В. Мальцева, Е.А. Антропова, П.С. Деменков, Т.В. Иванисенко, Ю.Л. Орлов, Х. Чао, М. Чэнь, Н.А. Колчанов, В.А. Иванисенко

918 ОБЗОР
Пограничные клетки корневого чехлика как регулятор ризосферной микробиоты.
Н.А. Омельянчук, В.А. Черенко, Е.В. Землянская

927 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Метод генных сетей и метаболомный анализ позволили выявить специфические пути изменения профиля аминокислот и ацилкарнитинов в плазме крови при болезни Паркинсона и сосудистом паркинсонизме. *А.А. Макарова, П.М. Мельникова, А.Д. Рогачев, П.С. Деменков, Т.В. Иванисенко, Е.В. Предтеченская, С.Ю. Карманов, В.В. Коваль, А.Г. Покровский, И.Н. Лаврик, Н.А. Колчанов, В.А. Иванисенко*

940 ОБЗОР
Онтологии в моделировании и анализе больших генетических данных.
Н.Л. Подколодный, О.А. Подколодная, В.А. Иванисенко, М.А. Марченко

950 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
PlantReg: реконструкция связей между регуляторными сетями транскрипционных факторов и контролируемыми ими признаками.
В.В. Лавреха, Н.А. Омельянчук, А.Г. Богомолов, Е.В. Землянская

960 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Поиск перспективных генетических маркеров, ассоциированных с молекулярными механизмами снижения устойчивости риса к *Rhizoctonia solani* при избытке азотных удобрений, методом реконструкции и анализа генных сетей.
Е.А. Антропова, А.Р. Волянская, А.В. Адамовская, П.С. Деменков, И.В. Яцык, Т.В. Иванисенко, Ю.Л. Орлов, Х. Чао, М. Чэнь, В.А. Иванисенко

974 ОБЗОР
Методы реконструкции генных регуляторных сетей на основе транскриптомных данных отдельных клеток. *М.А. Рыбаков, Н.А. Омельянчук, Е.В. Землянская*

982 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Сопоставление показателей мозговой активности у китайских и российских студентов в условиях распознавания информации, отнесенной к себе и другим людям.
Ц. Сы, Ц. Тянь, В.А. Савостьянов, Д.А. Лебедкин, А.В. Бочаров, А.Н. Савостьянов

Биомедицина

993 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Концепция природной реконструкции генома. Часть 2. Влияние фрагментов экстраклеточной двуцепочечной ДНК на гемопоэтические стволовые клетки.
В.С. Рузанова, С.Г. Ошихмина, А.С. Проскурина, Г.С. Риттер, С.С. Кирикович, Е.В. Левитес, Я.Р. Ефремов, Т.В. Карамышева, М.И. Мещанинова, А.Л. Мамаев, О.С. Таранов, А.С. Богачев, С.В. Сидоров, С.Д. Никонов, О.Ю. Леплина, А.А. Останин, Е.Р. Черных, Н.А. Колчанов, Е.В. Долгова, С.С. Богачев

1008 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
База данных о генах и белках, ассоциированных с нарушениями метаболизма глюкозы (GlucoGenes®): описание и возможности применения в биоинформатических исследованиях.
В.В. Климонтов, К.С. Шишин, Р.А. Иванов, М.П. Пономаренко, К.А. Золотарева, С.А. Лашин

1018 ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ
Ассоциация аутистических личностных черт у неклинических испытуемых с показателями ЭЭГ в условиях просмотра видеозаписей лица.
А.Н. Савостьянов, Д.А. Кулешов, Д.И. Клемешова, М.С. Власов, А.Е. Сапрыгин

1025 Алфавитный указатель авторов статей, опубликованных в журнале в 2024 г.

© Сибирское отделение Российской академии наук, 2024
© Институт цитологии и генетики СО РАН, 2024
Вавиловский журнал генетики и селекции, 2024

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII • VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2024 • 28 • 8

807 FROM THE EDITOR
N.A. Kolchanov

Genomics and transcriptomics

808 ORIGINAL ARTICLE
Candidate SNP markers of changes in the expression levels of the human *SCN9A* gene as a hub gene for pain generation, perception, response and anesthesia. *P.A. Dotsenko, K.A. Zolotareva, R.A. Ivanov, I.V. Chadaeva, N.L. Podkolodnyy, V.A. Ivanisenko, P.S. Demenkov, S.A. Lashin, M.P. Ponomarenko*

822 ORIGINAL ARTICLE
MetArea: a software package for analysis of the mutually exclusive occurrence in pairs of motifs of transcription factor binding sites based on ChIP-seq data.
V.G. Levitsky, A.V. Tsukanov, T.I. Merkulova

834 ORIGINAL ARTICLE
Computer analysis shows differences between mitochondrial miRNAs and other miRNAs. *P.S. Vorozheykin, I.I. Titov*

Evolutionary biology

843 ORIGINAL ARTICLE
A novel approach to analyzing the evolution of SARS-CoV-2 based on visualization and clustering of large genetic data compactly represented in operative memory. *A.Yu. Palyanov, N.V. Palyanova*

854 ORIGINAL ARTICLE
Search for and functional annotation of multi-domain PLA2 family proteins in flatworms. *M.E. Bocharnikova, I.I. Turnaev, D.A. Afonnikov*

864 ORIGINAL ARTICLE
Reconstruction and computer analysis of the structural and functional organization of the gene network regulating cholesterol biosynthesis in humans and the evolutionary characteristics of the genes involved in the network.
A.D. Mikhailova, S.A. Lashin, V.A. Ivanisenko, P.S. Demenkov, E.V. Ignatieva

874 ORIGINAL ARTICLE
Orthoweb: software package for evolutionary analysis of gene networks.
R.A. Ivanov, A.M. Mukhin, F.V. Kazantsev, Z.S. Mustafin, D.A. Afonnikov, Y.G. Matushkin, S.A. Lashin

Systems computational biology

882 ORIGINAL ARTICLE
Investigation of metabolic features of glioblastoma tissue and the peritumoral environment using targeted metabolomics screening by LC-MS/MS and gene network analysis. *N.V. Basov, A.V. Adamovskaya, A.D. Rogachev, E.V. Gaisler, P.S. Demenkov, T.V. Ivanisenko, A.S. Venzel, S.V. Mishinov, V.V. Stupak, S.V. Cheresiz, O.S. Oleshko, E.A. Butikova, A.E. Osechkova, Yu.S. Sotnikova, Y.V. Patrushev, A.S. Pozdnyakov, I.N. Lavrik, V.A. Ivanisenko, A.G. Pokrovsky*

897 ORIGINAL ARTICLE
A software module to assess the metabolic potential of mutant strains of the bacterium *Corynebacterium glutamicum*. *F.V. Kazantsev, M.F. Trofimova, T.M. Khlebodarova, Yu.G. Matushkin, S.A. Lashin*

904 ORIGINAL ARTICLE
Reconstruction and computational analysis of the microRNA regulation gene network in wheat drought response mechanisms.
M.A. Kleshchev, A.V. Maltseva, E.A. Antropova, P.S. Demenkov, T.V. Ivanisenko, Y.L. Orlov, H. Chao, M. Chen, N.A. Kolchanov, V.A. Ivanisenko

918 REVIEW
Root cap border cells as regulators of rhizosphere microbiota.
N.A. Omelyanchuk, V.A. Cherenko, E.V. Zemlyanskaya

927 ORIGINAL ARTICLE
Gene networks and metabolomic screening analysis revealed specific pathways of amino acid and acylcarnitine profile alterations in blood plasma of patients with Parkinson's disease and vascular parkinsonism. *A.A. Makarova, P.M. Melnikova, A.D. Rogachev, P.S. Demenkov, T.V. Ivanisenko, E.V. Predtechenskaya, S.Y. Karmanov, V.V. Koval, A.G. Pokrovsky, I.N. Lavrik, N.A. Kolchanov, V.A. Ivanisenko*

940 REVIEW
Ontologies in modelling and analysing of big genetic data. *N.L. Podkolodnyy, O.A. Podkolodnaya, V.A. Ivanisenko, M.A. Marchenko*

950 ORIGINAL ARTICLE
PlantReg: the reconstruction of links between transcription factor regulatory networks and biological processes under their control. *V.V. Lavrekha, N.A. Omelyanchuk, A.G. Bogomolov, E.V. Zemlyanskaya*

960 ORIGINAL ARTICLE
Computational identification of promising genetic markers associated with molecular mechanisms of reduced rice resistance to *Rhizoctonia solani* under excess nitrogen fertilization using gene network reconstruction and analysis methods.
E.A. Antropova, A.R. Volyanskaya, A.V. Adamovskaya, P.S. Demenkov, I.V. Yatsyk, T.V. Ivanisenko, Y.L. Orlov, Ch. Haoyu, M. Chen, V.A. Ivanisenko

974 REVIEW
Reconstruction of gene regulatory networks from single cell transcriptomic data.
M.A. Rybakov, N.A. Omelyanchuk, E.V. Zemlyanskaya

982 ORIGINAL ARTICLE
Comparison of brain activity metrics in Chinese and Russian students while perceiving information referencing self or others. *Q. Si, J. Tian, V.A. Savostyanov, D.A. Lebedkin, A.V. Bocharov, A.N. Savostyanov*

Biomedicine

993 ORIGINAL ARTICLE
A concept of natural genome reconstruction. Part 2. Effect of extracellular double-stranded DNA fragments on hematopoietic stem cells.
V.S. Ruzanova, S.G. Oshikhmina, A.S. Proskurina, G.S. Ritter, S.S. Kirikovich, E.V. Levites, Y.R. Efremov, T.V. Karamysheva, M.I. Meschaninova, A.L. Mamaev, O.S. Taranov, A.S. Bogachev, S.V. Sidorov, S.D. Nikonov, O.Y. Leplina, A.A. Ostanin, E.R. Chernykh, N.A. Kolchanov, E.V. Dolgova, S.S. Bogachev

1008 ORIGINAL ARTICLE
GlucoGenes®, a database of genes and proteins associated with glucose metabolism disorders, its description and applications in bioinformatics research.
V.V. Klimontov, K.S. Shishin, R.A. Ivanov, M.P. Ponomarenko, K.A. Zolotareva, S.A. Lashin

1018 ORIGINAL ARTICLE
Association of autistic personality traits in non-clinical subjects with EEG scores during the facial video viewing. *A.N. Savostyanov, D.A. Kuleshov, D.I. Klemeshova, M.S. Vlasov, A.E. Saprygin*

1025 Alphabetical author index for the list of papers published in the journal in 2024

© Siberian Branch RAS, 2024
© Institute of Cytology and Genetics, SB RAS, 2024

Вавиловский журнал генетики и селекции. 2024;28(8):807
doi 10.18699/vjgb-24-109

Уважаемые коллеги, дорогие читатели!

Представляем Вашему вниманию очередной выпуск «Вавиловского журнала генетики и селекции», посвященный вопросам биоинформатики и системной компьютерной биологии. С появлением массового секвенирования геномов началась разработка обширного арсенала методов биоинформатики для анализа структурно-функциональной организации геномов, генов и кодируемых ими РНК и белков, а также подходов системной компьютерной биологии, ориентированных на реконструкцию, анализ и моделирование генных сетей, контролирующих формирование фенотипических признаков организмов на основе информации, закодированной в геномах, и функционирующих на различных иерархических уровнях организации живых систем (начиная с геномов, генов, белков, метаболических путей и генных сетей, включая клетки и ткани, и заканчивая целостными организмами).

Методы биоинформатики и системной компьютерной биологии прочно вошли в арсенал исследовательских инструментов, используемых во всех областях наук о жизни. Однако в последнее десятилетие их значимость еще более возросла: в связи со стремительным развитием омиксных технологий (геномики, транскриптомики, протеомики, метаболомики) и других высокопроизводительных методов экспериментального исследования молекулярно-генетических систем и процессов в генетике произошел информационный взрыв. Она стала главным источником больших данных, перегнав по темпам роста не только все другие науки и технологии, но и мировые социальные сети. Огромные объемы и сложность накапливаемых в настоящее время больших генетических данных требуют создания информационно-программных комплексов, основанных на новом поколении методов биоинформатики и системной компьютерной биологии, использующих вычислительные конвейеры, реализующих сложные сценарии анализа и интегрирующих большое количество разнообразных программных продуктов и баз данных, полученных в том числе с помощью методов искусственного интеллекта.

Новая эра больших данных, в которую входят науки о жизни, требует трансформации базовых подходов биоинформатики и системной компьютерной биологии. В чем это проявляется? Во-первых, в широком применении методов искусственного интеллекта и их интеграции с классическими методами биоинформатики и системной компьютерной биологии. Во-вторых, в разработке на этой основе нового поколения информационно-программных систем для планирования экспериментов по проверке результатов компьютерных предсказаний, полученных при анализе больших генетических данных. Движение в этом направлении будет означать фактически смену базовой парадигмы исследований: от науки, направляемой гипотезами, к науке, направляемой большими данными. Биоинформатика и системная компьютерная биология, которым посвящен данный выпуск журнала, находятся на острие этого движения к новой биологии.

Научный редактор выпуска
академик Н.А. Колчанов,
научный руководитель ФИЦ ИЦиГ СО РАН

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):808-821

doi 10.18699/vjgb-24-89

Кандидатные SNP-маркеры изменения экспрессии гена *SCN9A* человека в качестве интегратора генерации, чувства, ответа на боль и анестезии

П.А. Доценко [1, 2, 3], К.А. Золотарева [1], Р.А. Иванов[1], И.В. Чадаева [1, 3], Н.Л. Подколодный [1, 4], В.А. Иванисенко [1, 2, 3], П.С. Деменков [1], С.А. Лашин [1, 2], М.П. Пономаренко [1, 3]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[4] Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия
pon@bionet.nsc.ru

Аннотация. В настоящей работе впервые проведен комплексный биоинформатический анализ генов человека, связанных с генерацией, чувством и ответом на боль наряду с обезболиванием, которые были представлены 568 генами человека согласно базе данных NCBI Gene (дата обращения 15.09.2024). Ген *SCN9A* человека (sodium voltage-gated channel α subunit 9) передачи сигналов о повреждении тканей от сенсорных нейронов в центральную нервную систему был единственным среди исследуемых 568 генов, который вовлечен во все анализируемые процессы как ген-интегратор для них. Сначала с использованием созданного нами ранее инструмента OrthoWeb для каждого гена оценили таксон ближайшего общего предка всех организмов, у которого расшифрована ДНК этого гена (т. е. индекс филостратиграфического возраста, PAI). Среднеарифметическая оценка PAI для всех анализируемых генов, а также его значение для гена *SCN9A*, интегратора генерации, чувства и ответа на боль наряду с анестезией, оказались равными 4. На эволюционной шкале Киотской энциклопедии генов и геномов (KEGG) это соответствует таксону Chordata, у одних из самых древних представителей которого произошла специализация центральной и периферической нервной системы. Далее с помощью созданной нами системы ANDSystem мы выявили фосфорилирование ионных каналов как краеугольного камня в генерации, чувстве, ответе на боль и обезболивание, которое определяет эффективность передачи сигналов из периферической в центральную нервную систему. Этот вывод согласуется с литературными данными о ключевой роли эффективной передачи сигналов периферической нервной системы в центральную при коррекции циркадного ритма человека через фактическую детекцию фоторецепторами смены ночной темноты на дневное освещение, а также при определении направления на источник звука слуховыми ядрами мозга, формировании ответа на холодовой стресс и при координации движений у человека. Затем с использованием ранее созданной нами базы данных Human_SNP_TATAdb был предложен 21 кандидатный SNP-маркер значимого увеличения и уменьшения экспрессии гена *SCN9A* человека. Наконец, отношение встречаемости этих SNP-маркеров сравнили с полногеномным отношением, которое было оценено консорциумом «1000 геномов». В результате обнаружено, что *SCN9A* как ген-интегратор генерации, чувства, ответа на боль наряду с анестезией подвержен естественному отбору против снижения его экспрессии для поддержания высокого уровня контроля состояния организма и параметров внешней среды.

Ключевые слова: человек; TBP; SNP; промотор; ген-интегратор; *SCN9A*; изменение экспрессии; генерация боли; чувство боли; ответ на боль; анестезия.

Для цитирования: Доценко П.А., Золотарева К.А., Иванов Р.А., Чадаева И.В., Подколодный Н.Л., Иванисенко В.А., Деменков П.С., Лашин С.А., Пономаренко М.П. Кандидатные SNP-маркеры изменения экспрессии гена *SCN9A* человека в качестве интегратора генерации, чувства, ответа на боль и анестезии. *Вавиловский журнал генетики и селекции*. 2024;28(8):808-821. doi 10.18699/vjgb-24-89

Финансирование. Работа выполнена при поддержке Федеральной научно-технической программы развития генетических технологий России. Авторы выражают благодарность ЦКП «Биоинформатика» за вычислительные ресурсы по бюджетному проекту FWNR-2022-0020.

Candidate SNP markers of changes in the expression levels of the human *SCN9A* gene as a hub gene for pain generation, perception, response and anesthesia

P.A. Dotsenko [1, 2, 3], K.A. Zolotareva [1], R.A. Ivanov[1], I.V. Chadaeva [1, 3], N.L. Podkolodnyy [1, 4], V.A. Ivanisenko [1, 2, 3], P.S. Demenkov [1], S.A. Lashin [1, 2], M.P. Ponomarenko [1, 3]

© Доценко П.А., Золотарева К.А., Иванов Р.А., Чадаева И.В., Подколодный Н.Л., Иванисенко В.А., Деменков П.С., Лашин С.А., Пономаренко М.П., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
[4] Institute of Computational Mathematics and Mathematical Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
pon@bionet.nsc.ru

Abstract. In this work, we for the first time performed a comprehensive bioinformatics analysis of 568 human genes that, according to the NCBI Gene database as on September 15, 2024, were associated with pain generation, perception and anesthesia. The *SCN9A* gene encoding the sodium voltage-gated channel α subunit 9 and expressed in sensory neurons for transferring signals to the central nervous system about tissue damage was the only one involved in all the processes of interest at once as a hub gene. First, with our tool called OrthoWeb, we estimated the phylostratigraphic age indices (PAIs) for each of the genes, that is, identified the taxon of the most recent common ancestor of the organisms for which that gene has been sequenced. The mean PAI for all genes under study, including *SCN9A* as a hub gene for pain generation, perception, response and anesthesia, was '4'. On the evolutionary scale by the Kyoto Encyclopedia of Genes and Genomes (KEGG), the ancestor is the phylum Chordata, some of the most ancient of which evolved the central and the peripheral nervous system. Next, with our tool called ANDSystem, we found that phosphorylation of ion channels is a centerpiece in pain generation, perception, response and anesthesia, on which the efficiency of signal transduction from the peripheral to the central system depends. This conclusion was consistent with literature data on a key role an efficient signal transduction from the peripheral to the central system for adjusting the human circadian rhythm through detection of a change from the dark of night to the light of day and for identification of the direction of the source of sound by auditory brainstem nuclei, for generating the response to cold stress and for physical coordination. 21 candidate SNP marker of significant SCN9A over- and underexpression. Finally, the ratio of SCN9A upregulating to downregulating SNPs was compared to that for all known human genes estimated by the 1000 Genomes Project Consortium. It was found that *SCN9A* as a hub gene for pain generation, perception, pain response and anesthesia is acted on by natural selection against its downregulation, to keep the nervous system highly informed on the status of the organism and the environment.

Key words: human; TBP; SNP; promoter; hub gene; *SCN9A*; expression change; pain generation; pain perception; pain response; anesthesia.

For citation: Dotsenko P.A., Zolotareva K.A., Ivanov R.A., Chadaeva I.V., Podkolodnyy N.L., Ivanisenko V.A., Demenkov P.S., Lashin S.A., Ponomarenko M.P. Candidate SNP markers of changes in the expression levels of the human *SCN9A* gene as a hub gene for pain generation, perception, response and anesthesia . *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):808-821. doi 10.18699/vjgb-24-89

Введение

В 2020 г. Международная ассоциация по изучению боли (International Association for the Study of Pain, IASP) путем голосования Совета единогласно приняла текущее определение боли как «неприятное сенсорное и эмоциональное переживание, связанное с фактическим или потенциальным повреждением тканей или напоминающее таковое» – дословно: "An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" (Raja et al., 2020). Дополнительно были приняты шесть ключевых примечаний для адекватного применения термина «боль» в зависимости от контекста (Raja et al., 2020). В первую очередь под болью было рекомендовано понимать личное эмоциональное переживание индивида, на которое могут влиять биологические, психологические и социальные факторы. Кроме того, боль отличается от возбуждаемой разнообразными стимулами пульсирующей активности чувствительных нервных волокон периферической и центральной нервной системы (ноцицепция и ноцицепция, в более узком смысле «физиологическая боль»). Представление о боли возникает у индивида как часть его жизненного опыта. Также Совет IASP рекомендовал учитывать мнение пациента об ощущении им боли. Несмотря на адаптивную роль боли, она может негативно влиять на социально-психологическое благополучие и функционирование организма человека. Наконец, словесное описание боли является одним из способов широкого спектра выражения чувства боли индивидом, неспособность которого выразить ощущение им боли не исключает возможности того, что он ее испытывает.

В связи с этим объектом изучения была «физиологическая боль», которую здесь и далее мы будем называть «болью», прежде всего, для краткости, а также из-за широкого использования термина «боль» именно в этом узком смысле такими признанными источниками научных данных, как свободно доступные базы данных NCBI Gene (Brown et al., 2015) и Gene Ontology (Gene Ontology Consortium, 2015), на данные которых мы опираемся в настоящей работе.

При этом впервые ведется комплексное биоинформатическое исследование боли одновременно с обезболиванием в качестве практической сервисной услуги анестезии в рамках прикладной медицины, когда лечение пациента с необходимостью требует полностью или частично понизить у него чувствительность к состоянию собственного организма и окружающей среды с использованием медикаментозных средств с целью поддержания гомеостаза согласно рекомендациям Ассоциации анестезиологов (Klein et al., 2021; Lucas et al., 2021). Актуальность изучения боли и анестезии настолько высока, что уже в текущем 2024 г. в базе данных PubMed (Lu, 2011) представлены соответ-

ственно 49 305 и 3782 оригинальных научных статьи о них (дата обращения 15.09.2024). Поэтому здесь с помощью ранее созданных нами свободно доступных web-сервисов OrthoWeb (Mustafin et al., 2020) и ANDSystem (Ivanisenko et al., 2015), а также базы данных Human_SNP_TATAdb (Filonov et al., 2023) были проанализированы 568 генов человека, связанных с генерацией, чувством, ответом на боль и анестезией согласно базе данных NCBI Gene (Brown et al., 2015) (дата обращения 15.09.2024). Полученные результаты были верифицированы в сравнении с результатами независимых web-сервисов PANTHER (Mi et al., 2021), DAVID (Sherman et al., 2022), STRING (Szklarczyk et al., 2023), Metascape (Zhou et al., 2019) и GeneMania (Warde-Farley et al., 2010), а также базы данных ClinVar (Landrum et al., 2014) и полногеномных оценок, которые были получены консорциумом «1000 геномов» (1000 Genomes Project Consortium et al., 2012), в рамках дилеммы Холдейна (Haldane, 1957) и теории нейтральной эволюции (Kimura, 1968).

Материалы и методы

Исследуемые гены человека. В работе было изучено 568 генов человека (n = 568). Список генов составлен с использованием диалоговой системы автоматизированного поиска данных по запросам в формате «*Homo sapiens» AND* «[признак гена]» к общедоступной базе данных NCBI Gene (Brown et al., 2015) (дата обращения 15.09.2024). При этом были установлены фильтры *Protein-coding genes*, *Genomic*, *Annotated genes*, *Ensembl*, *Current*, которые предназначены для отбора наиболее полно аннотированных белок-кодирующих генов человека.

Индекс филостратиграфического возраста (PAI) исследуемых генов человека. С помощью ранее созданного нами web-сервиса OrthoWeb (Mustafin et al., 2020) независимо для каждого исследуемого гена человека находили все биологические виды, для которых свободно доступны секвенированные ортологи данного гена, и на этой основе определяли «ближайшего общего предка» этих видов (Samet, 1985; Sun et al., 2008; Morozova et al., 2020) в качестве оценки филостратиграфического возраста (PAI) рассматриваемого гена по шкале Киотской энциклопедии генов и геномов KEGG (Kanehisa, Goto, 2000).

Ассоциативную сеть интеграции генерации, чувства, ответа на боль и анестезии реконструировали с использованием ранее созданной нами системы ANDSystem (Ivanisenko et al., 2015). Полученные результаты верифицировали в сравнении с результатами независимых свободно доступных web-сервисов PANTHER (Mi et al., 2021), DAVID (Sherman et al., 2022), STRING (Szklarczyk et al., 2023), Metascape (Zhou et al., 2019) и GeneMania (Warde-Farley et al., 2010). Степень согласия между результатами этих web-сервисов и ANDSystem (Ivanisenko et al., 2015) оценивали путем поиска соответствующих публикаций в свободно доступной базе данных PubMed (Lu, 2011).

Курируемая аннотация влияния снижения и повышения экспрессии генов человека на генерацию, чувство, ответ на боль и анестезию. Влияние уменьшения и увеличения экспрессии гена *SCN9A* человека на изменение генерации, чувства, ответа на боль и анестезии оценивали посредством поиска соответствующих публикаций в свободно доступной базе данных PubMed (Lu, 2011).

Влияние вариантов однонуклеотидного полиморфизма (SNP) промоторов генов человека на уменьшение и увеличение уровня экспрессии этого гена. Оценки уровней статистической значимости уменьшения и увеличения экспрессии генов человека для минорных в сравнении с референсными аллелями SNP промоторов этих генов были взяты из базы знаний Human_SNP_TATAdb (Filonov et al., 2023).

Верификация оценок влияния SNP промоторов генов человека на их экспрессию. Выборочную верификацию полученных оценок *in silico* для влияния SNP промоторов генов человека на экспрессию этих генов провели с помощью баз данных ClinVar (Landrum et al., 2014) и PubMed (Lu, 2011), а также литературных данных консорциума проекта «1000 геномов» (Lowy-Gallego et al., 2019) для оценки встречаемости таких SNP во всем референсном геноме человека с учетом дилеммы Холдейна (Haldane, 1957) и теории нейтральной эволюции (Kimura, 1968).

Статистический анализ. Статистические критерии Колмогорова–Смирнова и биномиального распределения проверяли в пакете STATISTICA выпуск №12 (Statsoft™, США).

Результаты

Ген *SCN9A* человека интеграции генерации, чувства, ответа на боль и анестезии

В настоящей работе были изучены 568 генов человека, выборка которых составлена с помощью базы данных NCBI Gene (Brown et al., 2015) (см. Материалы и методы). Из общей выборки 553 гена человека были связаны с болью, 231 – с генерацией боли, 84 – с чувством боли, 39 – с ответом на боль, тогда как 28 – с анестезией (рис. 1, *А*). На диаграмме Венна всех возможных пересечений вышеупомянутых групп генов друг с другом (см. рис. 1, *А*) бордовым шрифтом обозначен единственный общий для всех этих групп ген *SCN9A* человека, который кодирует субъединицу управляемого напряжением натриевого канала Nav1.7 типа IXα (sodium voltage-gated channel α subunit 9) в сенсорных нейронах для передачи ими сигналов о повреждении тканей в центральную нервную систему. Поэтому в рамках данного исследования было принято рассматривать ген *SCN9A* человека как интегратор генерации, чувства, ответа на боль и анестезии.

Группы генов человека, связанных с генерацией, чувством, ответом на боль и анестезией, неразличимы по индексам PAI

Для каждого из всех 568 генов человека, анализируемых в этой работе, мы оценили значения индекса филостратиграфического возраста (PAI). Гистограмма количества исследуемых генов для каждого из 16 временных интервалов шкалы PAI согласно Киотской энциклопедии генов и геномов KEGG (Kanehisa, Goto, 2000) показана на рис. 1, *Б*. Полученные эволюционные оценки PAI генов человека, связанных с генерацией, чувством, ответом на

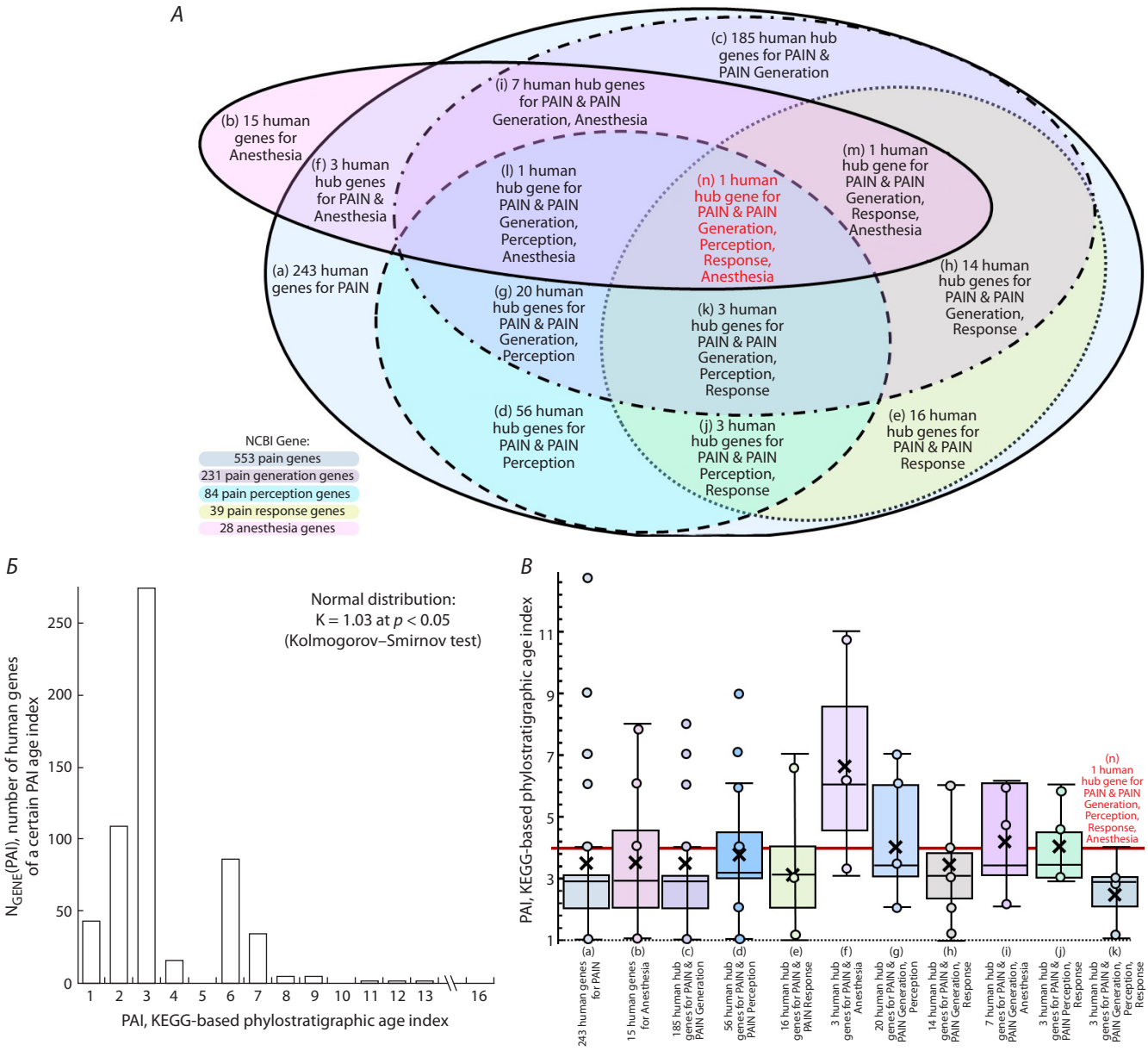


Рис. 1. Исследуемые гены человека, полученные по запросу «*Homo sapiens*» *AND* «[признак гена]» в базе данных NCBI Gene (Brown et al., 2015) с фильтрами *Protein-coding*, *Genomic*, *Annotated genes*, *Ensembl*, *Current*.

A – диаграмма Венна (всего 568 генов человека): для «[признака гена]» = "Pain" было найдено 553 гена человека, для "Generation of pain" – 231, для "Perception of pain" – 84, для "Response to pain" – 39, для "Anesthesia" – 28. Бордовым цветом обозначен единственный ген *SCN9A* человека интеграции генерации, чувства, ответа на боль и анестезии. *Б* – индекс филостратиграфического возраста PAI изученных генов, имеющий нормальное распределение (тест Колмогорова–Смирнова: K = 1.03, $p < 0.05$). *B* – диаграмма «ящик с усами»; высота – размах от 25 до 75 % квартиля, IQR; линия – медиана, 50 % квартиль; крест – среднеарифметическое; планка ошибок "I" – 95 % доверительный интервал; кружки – гены. Шкала PAI: 1 = Cellular organism, 4100 млн лет, Mr (Bell et al., 2015), 2 = Eukaryota, 1850 Mr (Leander, 2020), 3 = Metazoa, 665 Mr (Maloof et al., 2010a), 4 = Chordata, 541 Mr (Maloof et al., 2010b), 5 = Craniata, 535 Mr (Maloof et al., 2010b), 6 = Vertebrata, 525 Mr (Shu et al., 1999), 7 = Euteleostomi, 420 Mr (Diogo, 2007), 8 = Mammalia, 225 Mr (Datta, 2005), 9 = Eutheria, 160 Mr (Luo et al., 2011), 10 = Euarchontoglires, 65 Mr (Kumar et al., 2013), 11 = Primates, 55 Mr (Chatterjee et al., 2009), 12 = Haplorrhini, 50 Mr (Dunn et al., 2016), 13 = Catarrhini, 44 Mr (Harrison, 2013), 14 = Hominidae, 17 Mr (Hey, 2005), 15 = *Homo*, 2.8 Mr (Schrenk et al., 2014), 16 = *Homo sapiens*, 0.35 Mr (Scerri et al., 2018).

боль и анестезией, статистически значимо отвечают нормальному распределению (критерий Колмогорова–Смирнова: K = 1.03, $p < 0.05$). В рамках центральной предельной теоремы (Kwak, Kim, 2017) это может означать, что оценки PAI отражают интеграцию большого разнообразия жизненно важных признаков боли. Поэтому мы сфокусировали данную работу на гене *SCN9A* человека в качестве интегратора генерации, чувства, ответа на боль и анестезию.

Гипотетическая связь эволюционной оценки PAI генов человека, ассоциированных с генерацией, чувством, ответом на боль и анестезией, верифицирована в виде диаграммы «ящик с усами» для пересечений между этими группами генов (см. рис. 1, *B*). Пересечения групп генов человека неразличимы по индексу PAI друг от друга, а также от *SCN9A* как гена-интегратора признаков боли у человека (см. рис. 1, *A*). Это стало еще одним аргументом для концентрации усилий на гене *SCN9A*.

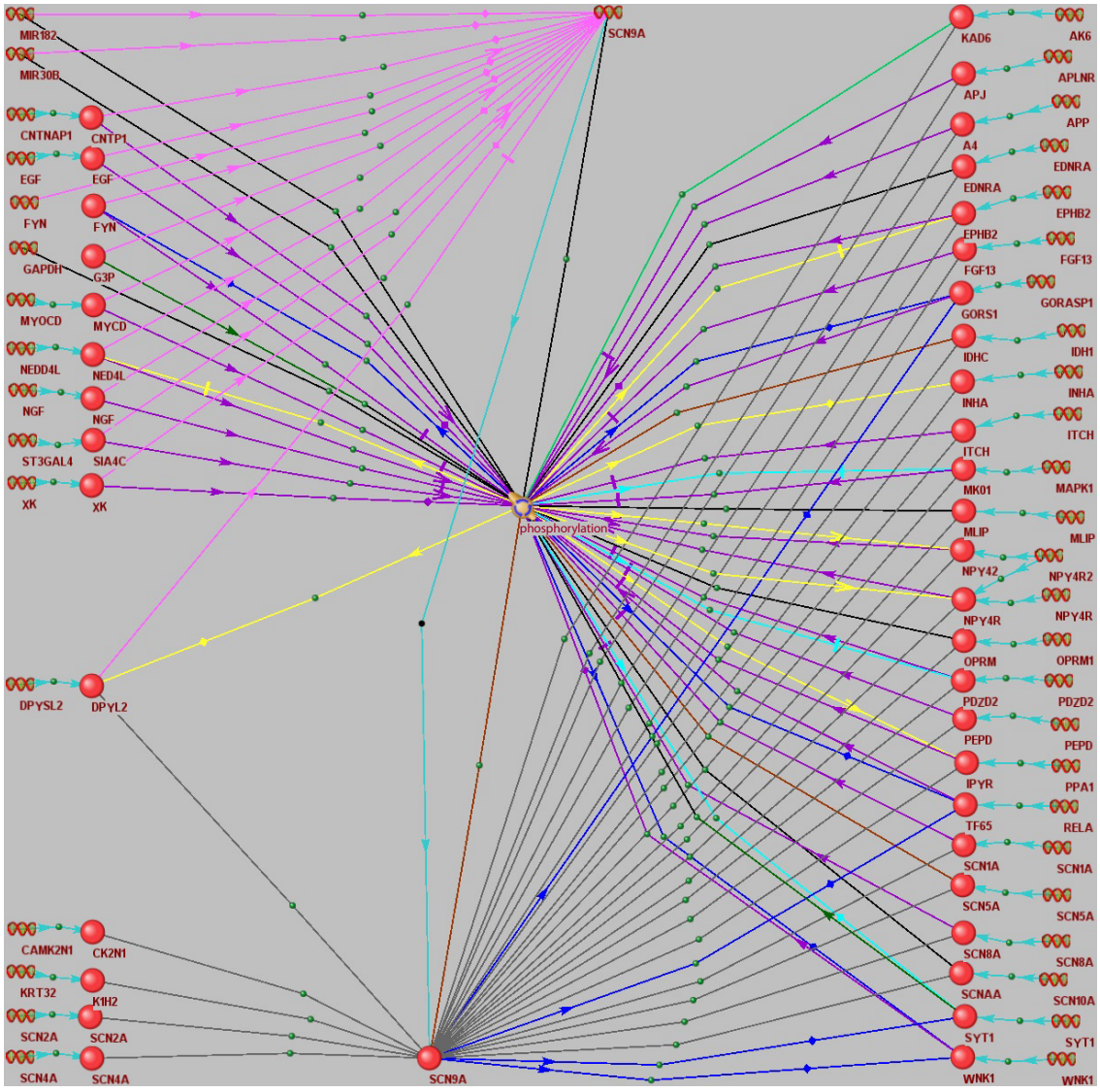


Рис. 2. Ассоциативная генная сеть функционирования гена *SCN9A* человека, его белкового продукта и их ближайших партнеров в организме человека, построенная ранее созданным нами инструментом ANDSystem (Ivanisenko et al., 2015) на основе автоматического анализа текстов свободно доступных научных статей и баз данных по запросу «[список генов] [Только их непосредственные связи] *Genes Proteins Pathway*» для [список генов] = «*SCN9A*».

Обозначения: – ген, ● – белок; – фосфорилирование как самый статистически значимый биологический процесс, в который достоверно вовлечены все найденные гены и белки ($P_{ADJ} < 10^{-13}$, точный тест Фишера с поправкой Бонферрони на множественные сравнения). Стрелки: острая – активация, тупая – ингибирование, ненаправленная – вовлеченность; желтая – активность, синяя – транспорт, черная – контакт, фиолетовая – функция, бордовый – регуляция, бирюзово-синий – экспрессия.

Ассоциативная сеть интеграции генерации, чувства, ответа на боль и анестезии

С использованием системы ANDSystem (Ivanisenko et al., 2015) была построена ассоциативная генная сеть функционирования *SCN9A* человека в качестве интегратора генерации, чувства, ответа на боль и анестезию (рис. 2). В центре вверху на этом рисунке читатель может видеть целевой ген *SCN9A* человека, в центре внизу – его белковый продукт, в центре посередине – фосфорилирование в качестве молекулярно-генетического процесса, который статистически достоверно наиболее часто упоминается в контексте этого гена в свободно доступной научной литературе и базах данных согласно оценке ANDSystem (Ivanisenko et al., 2015).

Слева в центре на рис. 2 находится единственный ген *DPYSL2* человека, который ассоциирован как с самим геном *SCN9A* человека, так и с его белком, а также вовлечен в фосфорилирование. Кроме того, слева внизу сгруппированы четыре гена и кодируемые ими белки, которые взаимодействуют с белком SCN9A человека, а слева вверху – 11 генов человека и их продукты, взаимодействующие с геном *SCN9A* и вовлеченные в фосфорилирование. Наконец, остальные 25 генов человека и кодируемые ими белки взаимодействуют с белком SCN9A и тоже вовлечены в фосфорилирование (см. рис. 2, справа). Всего на рис. 2 представлено 42 гена человека, 14 из которых были в исходном списке из 568 генов, связанных с генерацией, чувством, ответом на боль и анестезией (см. рис. 1).

Пересечение списков из 42 и 568 генов является достоверным в масштабе референсного генома человека, в котором содержится 19424 аннотированных белок-кодирующих гена согласно базе данных NCBI Gene (Brown et al., 2015), при фильтрах *Ensembl*, *Current*, *Protein-coding genes*, *Genomic*, *Annotated genes* (дата обращения 20.08.2024): $p < 10^{-6}$ по критерию биномиального распределения.

Этот результат означает, что система ANDSystem (Ivanisenko et al., 2015) по одному гену *SCN9A* как интегратору генерации, чувства, ответа на боль и анестезии (см. рис. 1), статистически достоверно реконструировала список генов человека (см. рис. 2), которые связаны с этими биологическими процессами согласно базе данных NCBI Gene (Brown et al., 2015).

Верификация результата ANDSystem с использованием независимых web-сервисов оценки обогащения групп генов терминами Gene Ontology

Сравнение результата ANDSystem (Ivanisenko et al., 2015) – фосфорилирование как наиболее статистически значимый биологический процесс для генерации, чувства, ответа на боль и анестезии (см. рис. 2) – с результатами независимых web-сервисов оценки обогащения групп генов терминами Gene Ontology (Gene Ontology Consortium, 2015) представлено в табл. 1.

Таблица 1. Сравнение результата web-сервиса ANDSystem (Ivanisenko et al., 2015) с результатами других web-сервисов оценки обогащения групп генов терминами Gene Ontology (Gene Ontology Consortium, 2015)

№ п/п	Самый обогащенный термин Gene Ontology, GO			Связь фосфорилирования, найденного web-сервисом ANDSystem (Ivanisenko et al., 2015), с независимо найденным наилучшим термином GO
	Web-сервис	Раздел GO: ID	P_{ADJ}	
		Биологический процесс		
1	PANTHER (Mi et al., 2021)	GO:0086002 ~ cardiac muscle cell action potential involved in contraction	10^{-9}	Согласно исчерпывающему клиническому обзору (Iyer et al., 2007): фосфорилирование кальциевого канала в клетках сердечной мышцы усиливает приток Ca^{2+}, что увеличивает высвобождение Ca^{2+} из саркоплазматического ретикулума и в конечном итоге повышает мощность сокращения сердца
2	DAVID (Sherman et al., 2022)	GO:0086010 ~ membrane depolarization during action potential	10^{-9}	В рамках клеточной модели боли с использованием клеток линии HEK293T человека (Kerth et al., 2021): замена I848T в SCN9A создает новый сайт фосфорилирования, улучшающий чувствительность и возбудимость нейрона за счет увеличения диапазона (потенциала) степени деполяризации его мембраны
3	STRING (Szklarczyk et al., 2023)	GO:0043269 ~ regulation of ion transport	10^{-11}	В биомедицинской тканевой модели боли с использованием культуры спинальных ганглиев крысы (Stamboulian et al., 2010): фосфорилирование Scn9A регулирует транспорт ионов по механизму варьирования порога активации и длительности инактивации потенциал-зависимого натриевого канала
4	Metascape (Zhou et al., 2019)	GO:0044057 ~ regulation of system process	10^{-9}	Метаанализ свободно доступных информационных ресурсов и баз данных по традиционной китайской медицине (Shuyuan, Haoyu, 2023) указал на "GO:0042327 ~ positive regulation of phosphorylation" и "GO:0044057 ~ regulation of system process" среди наилучших GO-терминов, характеризующих лечение преждевременных сердечных сокращений на основе использования корня и корневища *Nardostachys jatamansi*
5	GeneMania (Warde-Farley et al., 2010)	GO:0034706 ~ sodium channel complex	10^{-18}	В биомедицинской тканевой модели боли с использованием нейронов Пуркинье мозжечка, быстро изолированных у двухнедельных мышей (Grieco et al., 2002): конститутивное фосфорилирование комплекса натриевых каналов нейронов Пуркинье необходимо для поддержания функционального блокирующего элемента и возобновления натриевого тока
		Молекулярная функция		
6	PANTHER (Mi et al., 2021)	GO:0005248 ~ voltage-gated sodium channel activity	10^{-10}	В рамках субклеточной модели боли с использованием препаратов мембраны клеток линии HEK293T человека (Sokolov et al., 2018): фосфорилирование *SCN9A* увеличивает проводимость этого зависимого от напряжения натриевого канала для ионов Na^+
7	DAVID (Sherman et al., 2022)		10^{-9}	
8	STRING (Szklarczyk et al., 2023)		10^{-9}	
		Клеточный компонент		
9	PANTHER (Mi et al., 2021)	GO:0001518 ~ voltage-gated sodium channel complex	10^{-12}	В рамках субклеточной модели боли с использованием препаратов мембраны клеток линии HEK293T человека (Sokolov et al., 2018): фосфорилирование *SCN9A* способствует связыванию вспомогательной β3-субъединицы, смещающей состояние устойчивой инактивации зависимого от напряжения натриевого канала к более быстрому восстановлению после инактивации в составе такого их комплекса
10	DAVID (Sherman et al., 2022)		10^{-10}	
11	STRING (Szklarczyk et al., 2023)		10^{-10}	

Примечание. P_{ADJ} – оценка уровня статистической значимости с поправкой на множественные сравнения, используемая web-сервисом оценки обогащения групп генов терминами Gene Ontology.

Таблица 2. Клиническое проявление уменьшения и увеличения уровня экспрессии гена *SCN9A* человека в изменениях генерации, чувства, ответа на боль и анестезии согласно базе данных PubMed (Lu, 2011)

№ п/п	Фенотипический признак	Изменения уровня экспрессии гена *SCN9A* человека	
		Уменьшение	Увеличение
1	Генерация боли	В модели невропатической боли с использованием мыши линии C57BL/6 (Palomes-Borrajo et al., 2021): обработка повреждения нерва препаратом JQ1 снизила частоту генерации боли за счет снижения уровня экспрессии *SCN9A*, понизившего возбудимость сенсорных нейронов	Согласно исчерпывающему обзору (Baker, Nassar, 2020): мутационный рост активности SCN9A увеличивает частоту генерации боли за счет усиления возбудимости сенсорных нейронов
2	Ощущение боли	В биомедицинской модели боли с использованием *Scn9a*-дефицитных мышей (Shields et al., 2018): снижение возбудимости сенсорных нейронов малого и среднего диаметров из-за дефицита натриевых тетродотоксин-чувствительных каналов в них	Согласно исчерпывающему обзору (Dib-Hajj et al., 2007): мутационный рост активности SCN9A снижает порог активации и замедляет деактивацию потенциал-зависимых натриевых каналов, что повышает возбудимость сенсорных нейронов и приводит к эритромелалгии и пароксизмальному острому болевому расстройству
3	Ответ на боль	В модели невропатической боли с использованием мыши линии C57BL/6 (Palomes-Borrajo et al., 2021): обработка повреждения нерва препаратом JQ1 увеличила время ответа на болевой стимул в сравнении с контролем при снижении уровня экспрессии *SCN9A*	В модели спонтанной боли с использованием CRISPR/Cas9-трансгенной мыши с мутацией R185H как клинического маркера пациентов с мелковолокнистой периферической невропатией (Xue et al., 2022): меньше время задержки между тепловым болевым воздействием на лапу или хвост и отдергиванием лапы или хвоста
4	Анестезия	В биомедицинской модели боли с использованием *Scn9a*-дефицитных мышей (Shields et al., 2018): снижение экспрессии гена *SCN9A* человека и ингибирование его белковых продуктов может иметь обезболивающий эффект	В рамках метаанализа транскриптомов опухолей в сравнении с соседними здоровыми тканями (Garate et al., 2021): избыток экспрессии *SCN9A* – клинический маркер опухоли, отражающий специфический тип опухолевой боли и, соответственно, вспомогательное обезболивание наряду с основной противоопухолевой терапией (Cui et al., 2011)

Например, в первой строке таблицы можно видеть, что web-сервис PANTHER (Mi et al., 2021) для 42 генов человека, показанных на рис. 2, выявил “GO:0086002 ~ cardiac muscle cell action potential involved in contraction” как самый статистически значимый биологический процесс, куда вовлечены эти 42 гена ($P_{ADJ} < 10^{-9}$, оценка уровня значимости с поправкой на множественные сравнения). В последней ячейке этой строки цитируется обзор (Iyer et al., 2007): фосфорилирование кальциевого канала клеток сердечной мышцы усиливает приток Ca^{2+}, увеличивая высвобождение Ca^{2+} из саркоплазматического ретикулума и мощность сокращений сердца. В этом смысле результаты ANDSystem (Ivanisenko et al., 2015) и PANTHER (Mi et al., 2021) для 42 генов человека, показанных на рис. 2, согласуются между собой.

Всего в табл. 1 представлено 11 аналогичных соответствий между результатами ANDSystem и пяти независимых web-сервисов (PANTHER, DAVID, STRING, Metascape и GeneMania) оценки обогащения групп генов терминами Gene Ontology (Gene Ontology Consortium, 2015).

Влияние снижения и повышения экспрессии гена *SCN9A* человека как интегратора генерации, чувства, ответа на боль и анестезии на изменение этих признаков

На этом этапе исследования мы сначала с использованием текстовых запросов к свободно доступной базе данных PubMed (Lu, 2011) сделали курируемую аннотацию снижения и повышения уровня экспрессии гена *SCN9A* человека, сопоставив им статьи об их клинических проявлениях в изменениях генерации, чувства, ответа на боль и анестезии (табл. 2).

Затем в базе данных Human_SNP_TATAdb (Filonov et al., 2023) был найден 21 кандидатный SNP-маркер достоверного изменения сродства TBP к промоторам этого гена и, соответственно, уровней его экспрессии (табл. 3). Четыре из 21 SNP-маркера значимого изменения экспрессии гена *SCN9A* имеют известные клинические проявления (см. табл. 3), согласно базе данных ClinVar (Landrum et al., 2014). На примере одного из этих четырех клинических SNP-маркеров боли, rs201905758:T, показано (рис. 3), каким образом был выявлен этот SNP-маркер в автоматическом режиме работы web-сервиса SNP_TATA_Comparator (Ponomarenko et al., 2015), использующего библиотеку BioPerl (Stajich et al., 2002) для доступа к базам данных Ensembl (Zerbino et al., 2015) и dbSNP (Day, 2010) как официальным хранилищам референсных генома и вариома человека соответственно. Согласно базе данных ClinVar (Landrum et al., 2014), четыре из набора 21 SNP были клинически доказанными маркерами пароксизмального острого болевого расстройства (PEPD), нейропатии мелких волокон (SFN), первичной эритромелалгии (PE) и врожденной нечувствительности к боли из-за каналопатии (CIP) (см. табл. 3).

Все четыре клинически доказанных SNP-маркера гена *SCN9A* человека повышают уровни его экспрессии как гена-интегратора генерации, чувства, ответа на боль и анестезии (см. табл. 3, последняя колонка “Δ”). Поэтому

Таблица 3. Кандидатные маркеры SNP в проксимальных районах промоторов длиной 90 п. о. перед стартами транскриптов гена *SCN9A* человека, связанного с интеграцией, генерацией, чувством, ответом на боль и анестезией, согласно анализу *in silico* в данной работе, как показано на рис. 3 и документировано в базе данных Human_SNP_TATAdb (Filonov et al., 2023)

№ п/п	Кандидатный SNP-маркер				K_D, нМ, *in silico*		Уровень значимости			
	dbSNP ID:min (Day, 2010)	5′ фланг	WT →min	3 фланг	WT	min				
					$M_0 \pm SEM$	$M_0 \pm SEM$	Z	*p*	ρ	Δ
1	rs1341944281:G	gttttctaat	A→G	gttgatttcc	3.32 ± 0.34	6.41 ± 0.52	10.04	10^{-6}	A	
2	rs1470018720:C	ccgggcgcgc	T→C	ggggtgggga	86.89 ± 6.63	104.57 ± 8.03	3.42	10^{-3}	B	
3	rs1477103793:C	gcgcgctggg	A→C	tggggacccg	86.89 ± 6.63	120.27 ± 8.57	6.23	10^{-6}	A	↓
4	rs1559004384:G	atttcctgtt	T→G	tcattgtgtt	3.32 ± 0.34	3.86 ± 0.42	2.01	0.05	D	
5	rs933017443:C	gcggggctgc	T→C	ccctcgggga	56.13 ± 5.20	120.27 ± 8.57	13.04	10^{-6}	A	
6	rs1028575943:A	cgcgctggga	G→A	ggggacccgg	86.89 ± 6.63	66.72 ± 4.61	5.13	10^{-6}	A	
7	rs1038516207:A	gagtggagga	G→A	cgcgctggga	86.89 ± 6.63	69.11 ± 4.82	4.43	10^{-3}	B	
8	rs1282480960:G	ctaatattaa	C→G	tttcctgttt	3.32 ± 0.34	2.67 ± 0.27	3.03	10^{-2}	C	
9	rs1284056769:A	gagggagcaa	G→A	agggagggag	86.89 ± 6.63	75.42 ± 5.45	2.70	10^{-2}	C	
10	rs1343738748:T	gggagcaagg	G→T	ggagggaggg	86.89 ± 6.63	63.02 ± 5.10	5.78	10^{-6}	A	
11	rs1410144156:A	gctgggagga	G→A	gacccgggcg	86.89 ± 6.63	71.32 ± 4.53	3.98	10^{-3}	B	
12	rs1697331114:A	tgattattat	C→A	taagcaaaca	3.32 ± 0.34	2.37 ± 0.26	4.45	10^{-3}	B	
13	rs1700681124:T	gggctgctac	C→T	tcggggaggc	56.13 ± 5.20	35.36 ± 3.15	7.20	10^{-6}	A	
14	rs1700681309:A	gggaggcggg	G→A	agctgccctc	86.89 ± 6.63	40.86 ± 3.84	12.47	10^{-6}	A	
15	rs1700683197:A	agtggaggag	G→A	gcgctgggag	86.89 ± 6.63	30.32 ± 2.29	19.64	10^{-6}	A	↑
16	rs1700683375:A	gggaggagtg	G→A	ccgggcgcgc	86.89 ± 6.63	60.80 ± 4.59	6.65	10^{-6}	A	
17	rs890040570:A	cggcgcagct	G→A	aggaggcaaa	86.89 ± 6.63	64.68 ± 7.23	4.36	10^{-3}	B	
	База данных ClinVar (Landrum et al., 2014): клинические SNP-маркеры пароксизмального острого болевого расстройства, нейропатии мелких волокон, первичной эритромелалгии и врожденной нечувствительности к боли из-за каналопатии									
18	rs148362057:A	gcagtctgct	T→A	gcaggagggg	91.71 ± 6.30	41.82 ± 3.93	13.50	10^{-6}	A	
19	rs1881440:T	gccctggcag	G→T	tccacgggcg	91.71 ± 6.30	41.75 ± 3.64	14.19	10^{-6}	A	
20	rs201905758:A	gctacctcca	C→A	gaggcggggc	56.13 ± 5.20	47.62 ± 4.43	2.51	0.05	D	
21	rs201905758:T	gctacctcca	C→T	gaggcggggc	56.13 ± 5.20	43.97 ± 4.65	3.48	10^{-3}	B	

Примечание. WT и min – анцестральный (норма) и минорный (патология) аллели SNP; K_D – равновесная константа диссоциации комплекса «TBP-промотор», выраженная в наномолях на литр, нМ; M_0 и SEM – контекстно-зависимая оценка *in silico* и ее стандартизированная ошибка; Z, *p* и ρ – значение Z-статистики Фишера и уровень ее статистической значимости, а также эвристическая приоритизация оценок *in silico* от наилучшей (A) до наихудшей (D) в алфавитном порядке; Δ – повышение (↑) или снижение (↓) экспрессии гена *SCN9A* человека.

с использованием базы данных PubMed (Lu, 2011) мы осуществили курируемую аннотацию влияния суперэкспрессии гена *SCN9A* на генерацию, чувство, ответ на боль и анестезию (табл. 4). Прежде всего, согласно широкому кругу исчерпывающих клинических обзоров, например (Dabby, 2012; Bennett, Woods, 2014; Shields et al., 2018; Taub, Woolf, 2024), при патологиях PEPD, SFN и PE избыток SCN9A повышает генерацию, чувство, ответ на боль, тогда как низкомолекулярные ингибиторы *SCN9A* являются обезболивающими.

Что касается CIP, то, согласно клиническим наблюдениям (Kim et al., 2015), вторичная нечувствительность боли чередуется с первичными для нее приступами гиперчувствительности боли при PEPD, SFN и PE из-за избытка SCN9A. Это выглядит, как если бы гипервозбудимость нейронов из-за избытка управляемых напряжением натриевых каналов *SCN9A* могла бы «разрядить в них батарейку», которую необходимо заряжать для восстановления генерации, чувства и ответа на боль. В этом смысле оценки *in silico* суперэкспрессии *SCN9A* при всех клинически доказанных SNP-маркерах боли при PEPD, SFN, PE и CIP согласуются с проявлениями избытка *SCN9A* у пациентов с этими патологиями.

Сравнение встречаемости кандидатных SNP-маркеров изменения экспрессии гена *SCN9A* человека согласно этой работе с полногеномной встречаемостью таких SNP

В заключение этого исследования мы сравнили встречаемость кандидатных SNP-маркеров изменения экспрессии гена *SCN9A* человека (см. табл. 3) с полногеномной оценкой встречаемости таких SNP согласно проекту «1000 ге-

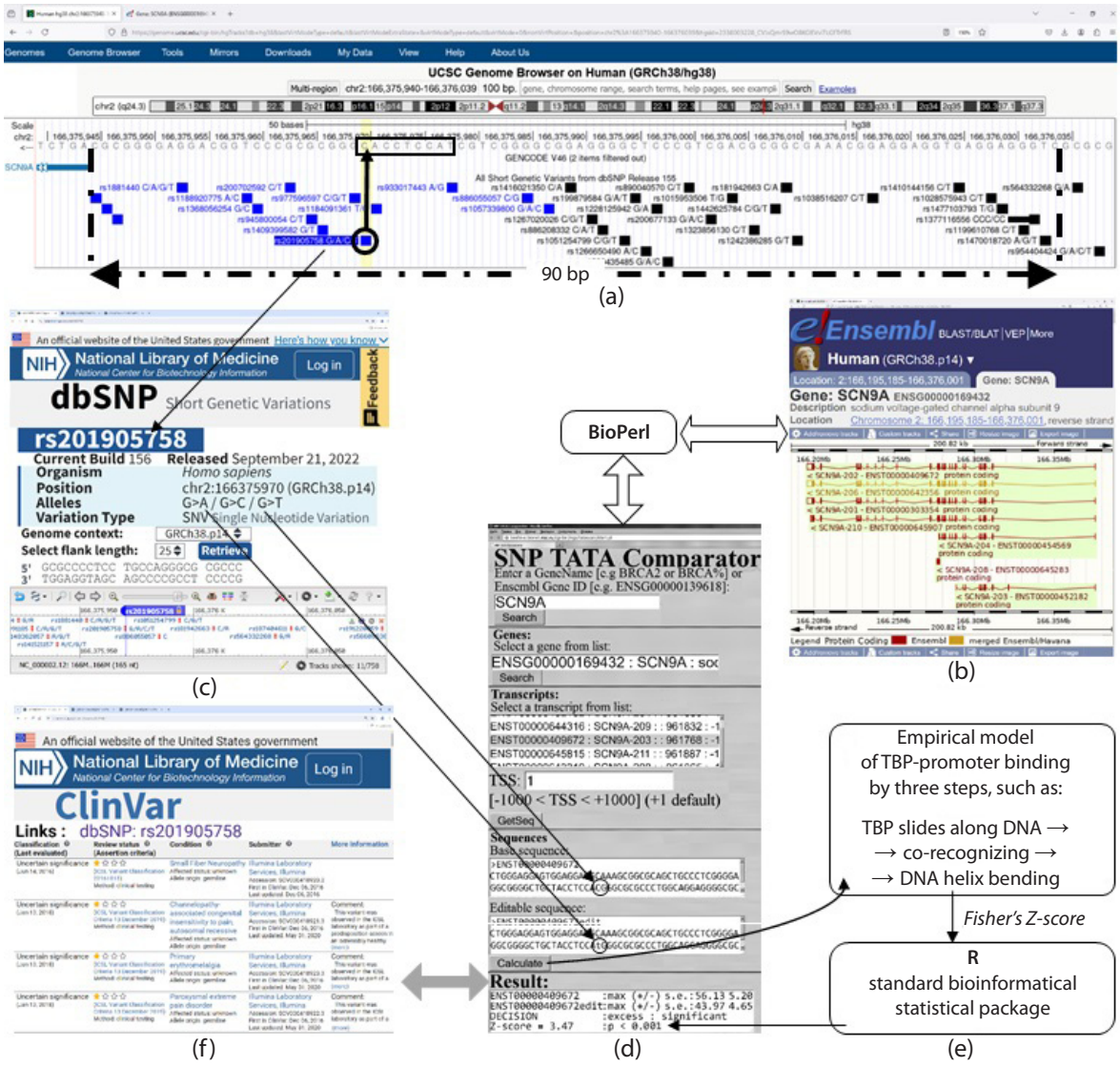


Рис. 3. Пример анализа кандидатного SNP-маркера rs201905758:T в проксимальном районе длиной 90 п. о. (двуглавая штрихпунктирная стрелка на панели (a)) перед стартом транскрипта SCN9A-203 с гена *SCN9A* человека согласно базе данных Ensembl (Zerbino et al., 2015) с использованием web-сервиса SNP_TATA_Comparator (Ponomarenko et al., 2015).

Панель (a) – визуализация анализируемого промотора с помощью web-сервиса UCSC Genome Browser (Raney et al., 2024); панель (b) – база данных Ensembl (Zerbino et al., 2015); панель (c) – описание SNP rs201905758 в базе данных dbSNP (Day, 2010); панели (d) и (e) – применение web-сервиса SNP_TATA_Comparator и принцип его работы соответственно (Ponomarenko et al., 2015); панель (f) – описание клинически доказанного SNP-маркера rs201905758:G⟶t патологий чувства боли согласно базе данных ClinVar (Landrum et al., 2014).

номов» (табл. 5). Индивидуальные геномы людей в сравнении с референсным геномом человека имеют в среднем 1000 SNP, среди которых 200 и 800 SNP соответствуют повышению и понижению сродства «TBP-промотор» и в итоге повышению и понижению уровней экспрессии генов человека с этими SNP (Kasowski et al., 2010; 1000 Genomes Project Consortium et al., 2012).

В рамках дилеммы Холдейна (Haldane, 1957) и теории нейтральной эволюции (Kimura, 1968) такое превышение повреждающих над улучшающими регуляторными SNP отражает нейтральный дрейф, от которого статистически значимо отличается встречаемость 21 кандидатного SNP-маркера изменения экспрессии гена *SCN9A* человека ($p < 10^{-6}$, критерий биномиального распределения) (см. табл. 5). Этот результат означает, что ген *SCN9A* подвергается давлению естественного отбора против снижения его экспрессии для поддержания высокого уровня контроля состояния организма и внешней среды.

Обсуждение

В настоящей работе с использованием ранее созданных нами свободно доступных web-сервисов OrthoWeb (Mustafin et al., 2020) и ANDSystem (Ivanisenko et al., 2015), а также базы данных Human_SNP_TATAdb (Filonov et al., 2023) мы впервые провели комплексное биоинформатическое исследование 568 генов человека, которые связаны с генерацией, чувством, ответом на боль и анестезией согласно базе данных NCBI Gene (Brown et al., 2015) (дата

Таблица 4. Влияние суперэкспрессии гена *SCN9A* человека на генерацию, чувство, ответ на боль и анестезию при пароксизмальном остром болевом расстройстве, первичной эритромелалгии, нейропатии мелких волокон и врожденной нечувствительности боли при каналопатии согласно PubMed (Lu, 2011)

Признак	Пароксизмальное острое болевое расстройство (PEPD)	Нейропатия мелких волокон (SFN)	Первичная эритромелалгия (PE)	Врожденная нечувствительность боли, связанная с каналопатией (CIP)
Генерация боли	Согласно исчерпывающему обзору (Drenth, Waxman, 2007): при мутациях усиления функции *SCN9A* у пациентов с PEPD боль генерируется дольше или повторно в ответ на холод и растяжение	Согласно исчерпывающему обзору (Hoeijmakers et al., 2012): при мутациях усиления функции *SCN9A* у пациентов с SFN периферические аксоны малого диаметра генерируют боль и дегенерируют	Анализ родословной китайской семьи с наследственной PE (Wu et al., 2017) выявил замену F826Y усиления функции белка SCN9A с гипергенерацией боли и устойчивости к обезболивающим	Согласно клиническому случаю в университетском госпитале Сент-Жюстин (Монреаль, Канада), у 6-летней девочки, родившейся у здоровых канадцев французского происхождения без близкого родства, нашли мутацию I234T усиления функции *SCN9A*, первичным проявлением которой были PEPD и PE с многократными ежедневными приступами боли с эритемой конечностей и потливостью, а также со вторичной CIP между этими приступами из-за превышения порога числа поляризации потенциал-зависимых натриевых каналов при гипервозбудимости нейронов, как если бы у них «разрядилась батарейка» (Kim et al., 2015), тогда как все эти симптомы были успешно сняты анестезией против первичных PEPD и PE. Столь парадоксальная сочетанность, с одной стороны, вторичной CIP и, с другой – первичных для нее PEPD, PE и SFN может иметь крайне опасное клиническое проявление при миотонии как паралич мышц вслед за их гиперсокращением (Kim et al., 2015). Что касается CIP в качестве первичной патологии, то ее наблюдают клинически у пациентов с полной потерей функции ионных натриевых каналов, включая такой канал *SCN9A* (Dabby, 2012; Bennett, Woods, 2014; Shields et al., 2018)
Ощущение боли	Согласно исчерпывающему обзору (Dabby, 2012): одной из двух самых частых форм клинической манифестации мутаций усиления функции *SCN9A* является рост ощущения боли пациентами с PEPD	Согласно исчерпывающему обзору (Taub, Woolf, 2024): при мутациях усиления функции *SCN9A* пациенты с SFN ощущают жар, покалывание, жжение и аллодинию в конечностях, что диагностируют из года в год все чаще и чаще (Dabby, 2012)	Согласно исчерпывающему обзору (Dabby, 2012): при мутациях усиления функции *SCN9A* растет ощущение боли с PE в качестве одной из двух самых частых форм клинической манифестации таких мутаций	
Ответ на боль	Согласно исчерпывающему обзору (Stephenson, 2013): у младенцев с PEPD наблюдается тоническое сокращение всего тела с эритематозным изменением цвета, как у арлекина (пятна)	В биомедицинской модели SFN с использованием трансгенных рыб *Danio rerio* с искусственной мутацией I228M или G856D для избытка Scn9A (Eijkenboom et al., 2019): активность личинок растет с ростом температуры окружающей среды	Согласно обзору (Renthal, 2020): жара, рост температуры тела, физическая нагрузка, тесные одежда и обувь, острая и горячая пища провоцируют приступы жжения, жара и эритемы в ногах, руках и на лице у пациентов с PE тем чаще, чем выше его тяжесть	
Анестезия	Согласно обзору (Hisama et al., 2020): при PEPD самое эффективное лекарство – блокатор натриевых каналов карбамазепин	В рамках комплексного экспериментально-биоинформатического исследования SFN (Shao et al., 2016) сначала контекстный анализ миРНК-30b показал, что мРНК *SCN9A* может быть ее мишенью, а затем с использованием крыс подтвердили, что избыток миРНК-30b в поврежденном нерве снижает боль	На фармацевтической модели PE с использованием клеток HEK293A человека (Cregg et al., 2014): низкие дозы мексилетина в качестве блокатора натриевых каналов могут нормализовать генерацию, чувство и ответ на боль	

обращения 15.09.2024). В результате мы обнаружили ген *SCN9A* человека в качестве гена-интегратора всех этих биологических процессов (см. рис. 1). Его индекс филостратиграфического возраста, PAI = 4, по шкале KEGG (Kanehisa, Goto, 2000) оказался неотличим от PAI генов человека, связанных с любой из исследуемых комбинаций признаков боли, и соответствовал таксону Chordata, в пределах которого у позвоночных нервная система разделилась на центральную и периферическую (Holland L.Z., Holland N.D., 1999).

В качестве ключевого молекулярно-генетического процесса при генерации и реакции на боль и анестезию было выявлено также фосфорилирование (см. рис. 2). Этот результат согласуется, прежде всего, с экспериментальными данными биомедицинской модели боли с использованием клеток линии HEK293T человека (Kerth et al., 2021). Авторы упомянутого исследования обнаружили, что замена I → T в позиции 848 белка SCN9A человека создает дополнительный сайт фосфорилирования этого белка, что сопровождается повышением чувствительности и возбудимости нейронов за счет увеличения диапазона (потенциала) деполяризации его мембраны.

Кроме того, наш вывод о важности фосфорилирования ионных каналов для генерации и реакции на боль и анесте-

Таблица 5. Сравнение величин встречаемости кандидатных SNP-маркеров повышения и снижения сродства TBP к промоторам гена *SCN9A* человека, выявленных в настоящей работе (см. рис. 3 и табл. 3) с полногеномной оценкой этих величин согласно оценкам проекта «1000 геномов»

Референсный геном человека: сборка GRCh38/hg38 (Lowy-Gallego et al., 2019), база данных dbSNP, выпуск 155 (Day, 2010)		Количество объектов					Нейтральный дрейф (Haldane, 1957; Kimura, 1968)
		N_{GENE}	N_{SNP}	N_{Δ}	$N_{\downarrow}$	$N_{\uparrow}$	Гипотеза H_0: $N_{\downarrow} \geq 4N_{\uparrow}$ (Kasowski et al., 2010), биномиальное распределение, *p*
Оценки частот встречаемости SNP-маркеров значимого повышения и снижения сродства «TBP-промотор»	Полногеномная оценка для всех генов человека (1000 Genomes Project Consortium et al., 2012)	30 000	100 000	1000	800	200	>0.50
	Частная оценка только для гена *SCN9A* человека (эта работа)	1	37#	21*	5*	16*	$<10^{-6}$

Примечание. N_{GENE} – количество исследуемых генов; N_{SNP} – количество исследуемых SNP; N_{Δ} – количество SNP, способных к повышению ($N_{\downarrow}$) и к снижению ($N_{\uparrow}$) сродства TBP к промоторам генов человека. # См. рис. 3, *A*. * См. табл. 2.

зию согласуется (см. табл. 1) с литературными данными о важности фосфорилирования кальциевого канала клеток миокарда (Iyer et al., 2007), фосфорилирования натриевых каналов нейронов Пуркинье в мозжечке для координации движений (Grieco et al., 2002).

В развитие этих примеров мы нашли в базе данных PubMed (Lu, 2011) клеточную модель циркадного ритма с использованием фоторецепторов курицы, в рамках которой повышение фосфорилирования ионных каналов колбочек сетчатки в ответ на рост освещенности днем после темноты ночью было контрольным событием циркадного ритма у этого модельного животного (Chae et al., 2007). При изучении патологий зрения у крыс было установлено, что фосфорилирование ионных каналов зрительного нерва регулирует проводящие пути зрительной системы (Ogata et al., 2022). Кроме того, фосфорилирование калиевого канала слуховых нейронов лежит в основе идентификации слуховыми ядрами ствола мозга направления на источник звука за счет микросекундных задержек в регистрации сигналов от него (Song et al., 2005). Уровень фосфорилирования канала SNAP-25 в миндалевидном теле, коре головного мозга и гиппокампе повышался с ростом интенсивности холодового стресса в исследовании стресс-ответа у мышей (Yamamori et al., 2014). Вместе все это убедительно свидетельствует о ключевой роли фосфорилирования ионных каналов в специализации центральной и периферической нервной системы в целом, а также при генерации, чувстве, ответе на боль и анестезии.

На заключительном шаге с использованием базы данных Human_SNP_TATAdb (Filonov et al., 2023) был предложен 21 кандидатный SNP-маркер достоверного изменения экспрессии гена *SCN9A* человека, кодирующего регулируемый напряжением натриевый канал передачи сигналов о повреждении тканей от сенсорных нейронов в центральную нервную систему (см. табл. 3). В базе данных ClinVar (Landrum et al., 2014) были найдены описания клинических *in vivo* проявлений для 4 из 21 предсказанного SNP-маркера, которые гармонировали с нашими оценками *in silico* (см. табл. 4). Сравнение встречаемости выявленных SNP в промоторе *SCN9A* с полногеномной оценкой встречаемости таких SNP, полученной консорциумом «1000 геномов» в 2012 г., позволяет сделать вывод о действии естественного отбора против снижения экспрессии *SCN9A* (см. табл. 5), что свидетельствует об адаптивной роли генерации боли и ее ощущения, а также ответа на боль и обезболивания как анестезии (Raja et al., 2020).

Таким образом, представленные результаты соответствуют результатам независимых авторов, а в ряде случаев уточняют и обобщают их.

Заключение

Впервые проведено комплексное биоинформатическое исследование 568 генов человека, ассоциированных с болью и анестезией согласно базе данных NCBI Gene (Brown et al., 2015). Среди этих генов был обнаружен ген *SCN9A*, кодирующий субъединицу IXα ионного канала Nav1.7 передачи сигналов о повреждении тканей от периферической нервной системы в центральную, который единственный был вовлечен во все изучаемые процессы в качестве гена-интегратора. С помощью web-сервиса OrthoWeb (Mustafin et al., 2020) для *SCN9A* оценили индекс филостратиграфического возраста, PAI. Его значение (PAI = 4) соответствует таксону Chordata, когда произошла специализация центральной и периферической нервной систем (Holland L.Z., Holland N.D., 1999). Ассоциативную сеть гена *SCN9A* реконструировали с использованием инструмента ANDSystem (Ivanisenko et al., 2015), где фосфорилирование ионного канала *SCN9A* определяет эффективность передачи сигналов из периферической в центральную нервную систему в качестве краеугольного камня генерации, чувства, ответа на боль и обезболивания. Наконец, поиск в базе данных Human_SNP_TATAdb (Filonov et al., 2023) выявил 21 кандидатный SNP-маркер статистически значимого изменения экспрессии *SCN9A*. Отношение встречаемости аллелей этих SNP, которые повышают и понижают уровень SCN9A, сравнили с известным таковым отношением для всех генов человека (1000 Genomes Project Consortium et al., 2012). В итоге получили свидетельство того, что генерация, чувство боли, ответ на боль и восприимчивость анестезии (Raja et al., 2020) – адаптивные признаки, эффективность которых контролирует естественный отбор.

Список литературы / References

1000 Genomes Project Consortium; Abecasis G.R., Auton A., Brooks L.D., DePristo M.A., Durbin R.M., Handsaker R.E., Kang H.M., Marth G.T., McVean G.A. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. doi 10.1038/nature11632

Baker M.D., Nassar M.A. Painful and painless mutations of *SCN9A* and *SCN11A* voltage-gated sodium channels. *Pflugers Arch*. 2020; 472(7):865-880. doi 10.1007/s00424-020-02419-9

Bell E.A., Boehnke P., Harrison T.M., Mao W.L. Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc. Natl. Acad. Sci. USA*. 2015;112(47):14518-14521. doi 10.1073/pnas.1517557112

Bennett D.L., Woods C.G. Painful and painless channelopathies. *Lancet Neurol*. 2014;13(6):587-599. doi 10.1016/S1474-4422(14)70024-9

Brown G.R., Hem V., Katz K.S., Ovetsky M., Wallin C., Ermolaeva O., Tolstoy I., Tatusova T., Pruitt K.D., Maglott D.R., Murphy T.D. Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res*. 2015;43(D1):D36-D42. doi 10.1093/nar/gku1055

Chae K.S., Ko G.Y., Dryer S.E. Tyrosine phosphorylation of cGMP-gated ion channels is under circadian control in chick retina photoreceptors. *Invest. Ophthalmol. Vis. Sci*. 2007;48(2):901-906. doi 10.1167/iovs.06-0824

Chatterjee H.J., Ho S.Y., Barnes I., Groves C. Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol. Biol*. 2009;9:259. doi 10.1186/1471-2148-9-259

Cregg R., Cox J.J., Bennett D.L., Wood J.N., Werdehausen R. Mexiletine as a treatment for primary erythromelalgia: normalization of biophysical properties of mutant L858F $Na_V1.7$ sodium channels. *Br. J. Pharmacol*. 2014;171(19):4455-4463. doi 10.1111/bph.12788

Cui J.H., Kim W.M., Lee H.G., Kim Y.O., Kim C.M., Yoon M.H. Antinociceptive effect of intrathecal cannabinoid receptor agonist WIN 55,212-2 in a rat bone tumor pain model. *Neurosci. Lett*. 2011; 493(3):67-71. doi 10.1016/j.neulet.2010.12.052

Dabby R. Pain disorders and erythromelalgia caused by voltage-gated sodium channel mutations. *Curr. Neurol. Neurosci. Rep.* 2012;12(1): 76-83. doi 10.1007/s11910-011-0233-8

Datta P.M. Earliest mammal with transversely expanded upper molar from the Late Triassic (Carnian) Tiki formation, South Rewa Gondwana Basin, India. *J. Vertebr. Paleontol*. 2005;25(1):200-207. doi 10.1671/0272-4634(2005)025[0200:EMWTEU]2.0.CO;2

Day I.N. dbSNP in the detail and copy number complexities. *Hum. Mutat*. 2010;31(1):2-4. doi 10.1002/humu.21149

Dib-Hajj S.D., Cummins T.R., Black J.A., Waxman S.G. From genes to pain: $Na_v1.7$ and human pain disorders. *Trends Neurosci*. 2007; 30(11):555-463. doi 10.1016/j.tins.2007.08.004

Diogo R. The Origin of Higher Clades: Osteology, Myology, Phylogeny and Evolution of Bony Fishes and the Rise of Tetrapods. NY: CRC Press, 2007. doi 10.1201/9780429063978

Drenth J.P., Waxman S.G. Mutations in sodium-channel gene *SCN9A* cause a spectrum of human genetic pain disorders. *J. Clin. Invest*. 2007;117(12):3603-3609. doi 10.1172/JCI33297

Dunn R.H., Rose K.D., Rana R.S., Kumar K., Sahni A., Smith T. New euprimate postcrania from the early Eocene of Gujarat, India, and the strepsirrhine-haplorhine divergence. *J. Hum. Evol*. 2016;99: 25-51. doi 10.1016/j.jhevol.2016.06.006

Eijkenboom I., Sopacua M., Otten A.B.C., Gerrits M.M., Hoeijmakers J.G.J., Waxman S.G., Lombardi R., Lauria G., Merkies I.S.J., Smeets H.J.M., Faber C.G., Vanoevelen J.M.; PROPANE Study Group. Expression of pathogenic *SCN9A* mutations in the zebrafish: a model to study small-fiber neuropathy. *Exp. Neurol*. 2019;311: 257-264. doi 10.1016/j.expneurol.2018.10.008

Filonov S.V., Podkolodnyy N.L., Podkolodnaya O.A., Tverdokhleb N.N., Ponomarenko P.M., Rasskazov D.A., Bogomolov A.G., Ponomarenko M.P. Human_SNP_TATAdb: a database of SNPs that statistically significantly change the affinity of the TATA-binding protein to human gene promoters: genome-wide analysis and use cases. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2023;27(7):728-736. doi 10.18699/VJGB-23-85

Garate J., Maimo-Barcelo A., Bestard-Escalas J., Fernandez R., Perez-Romero K., Martinez M.A., Payeras M.A., Lopez D.H., Fernandez J.A., Barcelo-Coblijn G. A drastic shift in lipid adducts in colon cancer detected by MALDI-IMS exposes alterations in specific K+ channels. *Cancers* (*Basel*). 2021;13(6):1350. doi 10.3390/cancers13061350

Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res*. 2015;43(D1):D1049-1056. doi 10.1093/nar/gku1179

Grieco T.M., Afshari F.S., Raman I.M. A role for phosphorylation in the maintenance of resurgent sodium current in cerebellar purkinje neurons. *J. Neurosci*. 2002;22(8):3100-3107. doi 10.1523/jneurosci.22-08-03100.2002

Haldane J.B.S. The cost of natural selection. *J. Genet*. 1957;55:511-524. doi 10.1007/bf02984069

Harrison T. Catarrhine origins. In: A Companion to Paleoanthropology. NY: Blackwell Publ. Ltd., 2013;376-396

Hey J. The ancestor's tale A pilgrimage to the dawn of evolution. *J. Clin. Invest*. 2005;115(7):1680. doi 10.1172/JCI25761

Hisama F.M., Dib-Hajj S.D., Waxman S.G. SCN9A Neuropathic pain syndromes. 2006 May 6 [Updated 2020 Jan 23]. In: GeneReviews®. [Internet]. Seattle (WA): Univ. of Washington, Seattle; 1993-2024. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1163/

Hoeijmakers J.G., Merkies I.S., Gerrits M.M., Waxman S.G., Faber C.G. Genetic aspects of sodium channelopathy in small fiber neuropathy. *Clin. Genet*. 2012;82(4):351-358. doi 10.1111/j.1399-0004.2012.01937.x

Holland L.Z., Holland N.D. Chordate origins of the vertebrate central nervous system. *Curr. Opin. Neurobiol*. 1999;9(5):596-602. doi 10.1016/S0959-4388(99)00003-3

Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol*. 2015;9(Suppl.2):S2. doi 10.1186/1752-0509-9-S2-S2

Iyer V., Edelman E.R., Lilly L.S. Basic cardiac structure and function. In: Pathophysiology of Heart Disease. Baltimore: Lippincott Williams & Wilkins, 2007;1-28

Kanehisa M., Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi 10.1093/nar/28.1.27

Kasowski M., Grubert F., Heffelfinger C., Hariharan M., Asabere A., Waszak S.M., Habegger L., Rozowsky J., Shi M., Urban A.E., Hong M.Y., Karczewski K.J., Huber W., Weissman S.M., Gerstein M.B., Korbel J.O., Snyder M. Variation in transcription factor binding among humans. *Science*. 2010;328(5975):232-235. doi 10.1126/science.1183621

Kerth C.M., Hautvast P., Korner J., Lampert A., Meents J.E. Phosphorylation of a chronic pain mutation in the voltage-gated sodium channel Nav1.7 increases voltage sensitivity. *J. Biol. Chem*. 2021; 296:100227. doi 10.1074/jbc.RA120.014288

Kim D.T., Rossignol E., Najem K., Ospina L.H. Bilateral congenital corneal anesthesia in a patient with *SCN9A* mutation, confirmed primary erythromelalgia, and paroxysmal extreme pain disorder. *J. AAPOS*. 2015;19(5):478-479. doi 10.1016/j.jaapos.2015.05.015

Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968; 217(5129):624-626. doi 10.1038/217624a0

Klein A.A., Meek T., Allcock E., Cook T.M., Mincher N., Morris C., Nimmo A.F., Pandit J.J., Pawa A., Rodney G., Sheraton T., Young P. Recommendations for standards of monitoring during anaesthesia and recovery 2021: Guideline from the Association of Anaesthetists. *Anaesthesia*. 2021;76(9):1212-1223. doi 10.1111/anae.15501

Kumar V., Hallstrom B.M., Janke A. Coalescent-based genome analyses resolve the early branches of the euarchontoglires. *PLoS One*. 2013;8(4):e60019. doi 10.1371/journal.pone.0060019

Kwak S.G., Kim J.H. Central limit theorem: the cornerstone of modern statistics. *Korean J. Anesthesiol*. 2017;70(2):144-156. doi 10.4097/kjae.2017.70.2.144

Landrum M.J., Lee J.M., Riley G.R., Jang W., Rubinstein W.S., Church D.M., Maglott D.R. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(D1):D980-D985. doi 10.1093/nar/gkt1113

Leander B.S. Predatory protists. *Curr. Biol*. 2020;30(10):R510-R516. doi 10.1016/j.cub.2020.03.052

Lowy-Gallego E., Fairley S., Zheng-Bradley X., Ruffier M., Clarke L., Flicek P.; 1000 Genomes Project Consortium. Variant calling on the GRCh38 assembly with the data from phase three of the 1000 Genomes Project. *Wellcome Open Res*. 2019;4:50. doi 10.12688/wellcomeopenres.15126.2

Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database* (*Oxford*). 2011;2011:baq036. doi 10.1093/database/baq036

Lucas D.N., Russell R., Bamber J.H., Elton C.D. Recommendations for standards of monitoring during anaesthesia and recovery 2021. *Anaesthesia*. 2021;76(10):1426-1427. doi 10.1111/anae.15528

Luo Z.X., Yuan C.X., Meng Q.J., Ji Q. A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature*. 2011; 476(7361):442-445. doi 10.1038/nature10291

Maloof A.C., Porter S.M., Moore J.L., Dudas F.O., Bowring S.A., Higgins J.A., Fike D.A., Eddy M.P. The earliest Cambrian record of animals and ocean geochemical change. *Geol. Soc. Am. Bull*. 2010a;122(11-12):1731-1774. doi 10.1130/B30346.1

Maloof A.C., Rose C.V., Beach R., Samuels B.M., Calmet C.C., Erwin D.H., Poirier G.R., Yao N., Simons F.J. Possible animal-body fossils in pre-Marinoan limestones from South Australia. *Nat. Geosci*. 2010b;3:653-659. doi 10.1038/ngeo934

Mi H., Ebert D., Muruganujan A., Mills C., Albou L.P., Mushayamaha T., Thomas P.D. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res*. 2021;49(D1):D394-D403. doi 10.1093/nar/gkaa1106

Morozova O.V., Alekseeva A.E., Sashina T.A., Brusnigina N.F., Epifanova N.V., Kashnikov A.U., Zverev V.V., Novikova N.A. Phylodynamics of G4P[8] and G2P[4] strains of rotavirus A isolated in Russia in 2017 based on full-genome analyses. *Virus Genes*. 2020; 56(5):537-545. doi 10.1007/s11262-020-01771-3

Mustafin Z., Mukhin A., Afonnikov D., Matushkin Y., Lashin S. OrthoWeb – web application for macro- and microevolutionary analysis of genes. In: Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2020). Novosibirsk, 2020; 228-229. doi 10.18699/BGRS/SB-2020-145

Ogata G., Partida G.J., Fasoli A., Ishida A.T. Calcium/calmodulin-dependent protein kinase II associates with the K^+ channel isoform Kv4.3 in adult rat optic nerve. *Front. Neuroanat*. 2022;16:958986. doi 10.3389/fnana.2022.958986

Palomes-Borrajo G., Badia J., Navarro X., Penas C. Nerve excitability and neuropathic pain is reduced by BET protein inhibition after spared nerve injury. *J. Pain*. 2021;22(12):1617-1630. doi 10.1016/j.jpain.2021.05.005

Ponomarenko M., Rasskazov D., Arkova O., Ponomarenko P., Suslov V., Savinkova L., Kolchanov N. How to use SNP_TATA_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene's promoter via a change in affinity of the TATA-binding protein for this promoter. *Biomed. Res. Int*. 2015; 2015:359835. doi 10.1155/2015/359835

Raja S.N., Carr D.B., Cohen M., Finnerup N.B., Flor H., Gibson S., Keefe F.J., Mogil J.S., Ringkamp M., Sluka K.A., Song X.J., Stevens B., Sullivan M.D., Tutelman P.R., Ushida T., Vader K. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. *Pain*. 2020;161(9): 1976-1982. doi 10.1097/j.pain.0000000000001939

Raney B.J., Barber G.P., Benet-Pagès A., Casper J., Clawson H., Cline M.S., Diekhans M., Fischer C., Navarro Gonzalez J., Hickey G., Hinrichs A.S., Kuhn R.M., Lee B.T., Lee C.M., Le Mercier P., Miga K.H., Nassar L.R., Nejad P., Paten B., Perez G., Schmelter D., Speir M.L., Wick B.D., Zweig A.S., Haussler D., Kent W.J., Haeussler M. The UCSC Genome Browser database: 2024 update. *Nucleic Acids Res*. 2024;52(D1):D1082-D1088. doi 10.1093/nar/gkad987

Renthal W. Pain genetics. In: Rosenberg R.N., Pascual J.M. (Eds) Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease. Amsterdam: Elsevier, 2020;397-410. doi 10.1016/b978-0-12-813866-3.00023-0

Samet H. A top-down quadtree traversal algorithm. *IEEE Trans. Pattern Anal. Mach. Intell*. 1985;7(1):94-98. doi 10.1109/tpami.1985.4767622

Scerri E.M.L., Thomas M.G., Manica A., Gunz P., Stock J.T., Stringer C., Grove M., Groucutt H.S., Timmermann A., Rightmire G.P., d'Errico F., Tryon C.A., Drake N.A., Brooks A.S., Dennell R.W., Durbin R., Henn B.M., Lee-Thorp J., deMenocal P., Petraglia M.D., Thompson J.C., Scally A., Chikhi L. Did our species evolve in subdivided populations across Africa, and why does it matter? *Trends Ecol. Evol*. 2018;33(8):582-594. doi 10.1016/j.tree.2018.05.005

Schrenk F., Kullmer O., Bromage T. The earliest putative homo fossils. In: Henke W., Tattersall I. (Eds) Handbook of Paleoanthropology. Berlin: Springer, 2014;1-19. doi 10.1007/978-3-540-33761-4_52

Shao J., Cao J., Wang J., Ren X., Su S., Li M., Li Z., Zhao Q., Zang W. MicroRNA-30b regulates expression of the sodium channel Nav1.7 in nerve injury-induced neuropathic pain in the rat. *Mol. Pain*. 2016; 12:1744806916671523. doi 10.1177/1744806916671523

Sherman B.T., Hao M., Qiu J., Jiao X., Baseler M.W., Lane H.C., Imamichi T., Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*. 2022;50(W1):W216-W221. doi 10.1093/nar/gkac194

Shields S.D., Deng L., Reese R.M., Dourado M., Tao J., Foreman O., Chang J.H., Hackos D.H. Insensitivity to pain upon adult-onset deletion of Nav1.7 or its blockade with selective inhibitors. *J. Neurosci*. 2018;38(47):10180-10201. doi 10.1523/jneurosci.1049-18.2018

Shu D.-G., Luo H.-L., Conway-Morris S., Zhang X.-L., Hu S.-X., Chen L., Han J., Zhu M., Li Y., Chen L.-Z. Lower Cambrian vertebrates from south China. *Nature*. 1999;402(6757):42-46. doi 10.1038/46965

Shuyuan L., Haoyu C. Mechanism of Nardostachyos Radix et Rhizoma–Salidroside in the treatment of premature ventricular beats based on network pharmacology and molecular docking. *Sci. Rep*. 2023;13(1):20741. doi 10.1038/s41598-023-48277-0

Sokolov M.V., Henrich-Noack P., Raynoschek C., Franzen B., Larsson O., Main M., Dabrowski M. Co-expression of β subunits with the voltage-gated sodium channel $Na_V1.7$: the importance of subunit association and phosphorylation and their effects on channel pharmacology and biophysics. *J. Mol. Neurosci*. 2018;65(2):154-166. doi 10.1007/s12031-018-1082-6

Song P., Yang Y., Barnes-Davies M., Bhattacharjee A., Hamann M., Forsythe I.D., Oliver D.L., Kaczmarek L.K. Acoustic environment determines phosphorylation state of the Kv3.1 potassium channel in auditory neurons. *Nat. Neurosci*. 2005;8(10):1335-1342. doi 10.1038/nn1533

Stajich J.E., Block D., Boulez K., Brenner S.E., Chervitz S.A., Dagdigian C., Fuellen G., Gilbert J.G., Korf I., Lapp H., Lehvaslaiho H., Matsalla C., Mungall C.J., Osborne B.I., Pocock M.R., Schattner P., Senger M., Stein L.D., Stupka E., Wilkinson M.D., Birney E. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res*. 2002;12(10):1611-1618. doi 10.1101/gr.361602

Stamboulian S., Choi J.S., Ahn H.S., Chang Y.W., Tyrrell L., Black J.A., Waxman S.G., Dib-Hajj S.D. ERK1/2 mitogen-activated protein kinase phosphorylates sodium channel $Na_v1.7$ and alters its gating properties. *J. Neurosci*. 2010;30(5):1637-1647. doi 10.1523/jneurosci.4872-09.2010

Stephenson J.B. Syncopes and other paroxysmal events. *Handb. Clin. Neurol*. 2013;112:861-866. doi 10.1016/B978-0-444-52910-7.00007-6

Sun G.L., Shen W., Wen J.F. Triosephosphate isomerase genes in two trophic modes of euglenoids (euglenophyceae) and their phylogenetic analysis. *J. Eukaryot. Microbiol*. 2008;55(3):170-177. doi 10.1111/j.1550-7408.2008.00324.x

Szklarczyk D., Kirsch R., Koutrouli M., Nastou K., Mehryary F., Hachilif R., Gable A.L., Fang T., Doncheva N.T., Pyysalo S., Bork P., Jensen L.J., von Mering C. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638-D646. doi 10.1093/nar/gkac1000

Taub D.G., Woolf C.J. Age-dependent small fiber neuropathy: mechanistic insights from animal models. *Exp. Neurol*. 2024;377:114811. doi 10.1016/j.expneurol.2024.114811

Warde-Farley D., Donaldson S.L., Comes O., Zuberi K., Badrawi R., Chao P., Franz M., Grouios C., Kazi F., Lopes C.T., Maitland A., Mostafavi S., Montojo J., Shao Q., Wright G., Bader G.D., Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010;38(W2):W214-W220. doi 10.1093/nar/gkq537

Wu B., Zhang Y., Tang H., Yang M., Long H., Shi G., Tang J., Shi X. A novel SCN9A mutation (F826Y) in primary erythromelalgia alters the excitability of Nav1.7. *Curr. Mol. Med*. 2017;17(6):450-457. doi 10.2174/1566524017666171009105029

Xue Y., Kremer M., Muniz Moreno M.D.M., Chidiac C., Lorentz R., Birling M.C., Barrot M., Herault Y., Gaveriaux-Ruff C. The human *SCN9A*^{R185H} point mutation induces pain hypersensitivity and spontaneous pain in mice. *Front. Mol. Neurosci*. 2022;15:913990. doi 10.3389/fnmol.2022.913990

Yamamori S., Sugaya D., Iida Y., Kokubo H., Itakura M., Suzuki E., Kataoka M., Miyaoka H., Takahashi M. Stress-induced phosphorylation of SNAP-25. *Neurosci. Lett*. 2014;561:182-187. doi 10.1016/j.neulet.2013.12.044

Zerbino D.R., Wilder S.P., Johnson N., Juettemann T., Flicek P.R. The ensemble regulatory build. *Genome Biol*. 2015;16(1):56. doi 10.1186/s13059-015-0621-5

Zhou Y., Zhou B., Pache L., Chang M., Khodabakhshi A.H., Tanaseichuk O., Benner C., Chanda S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun*. 2019;10(1):1523. doi 10.1038/s41467-019-09234-6

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 23.09.2024. После доработки 13.11.2024. Принята к публикации 13.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):822-833

doi 10.18699/vjgb-24-90

Программный комплекс MetArea для анализа взаимоисключающей встречаемости в парах мотивов сайтов связывания транскрипционных факторов по данным ChIP-seq

В.Г. Левицкий [1, 2], А.В. Цуканов [1], Т.И. Меркулова [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

levitsky@bionet.nsc.ru

Аннотация. Технология ChIP-seq, основанная на иммунопреципитации хроматина (ChIP), позволяет картировать набор геномных локусов (пиков), содержащих сайты связывания (СС) для исследуемого (целевого) транскрипционного фактора (ТФ). ТФ может распознавать несколько структурно различных мотивов СС. Мультибелковый комплекс, картируемый в эксперименте ChIP-seq, включает целевой и другие «партнерские» ТФ, связанные белок-белковыми взаимодействиями. Не все из этих ТФ связываются с ДНК напрямую. Поэтому и целевой, и партнерские ТФ распознают обогащенные мотивы СС в пиках. Для поиска обогащенных мотивов по данным ChIP-seq применяется подход *de novo* поиска. Для пары обогащенных мотивов СС ТФ в наборе пиков может быть обнаружена совместная или взаимоисключающая встречаемость: совместная отражает более частое нахождение двух мотивов СС ТФ в одних пиках, а взаимоисключающая – в разных пиках. Мы предлагаем программный комплекс (ПК) MetArea для выявления пар мотивов СС ТФ со взаимоисключающей встречаемостью по данным ChIP-seq. ПК MetArea предназначен для предсказания структурного разнообразия мотивов СС одного ТФ и функциональной связи мотивов СС разных ТФ. Функциональная связь мотивов двух разных ТФ предполагает, что они взаимозаменяемы в составе мультибелкового комплекса, который использует СС этих ТФ для прямого связывания с ДНК в различных пиках. ПК MetArea рассчитывает оценки точности распознавания pAUPRC (частичная площадь под кривой Precision–Recall) для каждого из двух входных одиночных мотивов, определяет их «объединенный» мотив и оценивает точность для него. Целью анализа является поиск пар одиночных мотивов A и B, для которых точность объединенного мотива A&B выше точностей обоих одиночных мотивов.

Ключевые слова: *de novo* поиск мотивов; кривая PR; площадь под кривой; структурные варианты мотивов сайтов связывания транскрипционных факторов; кооперативное действие транскрипционных факторов.

Для цитирования: Левицкий В.Г., Цуканов А.В., Меркулова Т.И. Программный комплекс MetArea для анализа взаимоисключающей встречаемости пар мотивов сайтов связывания транскрипционных факторов по данным ChIP-seq. *Вавиловский журнал генетики и селекции*. 2024;28(8):822-833. doi 10.18699/vjgb-24-90

Финансирование. Работа поддержана государственным бюджетным проектом № FWNR-2022-0020 Института цитологии и генетики СО РАН.

Благодарности. Разработка программного пакета и анализ данных проведены с использованием вычислительных ресурсов ЦКП «Биоинформатика» (при поддержке бюджетного проекта № FWNR-2022-0020).

MetArea: a software package for analysis of the mutually exclusive occurrence in pairs of motifs of transcription factor binding sites based on ChIP-seq data

V.G. Levitsky [1, 2], A.V. Tsukanov [1], T.I. Merkulova [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia

levitsky@bionet.nsc.ru

Abstract. ChIP-seq technology, which is based on chromatin immunoprecipitation (ChIP), allows mapping a set of genomic loci (peaks) containing binding sites (BS) for the investigated (target) transcription factor (TF). A TF may recognize several structurally different BS motifs. The multiprotein complex mapped in a ChIP-seq experiment includes target and other "partner" TFs linked by protein-protein interactions. Not all these TFs bind to DNA directly. Therefore, both target and partner TFs recognize enriched BS motifs in peaks. A *de novo* search approach is used to search for enriched TF BS motifs in ChIP-seq data. For a pair of enriched BS motifs of TFs, the co-occurrence or mutually exclusive occurrence can be detected from a set of peaks: the co-occurrence reflects a more frequent occurrence of two motifs in the same peaks, while the mutually exclusive means their more frequent detection in different peaks. We propose

© Левицкий В.Г., Цуканов А.В., Меркулова Т.И., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

the MetArea software package to identify pairs of TF BS motifs with the mutually exclusive occurrence in ChIP-seq data. MetArea was designed to predict the structural diversity of BS motifs of the same TFs, and the functional relation of BS motifs of different TFs. The functional relation of the motifs of the two distinct TFs presumes that they are interchangeable as part of a multiprotein complex that uses the BS of these TFs to bind directly to DNA in different peaks. MetArea calculates the estimates of recognition performance pAUPRC (partial area under the Precision–Recall curve) for each of the two input single motifs, identifies the "joint" motif, and computes the performance for it too. The goal of the analysis is to find pairs of single motifs A and B for which the accuracy of the joint A&B motif is higher than those of both single motifs.

Key words: *de novo* motif search; PR curve; area under curve; structural variants of transcription factor binding site motifs; cooperative action of transcription factors.

For citation: Levitsky V.G., Tsukanov A.V., Merkulova T.I. MetArea: a software package for analysis of the mutually exclusive occurrence in pairs of motifs of transcription factor binding sites based on ChIP-seq data. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):822-833. doi 10.18699/vjgb-24-90

Введение

Транскрипционные факторы (ТФ) – это белки, обладающие способностью специфического связывания ДНК и регулирующие таким образом транскрипцию генов. К ТФ относится около 1600 белков человека (Lambert et al., 2018). Сайты связывания (СС) ТФ в геномной ДНК эукариот представляют собой короткие участки длиной обычно от 6 до 20 пар нуклеотидов (п. н.) (Vorontsov et al., 2024). Как правило, ТФ способны связываться не с одной последовательностью ДНК, а со многими сходными. Мотив СС ТФ в ДНК – это общее представление доступного разнообразия таких сходных последовательностей (D'haeseleer, 2006). Установить четко закономерности, определяющие аффинность нуклеотидных последовательностей геномной ДНК к ТФ, очень трудно: даже умеренно консервативными в мотивах СС ТФ, т. е. неизменными в большинстве природных СС, является всего лишь несколько позиций нуклеотидов, число их обычно гораздо меньше половины длины мотива. Разнообразие мотивов СС ТФ *in vivo* пока слабо изучено из-за огромного разнообразия механизмов связывания ТФ с ДНК, включающего, наряду с прямым взаимодействием ТФ с ДНК, механизмы связывания при помощи или через посредника в составе мультибелкового комплекса с другими ТФ, использование пространственной структуры ДНК в составе нуклеосомы и т. п. (Morgunova, Taipale, 2017; Levitsky et al., 2020; Zeitlinger, 2020).

Самой популярной моделью мотивов СС ТФ является традиционная позиционная весовая матрица (ПВМ) (Wasserman, Sandelin, 2004; Tognon et al., 2023). Она оценивает аффинность сайта как сумму вкладов (весов) всех его позиций, где вес каждой позиции задан типом ее нуклеотида. Альтернативные модели мотива способны дополнить предсказания модели ПВМ (Levitsky et al., 2007; Siebert, Söding, 2016; Tsukanov et al., 2022), т. е. предсказывать СС ТФ в таких локусах генома, где модель ПВМ этого не делает. Общим отличием всех альтернативных моделей мотива от традиционной модели ПВМ является участие в оценке аффинности сайтов вкладов зависимостей частот нуклеотидов в разных позициях мотива.

Способность каждого ТФ взаимодействовать с ДНК обеспечивается его ДНК-связывающим доменом (ДСД). Структура ДСД ТФ определяет варианты мотивов его СС (Wingender, 2013; Lambert et al., 2018; Nagy G., Nagy L., 2020). Иерархическая классификация ТФ по структуре ДСД в базе данных TFClass (Wingender, 2013; Wingender et al., 2013, 2015, 2018) определяет классы ТФ по структуре ДСД. Например, в базе данных Hocomoco (Vorontsov et al., 2024) аннотированы мотивы СС 949 разных ТФ человека. Эти ТФ относятся к 34 классам, однако на 10 классов, имеющих не менее 10 ТФ, приходится 858 ТФ (более 90 % от всех 949 ТФ), а три крупнейших класса – C2H2 zinc finger factors {2.3}, Homeo domain factors {3.1} и Basic helix-loop-helix factors (bHLH) {1.2} – насчитывают 373, 184 и 76 ТФ соответственно. Выравнивание последовательностей ДСД ТФ определяет семейства и подсемейства ТФ ниже классов по иерархии.

Транскрипционные факторы эукариот взаимодействуют с ДНК *in vivo* в составе мультибелковых комплексов, включающих несколько ТФ. ТФ в составе таких комплексов можно назвать «партнерскими», так как между ними есть белок-белковые взаимодействия. Общее (кооперативное) действие нескольких ТФ на регуляторный район гена способно менять локальное окружение хроматина и регулировать транскрипцию гена (Morgunova, Taipale, 2017; Zeitlinger, 2020; Georgakopoulos-Soares et al., 2023). Для ТФ многих классов характерна возможность связываться с совершенно структурно разными СС (Rogers et al., 2019; Vorontsov et al., 2024). Например, ТФ класса Nuclear receptors with C4 zinc fingers {2.1} (Ядерные рецепторы с цинковыми пальцами С4) могут связываться и как мономеры, и как димеры. В случае димера СС включает два полусайта, спейсер между которыми и ориентация цепей ДНК могут варьировать. ТФ класса Basic leucine zipper factors (bZIP) {1.1} (Лейциновая застежка) связываются только как димеры, так что два полусайта всегда расположены в одной цепи ДНК, а спейсер почти не изменяется (Nagy G., Nagy L., 2020). Здесь и далее в фигурных скобках индексы обозначены согласно данным базы TFClass (Wingender et al., 2013, 2015, 2018). Есть несколько типов ДСД ТФ эукариот, которые могут функционировать как димеры, включающие пары близкородственных ТФ (Amoutzias et al., 2008). ТФ, похожие по структуре ДСД, часто распознают схожие мотивы СС ТФ (Lambert et al., 2018; Ambrosini et al., 2020). Единственное явное исключение из этого правила – мотивы СС ТФ класса C2H2 цинковых пальцев (C2H2 zinc finger factors {2.3}).

Идентификация СС ТФ в геномах значительно продвинулась с появлением методов высокопроизводительного массового секвенирования, в частности экспериментальной технологии ChIP-seq. Эта технология дает для целевого ТФ набор районов генома (пиков) длиной обычно

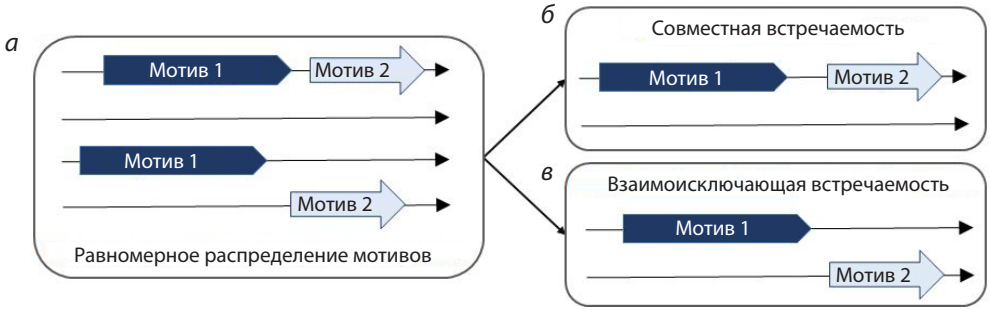


Рис. 1. Схема отличий терминов совместной и взаимоисключающей встречаемости мотивов СС ТФ.

Пусть частота встреч каждого из двух мотивов в пике 50 %. *а* – два мотива появляются в пиках независимо друг от друга, есть четыре равновероятных варианта картирования мотивов в пиках; *б* – совместная встречаемость: в пике есть только оба мотива или нет ни одного из двух; *в* – взаимоисключающая встречаемость: в пике есть только один мотив из двух. Стрелки от панели *а* к панелям *б* и *в* означают, что четыре варианта панели *а* точно разделяются на две группы по два варианта на панелях *б* и *в*.

несколько сотен пар нуклеотидов, где связывание мультибелкового комплекса многих ТФ, который включает и целевой ТФ, было экспериментально картировано. Поэтому два типа пиков отвечают за прямое и непрямое связывание целевого ТФ с геномной ДНК. Прямое связывание означает, что целевой ТФ взаимодействует с ДНК напрямую, а непрямое – что у целевого ТФ есть только белок-белковые взаимодействия с одним или несколькими партнерскими ТФ, которые, в свою очередь, взаимодействуют с ДНК напрямую. Из наличия прямого/непрямого связывания следует, что в пиках обогащены мотивы СС целевого/партнерских ТФ и что мотивы целевого ТФ есть только в части пиков. Для отражения роста содержания мотивов СС ТФ в геномных локусах, полученных по данным массового секвенирования ChIP-seq, применяется термин «обогащение» – повышенное содержание мотива СС ТФ по сравнению с ожидаемым по случайным причинам. Выборку последовательностей ДНК, по которой оценивается ожидаемое содержание мотива, называют негативной. Нами показано, что для пиков ChIP-seq более эффективно в негативную выборку брать любые участки генома, подходящие пикам по G/C-составу, чем использовать синтетические последовательности, полученные из пиков путем перемешивания нуклеотидов (Raditsa et al., 2024).

Когда по заданному набору данных массового секвенирования ChIP-seq определены обогащенные мотивы СС, на механизмы действия ТФ могут указать статистические закономерности встреч мотивов в парах. Ранее были определены понятия синергии и антагонизма мотивов в составе композиционного элемента (КЭ) – устойчивого сочетания для пары мотивов (Kel et al., 1995). Синергия означает, что результат действия пары ТФ значительно превосходит результат действия каждого из них в отдельности. При антагонизме ТФ, наоборот, мешают друг другу. Например, один из двух ТФ – активатор, а другой – репрессор, так что один из них вытесняет другого. К сожалению, понятия «синергия» и «антагонизм» относятся к устойчивой паре двух встречаемых в ДНК мотивов, и эти два случая невозможно различить по частотам совместных встреч пары мотивов.

С начала эпохи массового секвенирования СС ТФ прошло более 15 лет (Jonhson et al., 2007). Сегодня роль биоинформатического анализа полногеномных данных для понимания механизмов работы ТФ трудно переоценить. В случае данных ChIP-seq бионформатический анализ имеет дело не с отдельными локусами генома, а с набором сотен или даже тысяч таких локусов, в которых в целом наблюдается как прямое, так и непрямое связывание целевого ТФ. При переходе от отдельного рассмотрения частот двух мотивов СС ТФ в наборе пиков ChIP-seq к закономерностям связи между этими двумя мотивами целесообразно учитывать две возможности:

- два мотива чаще, чем ожидается по случайным причинам, встречаются вместе в одних пиках и реже по отдельности в разных пиках;
- два мотива встречаются чаще в разных пиках и реже в одних пиках.

Поэтому мы предлагаем для пары мотивов СС ТФ термины совместной и взаимоисключающей встречаемости (рис. 1).

Совместная встречаемость в паре мотивов отражает наличие КЭ, т. е. пары близко расположенных мотивов СС ТФ в ДНК, между которыми есть небольшой спейсер или которые перекрываются (Kel et al., 1995; Levitsky et al., 2019). Взаимоисключающая встречаемость в паре может означать, что либо эта пара представляет собой два структурных варианта СС одного ТФ (он связывается в разных пиках по-разному), либо это СС разных ТФ. Полагая, что два мотива СС соответствуют двум разным ТФ в составе одного мультибелкового комплекса, можно предположить замену одного ТФ, напрямую взаимодействующего с ДНК, на другой ТФ. Поэтому тренд расхождения мотивов СС двух ТФ по разным пикам может указывать на функциональную связь этих мотивов, в простейшем случае представляющую указанную выше замену. При совместной встречаемости, в случае как синергии, так и антагонизма, два ТФ связываются с ДНК вблизи друг от друга (по крайней мере некоторое время они могут быть в контакте даже при антагонизме); скорее всего, они входят в состав одного мультибелкового комплекса. При взаимоисключающей встречаемости, наоборот, мотивы

СС и соответствующие ТФ находятся в разных участках ДНК (разных пиках). Следовательно, мы предполагаем, что два мотива представляют альтернативные следы одной общей молекулярной функции ТФ:

- один ТФ распознаёт два различных по структуре мотива СС, или
- в составе мультибелкового комплекса связывание с ДНК происходит за счет разных ТФ и мотивов их СС.

Обе эти возможности иллюстрирует рис. 2.

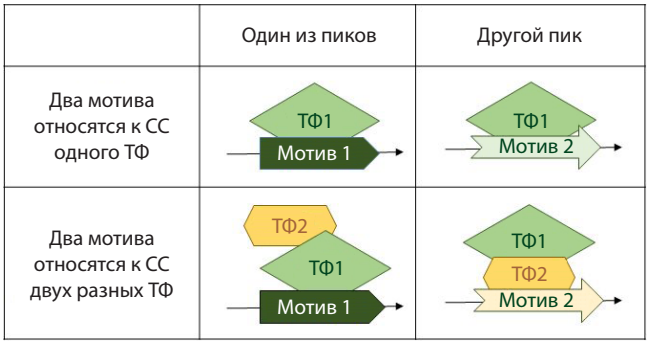


Рис. 2. Предполагаемое происхождение взаимоисключающей встречаемости двух мотивов СС ТФ в наборе пиков ChIP-seq.

Две колонки представляют два разных пика. Взаимоисключающая встречаемость в паре мотивов может означать, что либо пара мотивов объясняется двумя структурно разными мотивами одного ТФ (этот ТФ связывается с двумя мотивами в разных пиках), либо пара мотивов соответствует СС разных ТФ. В этом случае мы предполагаем, что в составе некоторого мультибелкового комплекса происходит замена взаимодействующего напрямую с ДНК одного ТФ на другой (ТФ1 на ТФ2).

Площадь под кривой ROC (AUC ROC, Area Under Curve) – это традиционная количественная оценка точности бинарного классификатора. Термин ROC (Receiver Operating Characteristic curve) означает кривая «Рабочая характеристика приемника». Для мотива СС ТФ кривая ROC определяется как зависимость доли распознанных последовательностей позитивной выборки (TPR, True Positive Rate) от доли распознанных последовательностей негативной выборки (FPR, False Positive Rate). Однако для моделей распознавания мотивов СС ТФ в данных ChIP-seq эффективно измерять FPR не как долю последовательностей негативной выборки, а как ожидаемую частоту мотива в ней. Это позволяет гораздо точнее оценить предсказания модели мотива на жестких и даже средних по жесткости порогах распознавания (Tsukanov et al., 2022). Для модели распознавания мотивов СС ТФ точность распознавания может рассчитываться как частичная площадь под кривой ROC (pAUC ROC) (Tsukanov et al., 2022). Значение pAUC ROC равно части площади под кривой, ограниченной максимально допустимой ожидаемой частотой мотива. Площадь под кривой ROC интегрирует доли пиков, обладающих предсказанными СС ТФ (доля верно предсказанных пиков, ось Y) в широком диапазоне порогов распознавания, исчисляемых как частота мотива в негативной выборке (ось X).

В этой работе мы предлагаем подход MetArea, который рассматривает два отдельных «одиночных» мотива, а также некий «объединенный» мотив, означающий появление любого их двух одиночных. Для распознавания объединенного мотива в последовательности ДНК достаточно распознать в ней хотя бы один из двух одиночных мотивов при заданном пороге ожидаемой частоты мотива. Точный расчет частоты такого объединенного мотива даже по одной последовательности ДНК представляет собой некоторые трудности из-за огромного разнообразия возможных перекрываний одиночных мотивов. Поэтому для оценки точности модели мотива мы разработали и применили меру точности «Частичная площадь под кривой PR (Precision–Recall)», для расчета которой нужно отслеживать только число распознанных последовательностей позитивной и негативных выборок.

Кривая PR – это зависимость меры Precision (отношения числа предсказанных последовательностей в позитивной выборке к числу предсказанных последовательностей в позитивной и негативной выборках) от меры Recall (отношения числа предсказанных последовательностей позитивной выборки к объему выборки). Кривая PR – это альтернатива более популярной кривой ROC (Davis, Goadrich, 2006; Keilwagen et al., 2019). Достоинством меры «Площадь под кривой PR» по сравнению с мерой «Площадь под кривой ROC» является соотношение вкладов мягких и жестких порогов распознавания, соответствующих предсказанным сайтам с низкой и высокой аффинностью. По сравнению с кривой ROC, кривая PR обеспечивает бóльшие вклады от сайтов с высокой аффинностью, чем от сайтов с низкой. Кривая ROC поступает наоборот. Согласно кривой PR, вклад от сайтов с некоторой низкой аффинностью может даже стремиться к нулю, если такие сайты не содержат специфического нуклеотидного контекста. Это связано с равными вероятностями распознавания объектов в позитивной и негативной выборках (Saito, Rehmsmeier, 2015).

Мы разработали программный комплекс (ПК) MetArea для выявления пар мотивов СС ТФ со взаимоисключающей встречаемостью. ПК MetArea производит расчеты оценок точности pAUPRC (частичная площадь под кривой PR) для каждого из двух входных одиночных мотивов, а также для их сочетания – «объединенного мотива». Это позволяет выявить взаимоисключающую встречаемость двух входных мотивов.

Материалы и методы

В анализе использовались данные ChIP-seq из базы данных GTRD (Kolmykov et al., 2021). По каждому эксперименту ChIP-seq в анализ взят набор из 1000 пиков наилучшего качества, согласно предобработке инструментом MACS2 (Zhang et al., 2008). В настоящей работе в анализ взяты обогащенные мотивы, полученные из результатов *de novo* поиска мотивов, и мотивы СС ТФ мыши *Mus musculus* из базы данных Hocomoco (https://hocomoco12.autosome.org/) (Vorontsov et al., 2024). *De novo* поиск мотивов традиционной ПВМ и альтернативной SiteGA моделей мотивов СС ТФ проведен с помощью STREME https://meme-suite.org/meme/tools/streme (Bailey, 2021) и https://github.com/parthian-sterlet/sitega (Tsukanov et al., 2022). Значимость сходства обогащенных мотивов из результатов *de novo* поиска мотивов STREME с мотивами известных ТФ из баз данных Hocomoco, Cis-BP (Weirauch et al., 2014) и JASPAR (Rauluseviciute et al., 2024) оценивалась

с помощью инструмента TomTom (https://meme-suite.org/meme/tools/tomtom) (Gupta et al., 2007). Для анализа ПК MetArea допускает также мотивы из баз данных Hocomoco и JASPAR, отбираемые согласно принятой ранее методике (ПК MCOT) (Levitsky et al., 2019). Самый лучший хит модели мотива должен иметь ожидаемую частоту в выборке всех промоторов генов генома, кодирующих белки, не менее 2E–5. Лучший хит дает предсказанный сайт с максимально возможным значением функции распознавания модели мотива.

Итого в состав ПК MetArea включены 1420/1142 мотива для 942/713 ТФ человека/мыши из базы данных Hocomoco и 556/151 мотив для 555/148 ТФ растений/насекомых из базы данных JASPAR. ПК MetArea доступен по адресу https://github.com/parthian-sterlet/metarea. Детальное описание алгоритма ПК MetArea см. далее в разделе результатов. В ПК MetArea для оценки сходства анализируемых мотивов модели ПВМ (матриц частот нуклеотидов) реализован подход из ПК MCOT (Levitsky et al., 2019).

Результаты

Общее описание ПК MetArea

ПК MetArea позволяет анализировать как пары мотивов традиционной модели ПВМ, так и пары мотивов традиционной ПВМ и альтернативной SiteGA моделей (Levitsky et al., 2007; Tsukanov et al., 2022). Общая схема работы конвейера ПК MetArea представлена на рис. 3.

Входными данными и параметрами ПК MetArea являются:

- Два мотива: сочетание двух мотивов модели ПВМ, заданными двумя матрицами частот нуклеотидов (МЧН), или сочетание мотива модели ПВМ, заданной МЧН, и модели мотива SiteGA, заданной своей весовой матрицей, см. https://github.com/parthian-sterlet/sitega (Tsukanov et al., 2022).
- Позитивная выборка в формате FASTA (набор пиков ChIP-seq, NF последовательностей, Number of Foreground sequences).
- Негативная выборка в формате FASTA (NB последовательностей, Number of Background sequences). Ее рекомендуется приготовить предварительно по позитивной выборке и полному геному с помощью ПК AntiNoise (Raditsa et al., 2024), https://github.com/parthian-sterlet/antinoise. Для каждой последовательности позитивной выборки по ее длине и G/C-составу в полном геноме случайным образом находится несколько последовательностей негативной выборки. Далее в анализе NF/NB = 5.
- Выборка промоторов всех генов генома, необходимая для определения порогов распознавания на основе расчета таблиц ‘Threshold vs. ERR’ («Порог функции распознавания vs. Частота мотива в выборке всех промоторов генома») для каждого из входных мотивов.
- Порог ERR_{MAX} максимальной ожидаемой частоты для каждого входного мотива (Expected Recognition Rate, ERR).
- Таблицы ‘Threshold vs. ERR’ для каждого входного мотива.

Максимальная частота мотива 0.01 означает, что специфичность СС соответствует одному сайту на сто нуклеотидных позиций. Рекомендуемый интервал порога ожидаемой частоты мотива ERR_{MAX} – от 0.001 до 0.01. Далее в расчетах принято значение параметра $ERR_{MAX} = 0.002$. Ранее мы использовали таблицы ‘Threshold vs. ERR’ для унификации порогов распознавания разных мотивов (Levitsky et al., 2019; Tsukanov et al., 2021, 2022). Каждый мотив вместе со своей таблицей ‘Threshold vs. ERR’ подается в файле бинарного формата, генерируемого компонентами ПК MetArea для расчета ожидаемых частот мотива для моделей мотива ПВМ и SiteGA.

Выходными данными ПК MetArea являются:

- Текстовый файл с кривыми PR для каждого из входных мотивов, а также их объединенного мотива.
- Текстовый файл со значениями оценок точности распознавания pAUPRC для каждого из входных мотивов, а также их объединенного мотива, значением отношения площадей под кривыми (см. ниже), оценкой сходства мотивов (только для пар мотивов модели ПВМ).

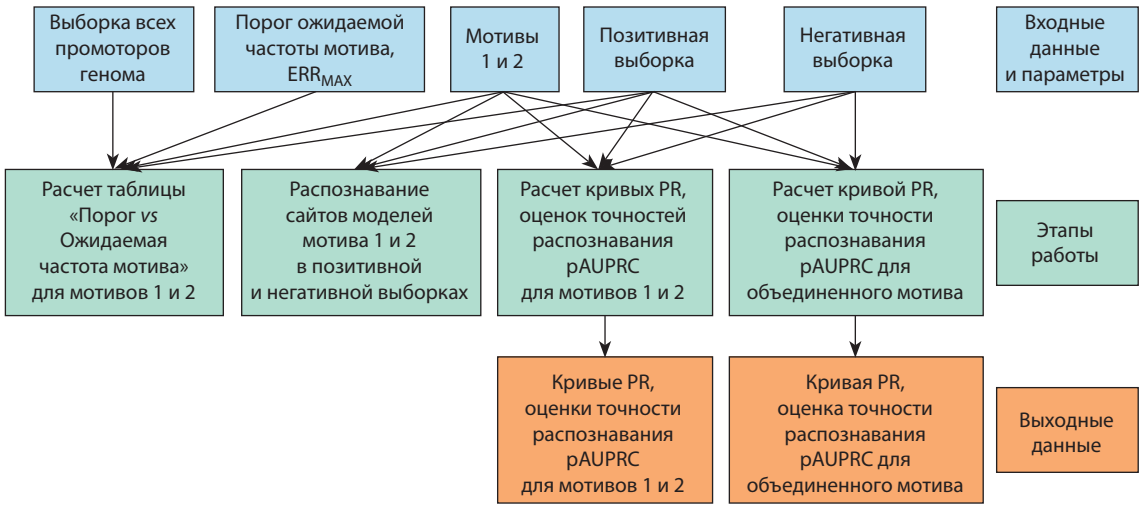


Рис. 3. Общая схема работы ПК MetArea.

Определение порогов распознавания для разных мотивов

Пороги функций распознавания каждого из двух входных мотивов согласно предварительно рассчитанным таблицам 'Threshold vs. ERR' переводятся в общую шкалу ожидаемой частоты мотива, ERR (Levitsky et al., 2019; Tsukanov et al., 2021, 2022). Это необходимо для построения PR кривой объединенного мотива. Ожидаемая частота мотива ERR для входных мотивов рассчитывается до порога ERR_{MAX}, так что все ожидаемые частоты удовлетворяют условию $ERR < ERR_{MAX}$.

Ожидаемую частоту мотива в выборке промоторов вычисляли следующим образом. Определяли значения функции распознавания мотива для каждого предсказанного сайта в выборке в каждой позиции и цепи ДНК. Затем для каждого порога функции распознавания ожидаемую частоту мотива вычисляли как отношение числа предсказанных СС, для которых значения этой функции равны порогу распознавания или выше него, к общему числу позиций, доступных для таких СС в выборке с учетом цепей ДНК.

Статистические метрики и кривая PR

Кривую PR, согласно (Davis, Goadrich, 2006), для модели мотива СС ТФ можно определить так: по оси X – отношение числа последовательностей позитивной выборки (пиков) с предсказанными сайтами к числу всех пиков (TPR, True Positive Rate, Recall, REC):

$$REC = \frac{TP}{TP+FN}. \quad (1)$$

Здесь TP/FN (True Positives/False Negatives) – число верно/неверно предсказанных последовательностей позитивной выборки (TP + FN = NF).

По оси Y кривой PR – отношение числа предсказанных последовательностей позитивной выборки к числу всех предсказанных последовательностей позитивной и негативной выборок (Precision, PREC), согласно (Davis, Goadrich, 2006):

$$PREC = \frac{TP}{TP+FP}. \quad (2)$$

Здесь FP (False Positives) – число предсказанных последовательностей негативной выборки. С учетом отличия объемов позитивной (NF) и негативной (NB) выборок мы поправили расчет величины Precision следующим образом:

$$PREC = \frac{TPR}{TPR+FPR} = \frac{TP/NF}{TP/NF+FP/NB} = \frac{TP}{TP+(NF/NB)\times FP}. \quad (3)$$

Здесь TPR и FPR – доли предсказанных последовательностей в позитивной и негативной выборках. Коэффициент NF/NB учитывает различие объемов негативной (NB) и позитивной (NF) выборок. Ожидаемое по случайным причинам число предсказанных последовательностей позитивной (TP) и негативной (FP) выборок пропорционально объемам выборок, NF и NB, соответственно. Целью замены формулы (2) на формулу (3) введением коэффициента NF/NB является унификация вида кривой PR для разных отношений объемов позитивной и негативной выборок.

Частичная площадь под кривой PR и отношение площадей под кривыми

Алгоритм MetArea использует таблицы «Порог функции распознавания vs. Частота мотива в выборке всех промоторов генома», описанные выше. Затем выполняется распознавание двух входных одиночных мотивов в позитивной и негативной выборках. Далее рассчитывается мера pAUPRC для одиночных мотивов, а также для объединенного мотива. Расчет частичной площади под кривой PR (pAUPRC) ограничен условиями, налагаемыми на меры Recall (ось X) и Precision (ось Y), т.е. площадь является частичной как по оси X, так и по оси Y (рис. 4).

Условием частичности площади под кривой PR по оси X является участие в расчетах меры pAUPRC части всего диапазона меры Recall от 0 до 1. Это условие означает, что в определении меры площади участвуют не все пики с предсказанными сайтами, а только те пики, наилучшие хиты в которых имеют ожидаемую частоту меньше порога, $ERR < ERR_{MAX}$ (см. рис. 4). Порог частоты $ERR_{MAX} = 0.002$ выбран более мягким, чем ранее использованный для анализа мотивов целевых ТФ порог $ERR_{MAX} = 0.001$

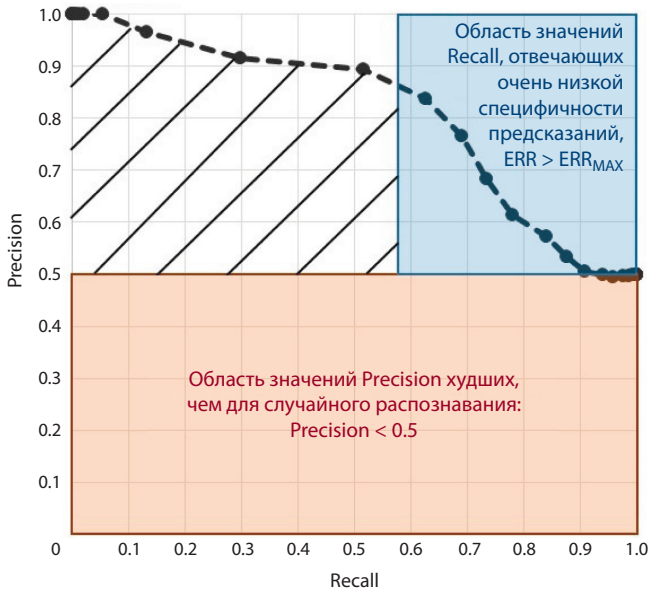


Рис. 4. Схема расчета частичной площади под кривой PR.

Ось X – мера Recall, вероятность предсказания последовательности позитивной выборки, Recall = TPR = TP/NF, формула (1). Ось Y – мера Precision, отношение вероятности предсказания последовательности позитивной выборки к сумме вероятностей предсказания последовательностей позитивной и негативной выборок, Precision = TPR/(TPR+FPR), формула (3). Розовая область отмечает значения Precision < 0.5, соответствующие предсказаниям худшим, чем предсказания случайной модели, равновероятно распознающей последовательности позитивной и негативной выборок. Precision > 0.5/Precision < 0.5 – области отбора в сторону позитивной/негативной выборок. Голубым показана область предсказанных последовательностей позитивной выборки с очень низкой специфичностью, соответствующих ожидаемой частоте мотива, большей порога, $ERR > ERR_{MAX}$. Для генерации данных примера негативной выборки взято нормальное распределение со средним и стандартным отклонением $(\mu_N, \sigma_N) = (5, 2.5)$, а позитивная выборка – это смесь 50 % на 50 % нормальных распределений $(\mu_{P1}, \sigma_{P1}) = (10, 1)$ и $(\mu_{P2}, \sigma_{P2}) = (5.5, 4)$. Эти распределения моделируют сайты, проходящие и не проходящие порог ERR_{MAX} ожидаемой частоты мотива. Штриховка обозначает область, по которой определяется метрика частичной площади под кривой pAUPRC.

(Tsukanov et al., 2022), поскольку ранее проводился анализ мотивов целевых ТФ ChIP-seq экспериментов, а ПК MetArea анализирует как мотивы СС целевых ТФ, так и менее консервативные мотивы СС партнерских ТФ.

Условием частичности площади под кривой PR по оси *Y* является вычитание из каждого значения меры Precision ее ожидаемого значения $PREC_{EXP}$ (см. рис. 4) (Saito, Rehmsmeier, 2015). Для модели, равновероятно распознающей данные позитивной и негативной выборки, PR кривая представляет собой горизонтальную линию:

$$PREC_{EXP} = \frac{NF}{NF+NB} = 0.5. \quad (4)$$

Значение $PREC_{EXP}$ постоянно и равно 0.5, поскольку ранее значение FP было нормализовано, так что объемы выборок в этой формуле уже можно считать равными. В итоге частичная площадь под кривой PR в ПК MetArea вычисляется как следующая сумма:

$$pAUPRC = \frac{2}{NF} \times \sum_{i=1}^{NI} \left[\left\{ \frac{PREC(i)+PREC(i-1)}{2} - PREC_{EXP} \right\} \times \right.$$
$$\left. \times \left\{ REC(i) - REC(i-1) \right\} \right]. \quad (5)$$

Здесь NI – самый мягкий порог, определяемый, как описано выше, по ожидаемым частотам и входному параметру ERR_{MAX}. Коэффициент 2/NF в формуле (5) необходим для нормировки значения pAUPRC на максимальное значение 1, так как при максимальных значениях меры Precision, равных 1, максимальное значение первого множителя под суммой, $\{(PREC(i)+PREC(i-1))/2 - PREC_{EXP}\}$, равно 0.5, а максимальное значение суммы всех вторых множителей, $\{REC(i)-REC(i-1)\}$, равно NF – объему позитивной выборки.

Критерий предсказания функциональной связи мотивов отражает повышение оценки точности объединенного мотива по сравнению с оценками точности одиночных мотивов и является количественной оценкой взаимоисключающей встречаемости в парах мотивов. Для пары мотивов A и B критерий требует более высокого значения оценки точности pAUPRC(A&B) объединенного мотива A&B по сравнению со значениями оценок точностей обоих одиночных мотивов, pAUPRC(A) и pAUPRC(B). Рассчитанное таким образом Отношение Площадей Под Кривыми (ОППК, Ratio of Areas Under Curves, RAUC) должно быть больше единицы:

$$RAUC(A, B) = \frac{pAUPRC(A\&B)}{Max\{pAUPRC(A), pAUPRC(B)\}} > 1. \quad (6)$$

Варианты применения ПК MetArea

Входными данными ПК MetArea могут быть мотивы СС ТФ, для которых предполагается обогащение в позитивной выборке по сравнению с негативной, например, такие мотивы – это результат *de novo* поиска мотивов (Bailey, 2021). Отдельные варианты применения ПК реализуют массовый анализ коллекций мотивов СС ТФ из баз данных Hocomoco и JASPAR. Анализ множества пар мотивов позволяет найти пары, для которых обнаруживается наибольший рост оценки точности распознавания pAUPRC при объединении мотивов. ПК MetArea допускает несколько вариантов применения, реализованных в виде отдельных программ. В следующих вариантах применения рассматривается модель мотива ПВМ:

- два заданных мотива;
- несколько заданных мотивов, для K мотивов проверяются все возможные $\{K \times (K-1)/2\}$ пары;
- заданный мотив против всех (M) мотивов СС известных ТФ из базы данных. Для заданного мотива проверяются все его M пар с мотивами из коллекции Hocomoco (человек, мышь) или JASPAR (растения, насекомые);
- все мотивы СС известных ТФ из базы данных. Из всех M мотивов известных ТФ из коллекции Hocomoco или JASPAR отбираются K мотивов с наивысшими оценками точности pAUPRC и проверяются все возможные пары этих мотивов, их всего $\{K \times (K-1)/2\}$.

Вариант применения для мотивов моделей ПВМ и SiteGA:

- мотив ПВМ и мотив SiteGA.

Далее приведены примеры результатов анализа данных ChIP-seq для разных вариантов применения ПК MetArea.

Анализ нескольких заданных мотивов модели ПВМ

Рассмотрим набор данных ChIP-seq для ТФ BHLHA15 (Hess et al., 2016) (GTRD PEAKS039234, GEO GSE86289) для поджелудочной железы мыши. *De novo* поиск мотивов с помощью инструмента STREME (Bailey, 2021) показал, что среди пяти мотивов с самым высоким обогащением четыре имеют значимое сходство (p-value < 0.001) (Gupta et al., 2007) с известными мотивами СС ТФ BHLHA15 из Hocomoco (BHA15.H12CORE.0.P.B, мотивы 1 и 5; BHA15.H12CORE.1.SM.B, мотивы 2 и 4). Мотивы 1/5 и 2/4 соответствуют консенсусу E-бокс CAnnTG со спейсерами GC и TA, поэтому они обозначены BHLHA15_GC_1/BHLHA15_GC_2 и BHLHA15_TA_1/BHLHA15_TA_2 соответственно. Мотив 3 имеет значимое сходство (p-value < <0.001) с мотивом СС ТФ CTCF (CTCF.H12CORE.0.P.B) (рис. 5, *а*).

Анализ значений оценок точности распознавания pAUPRC для одиночных мотивов и их попарных объединенных мотивов (см. рис. 5, *б*) проводится на основе соответствующих значений ОППК для пар мотивов (см. рис. 5, *в*), оценка сходства пар мотивов нужна для контроля значимо похожих мотивов (см. рис. 5, *г*). Высокие ОППК показаны для пар мотивов BHLHA15_GC_1/BHLHA15_TA2 и BHLHA15_GC_1/BHLHA15_TA1, кривые PR по ним показаны на рис. 5, *д*, *е*. Мотив CTCF имеет высокие ОППК с мотивами BHLHA15_GC1 и BHLHA15_TA2 (см. рис. 5, *в*). В паре мотивов BHLHA15_TA2 и CTCF обнаружено максимальное значение ОППК = 1.48 (см. рис. 5, *в*). Полученные результаты согласуются со способностью ТФ BHLHA15 связываться с ДНК только в составе димера ТФ класса bHLH (Amoutzias et al., 2008). Тренд к расхождению разных по структуре СС ТФ BHLHA15 в разные пики может означать, что в состав димера могут входить разные ТФ класса bHLH (включая и ТФ BHLHA15) и что на связывание димера влияют другие партнерские ТФ, формирующие с ТФ BHLHA15 мультибелковые комплексы, так что ДСД ТФ BHLHA15 принимает разную конформацию. Например, таким партнерским ТФ может быть CTCF, мотив СС

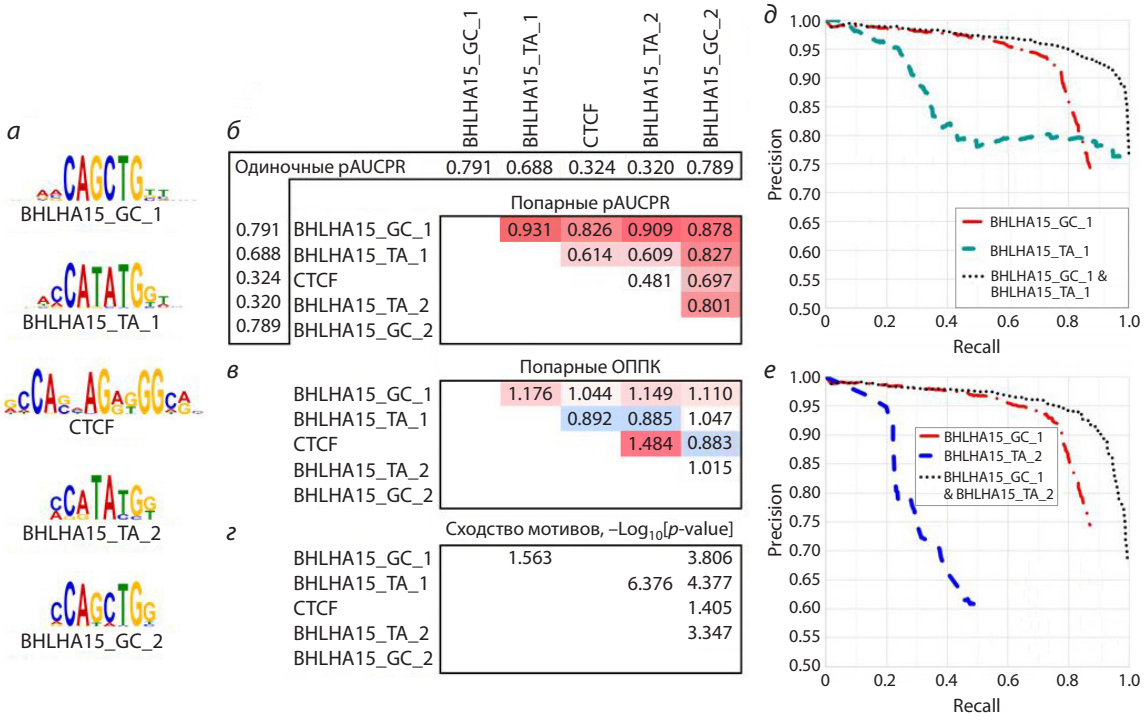


Рис. 5. Анализ пяти самых обогащенных мотивов из результатов *de novo* поиска мотивов (STREME) (Bailey, 2021) для набора данных ChIP-seq по ТФ BHLHA15 мыши (Hess et al., 2016, GTRD PEAKS039234, GEO GSM2299654/GSM2299655).

а – лого пяти мотивов, сортировка по значимости обогащения, полученной инструментом STREME; обозначения мотивов СС ТФ BHLHA15 согласно динуклеотиду в его спейсере в общем консенсусе CAnnTG; *б* – таблица попарных величин оценок точности pAUPRC объединенных мотивов, построенных по попарным сочетаниям мотивов; в заголовках указаны значения pAUPRC для одиночных мотивов, оттенки красного цвета отмечают максимальные значения pAUPRC объединенного мотива; *в* – таблица ОППК в парах мотивов, оттенки красного и синего цвета отмечают значения больше и меньше единицы; *г* – таблица значимостей сходства мотивов, $-\mathrm{Log}_{10}[p\text{-value}]$; *д*, *е* – кривые PR для одиночных мотивов и их попарных объединенных мотивов BHLHA15_GC_1/BHLHA15_TA1 и BHLHA15_GC_1/BHLHA15_TA2.

которого также обогащен (см. рис. 5, *а*). Согласно экспериментальным данным, (1) несколько ТФ из класса bHLH имеют белок-белковые взаимодействия с ТФ CTCF (база BIOGRID https://thebiogrid.org/); (2) анализ партнерских ТФ по геномной колокализации (Hu et al., 2020) подтверждает, что несколько ТФ класса bHLH колокализованы с ТФ CTCF в одних геномных локусах *in vivo*.

Анализ всех мотивов СС известных ТФ из базы данных

Рассмотрим набор данных ChIP-seq по ТФ AR (Androgene Receptor, рецептор андрогена) для простаты мыши (Chen et al., 2013) (GTRD PEAKS035588, GEO GSM1145307). На рис. 6 для этого набора данных ChIP-seq показана матрица попарных значений ОППК для 15 самых обогащенных мотивов СС ТФ согласно мере pAUPRC из всех 1142 мотивов СС ТФ мыши из базы данных Hocomoco. Семь из 15 мотивов принадлежат СС ТФ AR и его гомологам из подсемейства GR-like (NR3C) {2.1.1.1} семейства Steroid hormone receptors {2.1.1} класса Nuclear receptors with C4 zinc fingers {2.1}. Это семейство определяет целевой ТФ AR и вероятные мотивы его СС. Остальные восемь мотивов принадлежат СС ТФ из подсемейств FOXA {3.3.1.1}, FOXJ {3.3.1.10}, FOXM {3.3.1.13} и FOXP {3.3.1.16} одного семейства FOX {3.3.1} класса Fork head/winged helix factors {3.3}. ТФ этого семейства – предполагаемые партнерские ТФ для ТФ AR, например ТФ Foxa1, – известен для этой же ткани простаты (Yang, Yu, 2015).

Значения ОППК больше 1 получены почти для всех пар мотивов GR-like/FOX. Например, значение ОППК = = 1.03 в паре ANDR.H12CORE.0.P.B (ранг pAUPRC 1) и FOXA2.H12CORE.0.PSM.A (ранг 5) соответствует максимальному значению 0.853 pAUPRC по парам мотивов GR-like/FOX. Значения ОППК для пар мотивов GR-like/GR-like превышают 1 лишь для некоторых пар мотивов. Мотив ANDR.H12CORE.2.P.B (ранг 7), отличный по консенсусу от всех остальных мотивов GR-like (AAACA вместо GNACA, см. рис. 6, колонка Лого), имеет высокие значения ОППК. Также это единственный мотив, для которого ОППК больше 1 в парах со всеми остальными мотивами GR-like и FOX. В частности, среди пар мотивов GR-like/GR-like максимальное значение pAUPRC 0.876 при ОППК = 1.06 достигается в паре мотивов ANDR.H12CORE.0.P.B (ранг 1) и ANDR.H12CORE.2.P.B (ранг 7). Высокие значения ОППК в парах мотивов GR-like/GR-like обнаружены и для мотива MCR.H12CORE.1.SM.B, однако он имеет самый низкий ранг pAUPRC – 15. MCR.H12CORE.1.SM.B является мотивом связывания мономера, а не димера. Среди пар мотивов FOX/FOX значений ОППК больше 1 почти нет.

В целом высокие значения ОППК многих пар мотивов GR-like/GR-like позволяют предположить, что ТФ AR связывается в разных пиках с помощью различных структурных вариантов мотивов GR-like. Аналогичное предположение можно сделать и о связывании димера, вклю-

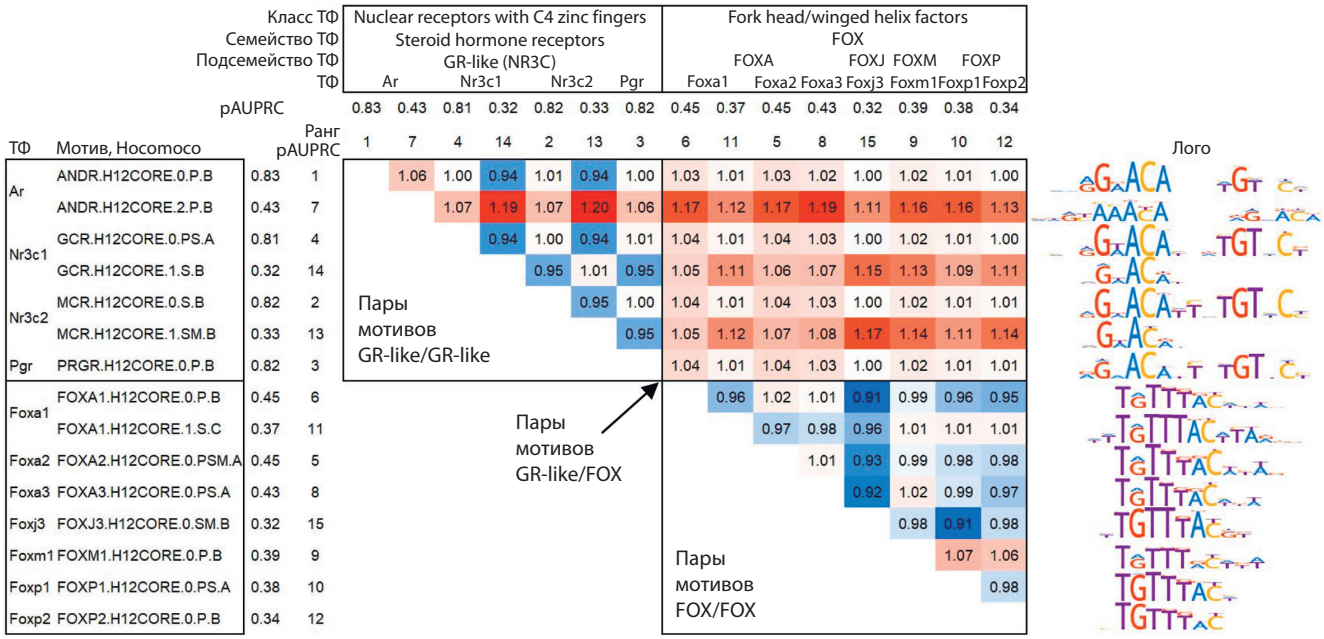


Рис. 6. Результаты анализа мотивов СС известных ТФ из базы данных Hocomoco для набора данных ChIP-seq по ТФ AR в простате мыши (Chen et al., 2013).

В анализ включены 15 самых обогащенных мотивов по оценкам точности pAUPRC. Их значения и ранги, а также имена ТФ из базы данных Hocomoco приведены в заголовках рядов и колонок. В заголовках рядов указаны идентификаторы мотивов из Hocomoco, а в заголовках колонок – имена классов, семейств и подсемейств ТФ. В таблице оттенками красного/синего цвета обозначены изменения ОППК в большую/меньшую сторону от нейтрального значения 1. В самой правой колонке указаны лого мотивов из базы данных Hocomoco. Черными рамками отмечены мотивы GR-like и FOX в заголовках рядов и колонок, а также пары мотивов СС ТФ GR-like/GR-like, GR-like/FOX и FOX/FOX в таблице.

чающего ТФ AR и ТФ семейства FOX на основе высоких значений ОППК для пар мотивов GR-like/FOX. Полученные результаты для ChIP-seq данных по ТФ AR означают, что связывание с ДНК ТФ AR происходит в составе димеров AR/AR и AR/Foxa1 (если в условиях эксперимента с мотивами FOX связывается именно ТФ Foxa1) и что оба ТФ допускают большое разнообразие разных структурных типов СС, так что разные пары мотивов расходятся по разным пикам.

Анализ пары мотивов моделей ПВМ и SiteGA

Рассмотрим набор данных ChIP-seq для ТФ E2F4 для дендритных клеток первичного врожденного иммунитета, полученных из костного мозга мыши, стимулированных патогенным компонентом липополисахаридом в течение 120 мин (Garber et al., 2012) (GTRD PEAKS035857, GEO GSM881061). Кривые PR для мотивов ПВМ, SiteGA и их объединенного мотива ПВМ & SiteGA, рассчитанные с помощью ПК MetArea, представлены на рис. 7. Значения pAUPRC для мотивов ПВМ, SiteGA и объединенного мотива ПВМ & SiteGA равны 0.457, 0.358 и 0.47 соответственно, значение ОППК объединенного мотива равно 1.028.

Модели мотивов ПВМ и SiteGA основаны на совершенно разных методологических принципах (Levitsky et al., 2007). Модель ПВМ представляет сайты с высокой аффинностью, определяемые самыми консервативными позициями и наиболее частыми нуклеотидами в них. Модель SiteGA подразумевает сайты, содержащие зависимости разных позиций, которые, по-видимому, возникают из-за кооперативных действий как минимум двух ТФ при связывании с ДНК (Morgunova, Taipale, 2017; Levitsky et al., 2020). Сайты модели SiteGA заметно менее консервативны, чем сайты модели ПВМ, модель SiteGA способна лучше, чем модель ПВМ, предсказывать сайты с низкой аффинностью (Tsukanov et al., 2022). Объединение моделей ПВМ и SiteGA позволяет улучшить распознавание сайтов с низкой аффинностью, что отражает бóльшая

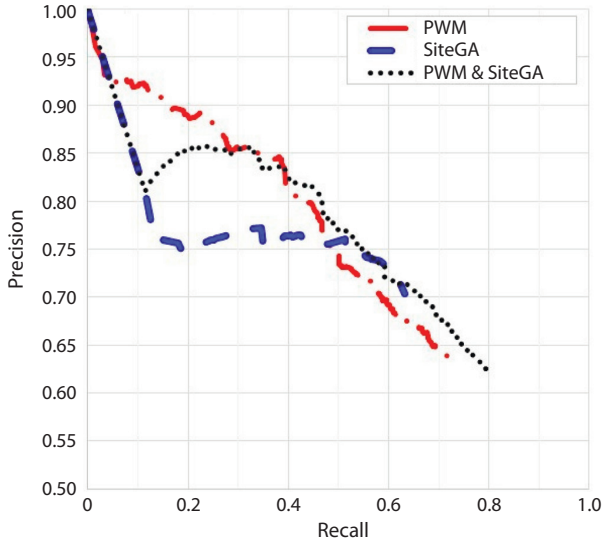


Рис. 7. Результаты анализа пары мотивов моделей ПВМ и SiteGA с помощью ПК MetArea.

Красным, синим и черным цветом показаны кривые PR для мотивов ПВМ, SiteGA и объединенного мотива ПВМ & SiteGA. В анализе использован набор данных ChIP-seq для ТФ E2F4 (GTRD PEAKS035857, GEO GSM881061).

протяженность кривой PR объединенного мотива ПВМ & SiteGA по оси *X* (Recall) по сравнению с каждым из одиночных мотивов ПВМ и SiteGA. Хотя объединенный мотив и имеет меньшие значения величины Precision (см. рис. 7, ось *Y*), чем модель ПВМ, более широкий интервал величин Recall (ось *X*) определяет рост меры pAUPRC объединенного мотива. Одиночные мотивы до порога ожидаемой частоты мотива $ERR_{MAX} = 0.002$ распознают 73.2 % (ПВМ) и 63.3 % (SiteGA) пиков, объединенный мотив распознает 79.9 %.

Гипотеза о том, что модели ПВМ и SiteGA представляют собой различные структурные варианты СС ТФ E2F4, подтверждается инструментом TomTom сравнения мотивов (p-value < 0.05) (Gupta et al., 2007): для модели ПВМ – по ее матрице частот нуклеотидов, а для модели SiteGA, как и ранее (Tsukanov et al., 2022), – по матрице частот нуклеотидов, построенной по предсказанным сайтам. На способность ТФ E2F4 связываться с разными структурными типами СС указывает также то, что в эксперименте M. Garber с коллегами (2012) в одних условиях определены геномные локусы связывания 25 ТФ и показано, что локусы ТФ E2F4 значимо перекрываются с локусами пяти ТФ: EGR2, EGR1, IRF2, ETS2 и E2F1. Следовательно, можно предполагать, что ТФ E2F4 входит в одни мультибелковые комплексы с этими ТФ. Поэтому в разных локусах ТФ E2F4 вынужден в большей или меньшей степени изменять свои СС, чтобы адаптироваться к СС партнерских ТФ.

Обсуждение

В нашей работе предложен новый подход MetArea для выявления взаимоисключающей встречаемости в парах мотивов СС ТФ на основе анализа наборов данных ChIP-seq. Если два мотива являются структурно различными мотивами СС одного ТФ в разных пиках одного набора, то взаимоисключающая встречаемость обусловлена тем, что в пиках этот ТФ предпочитает либо один, либо другой структурный тип СС, но реже наблюдаются два СС разной структуры в одном пике. Если мотивы СС относятся к двум разным ТФ, то взаимоисключающую встречаемость можно объяснить тем, что в составе мультибелкового комплекса, включающего оба ТФ, в разных пиках один или другой ТФ связывается с ДНК напрямую, но реже наблюдаются также СС обоих ТФ в одном пике.

В ходе разработки ПК MetArea мы отказались от применения метрики частичной площади под кривой ROC (pAUC ROC) (Levitsky, Tsukanov, 2024) и использовали метрику площади по кривой PR (Davis, Goadrich, 2006) для определения метрики частичной площади под кривой PR. Ранее высказано предположение (Davis, Goadrich, 2006), что применение метрики площади под кривой AUC ROC может быть некорректным, если реальные пороги распознавания классификатора должны быть достаточно жесткими, а преимущество одной модели относительно другой набирается в интервале более мягких порогов распознавания (на правом хвосте кривой ROC). Для корректного сравнения двух мотивов в таком случае вместо меры площади под кривой AUC ROC мы ранее использовали меру «Частичная площадь под кривой ROC, pAUC», которая вместо полноразмерного диапазона ошибок перепредсказания (False Positive Rate, доля в распознавании объектов негативной выборки, ось *X* кривой ROC) от 0 до 1 использует только некую ее левую часть, отбрасывая диапазон слишком больших ошибок перепредсказания. Этот подход нами реализован для сравнения точности распознавания мотивов СС ТФ моделей ПВМ, BaMM и SiteGA (Tsukanov et al., 2022), где пороги распознавания при вычислении оценки точности pAUC ROC были ограничены условием ERR < 0.001.

К сожалению, такой способ не подходит для расчета оценки точности объединенного мотива, необходимого при реализации подхода MetArea. Причина этого в определении частоты объединенного мотива (т.е. числа его хитов), которое возможно при условии неперекрывания хитов одиночных мотивов, а в случае их перекрывания частота объединенного мотива должна некоторым образом понижаться. Альтернативным способом избавления от завышенной оценки точности, даваемой мерой AUC ROC, является переход от кривой ROC к кривой PR и вычисление площади под PR (Davis, Goadrich, 2006; Keilwagen et al., 2019).

Ранее было предложено несколько подходов, нацеленных на выявление встречаемости разных мотивов СС ТФ или разных наборов мотивов в различных долях пиков одного набора пиков ChIP-seq. Инструмент DIVERSITY (Mitra et al., 2018) разбивает пики ChIP-seq одного набора на несколько неперекрывающихся групп, так что каждая группа описывается своим обогащенным мотивом из результатов *de novo* поиска. Позднее авторы допустили, что каждая группа представляется не одним мотивом, а комбинацией нескольких мотивов. Инструмент cisDIVERSITY (Biswas, Narlikar, 2021) для набора пиков проводит *de novo* поиск обогащенных мотивов с помощью модели ПВМ, а затем разделяет найденные мотивы по нескольким неперекрывающимся группам пиков так, что все группы составляют весь набор пиков. Каждый из мотивов имеет разные частоты по группам; например, в некоторых группах частота мотива выше, чем в других группах, а в других группах мотива может не быть. Задачи инструментов DIVERSITY/cisDIVERSITY и MetArea сходны в том, что происходит разделение разных мотивов по разным долям пиков. Однако инструменты DIVERSITY/cisDIVERSITY: (1) выявляют все разнообразие мотивов и разделяют все пики на группы, с целью найти для разных групп разные мотивы или их комбинации; (2) ограничены традиционной моделью мотива ПВМ. ПК MetArea: (1) рассматривает только пары мотивов, с целью за счет максимизации меры точности pAUPRC для объединенного мотива найти пары, наилучшим образом дополняющие друг друга; (2) рассматривает как традиционную ПВМ, так и альтернативные модели мотива СС ТФ.

Заключение

Нами разработан ПК MetArea, который по заданному набору пиков ChIP-seq рассчитывает меру точности «Частичная площадь под кривой PR» (pAUPRC) для двух входных одиночных мотивов СС ТФ, определяет по ним объединенный мотив и также рассчитывает меру pAUPRC для него. Создание объединенного мотива по двум одиночным мотивам и расчет для него оценки точно-

сти pAUPRC позволяют сравнить в единой шкале два одиночных мотива и их общее действие. Превышение оценки точности объединенного мотива над оценками точности обоих одиночных мотивов означает их взаимоисключающую встречаемость. Результаты анализа с помощью ПК MetArea позволяют предсказывать функциональную связь двух мотивов, а значит, и соответствующих им ТФ. В частности, ПК MetArea может предложить существенные аргументы за или против гипотезы о том, что два мотива являются структурными вариантами СС одного ТФ. Аналогично предлагается поддержка или опровержение гипотезы о том, что мотивы СС представляют два ТФ, вместе вовлеченных в регуляцию транскрипции генов в составе одного мультибелкового комплекса. В итоге для заданного набора ChIP-seq данных ПК MetArea предсказывает (1) структурное разнообразие СС отдельных ТФ и (2) пары мотивов СС разных ТФ, действующих для регуляции транскрипции генов в составе одних мультибелковых комплексов из многих ТФ.

Список литературы / References

Ambrosini G., Vorontsov I., Penzar D., Groux R., Forne O., Nikolaeva D.D., Ballester B., Grau J., Grosse I., Makeev V., Kulakovskiy I., Buche P. Insights gained from a comprehensive all-against-all transcription factor binding motif benchmarking study. *Genome Biol*. 2020;21:114. doi 10.1186/s13059-020-01996-3

Amoutzias G.D., Robertson D.L., Van de Peer Y., Oliver S.G. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci*. 2008;33(5):220-229. doi 10.1016/j.tibs.2008.02.002

Bailey T.L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics*. 2021;37:2834-2840. doi 10.1093/bioinformatics/btab203

Biswas A., Narlikar L. A universal framework for detecting cis-regulatory diversity in DNA regions. *Genome Res*. 2021;31(9):1646-1662. doi 10.1101/gr.274563.120

Chen Y., Chi P., Rockowitz S., Iaquinta P.J., Shamu T., Shukla S., Gao D., Sirota I., Carver B.S., Wongvipat J., Scher H.I., Zheng D., Sawyers C.L. ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat. Med*. 2013;19(8):1023-1029. doi 10.1038/nm.3216

Davis J., Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning. New York: Assoc. for Computing Machinery, 2006;233-240. doi 10.1145/1143844.1143874

D'haeseleer P. What are DNA sequence motifs? *Nat. Biotechnol*. 2006;24(4):423-425. doi 10.1038/nbt0406-423

Garber M., Yosef N., Goren A., Raychowdhury R., Thielke A., Guttman M., Robinson J., Minie B., Chevrier N., Itzhaki Z., Blecher-Gonen R., Bornstein C., Amann-Zalcenstein D., Weiner A., Friedrich D., Meldrim J., Ram O., Cheng C., Gnirke A., Fisher S., Friedman N., Wong B., Bernstein B.E., Nusbaum C., Hacohen N., Regev A., Amit I. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell*. 2012;47(5):810-822. doi 10.1016/j.molcel.2012.07.030

Georgakopoulos-Soares I., Deng C., Agarwal V., Chan C.S.Y., Zhao J., Inoue F., Ahituv N. Transcription factor binding site orientation and order are major drivers of gene regulatory activity. *Nat. Commun*. 2023;14:2333. doi 10.1038/s41467-023-37960-5

Gupta S., Stamatoyannopolous J.A., Bailey T.L., Noble W.S. Quantifying similarity between motifs. *Genome Biol*. 2007;8(2):R24. doi 10.1186/gb-2007-8-2-r24

Hess D.A., Strelau K.M., Karki A., Jiang M., Azevedo-Pouly A.C., Lee A.H., Deering T.G., Hoang C.Q., MacDonald R.J., Konieczny S.F. MIST1 links secretion and stress as both target and regulator of the unfolded protein response. *Mol. Cell. Biol*. 2016;36(23):2931-2944. doi 10.1128/MCB.00366-16

Hu G., Dong X., Gong S., Song Y., Hutchins A.P., Yao H. Systematic screening of CTCF binding partners identifies that BHLHE40 regulates CTCF genome-wide distribution and long-range chromatin interactions. *Nucleic Acids Res*. 2020;48(17):9606-9620. doi 10.1093/nar/gkaa705

Johnson D.S., Mortazavi A., Myers R.M., Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007;316(5830):1497-1502. doi 10.1126/science.1141319

Keilwagen J., Posch S., Grau J. Accurate prediction of cell type-specific transcription factor binding. *Genome Biol*. 2019;20(1):9. doi 10.1186/s13059-018-1614-y

Kel O.V., Romaschenko A.G., Kel A.E., Wingender E., Kolchanov N.A. A compilation of composite regulatory elements affecting gene transcription in vertebrates. *Nucleic Acids Res*. 1995;23(20):4097-4103. doi 10.1093/nar/23.20.4097

Kolmykov S., Yevshin I., Kulyashov M., Sharipov R., Kondrakhin Y., Makeev V.J., Kulakovskiy I.V., Kel A., Kolpakov F. GTRD: an integrated view of transcription regulation. *Nucleic Acids Res*. 2021;49(D1):D104-D111. doi 10.1093/nar/gkaa1057

Lambert S.A., Jolma A., Campitelli L.F., Das P.K., Yin Y., Albu M., Chen X., Taipale J., Hughes T.R., Weirauch M.T. The human transcription factors. *Cell*. 2018;172(4):650-665. doi 10.1016/j.cell.2018.01.029

Levitsky V.G., Ignatieva E.V., Ananko E.A., Turnaev I.I., Merkulova T.I., Kolchanov N.A., Hodgman T.C. Effective transcription factor binding site prediction using a combination of optimization, a genetic algorithm and discriminant analysis to capture distant interactions. *BMC Bioinformatics*. 2007;8(1):481. doi 10.1186/1471-2105-8-481

Levitsky V., Zemlyanskaya E., Oshchepkov D., Podkolodnaya O., Ignatieva E., Grosse I., Mironova V., Merkulova T. A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. *Nucleic Acids Res*. 2019;47:e139. doi 10.1093/nar/gkz800

Levitsky V., Oshchepkov D., Zemlyanskaya E., Merkulova T. Asymmetric conservation within pairs of co-occurred motifs mediates weak direct binding of transcription factors in ChIP-Seq data. *Int. J. Mol. Sci*. 2020;21(17):E6023. doi 10.3390/ijms21176023

Levitsky V.G., Tsukanov A.V. MetArea tool for predicting structural variability and cooperative binding of transcription factors in ChIP-seq data. In: 14th International Conference on Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2024). 2024;136-138. doi 10.18699/bgrs2024-1.2-17

Mitra S., Biswas A., Narlikar L. DIVERSITY in binding, regulation, and evolution revealed from high-throughput ChIP. *PLoS Comput. Biol*. 2018;14(4):e1006090. doi 10.1371/journal.pcbi.1006090

Morgunova E., Taipale J. Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol*. 2017;47:1-8. doi 10.1016/j.sbi.2017.03.006

Nagy G., Nagy L. Motif grammar: the basis of the language of gene expression. *Comput. Struct. Biotechnol. J*. 2020;18:2026-2032. doi 10.1016/j.csbj.2020.07.007

Raditsa V.V., Tsukanov A.V., Bogomolov A.G., Levitsky V.G. Genomic background sequences systematically outperform synthetic ones in de novo motif discovery for ChIP-seq data. *NAR Genom. Bioinform*. 2024;6(3):lqae090. doi 10.1093/nargab/lqae090

Rauluseviciute I., Riudavets-Puig R., Blanc-Mathieu R., Castro-Mondragon J.A., Ferenc K., Kumar V., Lemma R.B., Lucas J., Chèneby J., Baranasic D., Khan A., Fornes O., Gundersen S., Johansen M., Hovig E., Lenhard B., Sandelin A., Wasserman W.W., Parcy F., Mathelier A. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2024;52(D1):D174-D182. doi 10.1093/nar/gkad1059

Rogers J.M., Waters C.T., Seegar T.C.M., Jarrett S.M., Hallworth A.N., Blacklow S.C., Bulyk M.L. Bispecific forkhead transcription factor

FoxN3 recognizes two distinct motifs with different DNA shapes. *Mol. Cell*. 2019;74(2):245-253.e6. doi 10.1016/j.molcel.2019.01.019

Saito T., Rehmsmeier M. The Precision-Recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015;10(3):e0118432. doi 10.1371/journal.pone.0118432

Siebert M., Söding J. Bayesian Markov models consistently outperform PWMs at predicting motifs in nucleotide sequences. *Nucleic Acids Res*. 2016;44:6055-6069. doi 10.1093/nar/gkw521

Tognon M., Giugno R., Pinello L. A survey on algorithms to characterize transcription factor binding sites. *Brief. Bioinform*. 2023;24(3):bbad156. doi 10.1093/bib/bbad156

Tsukanov A.V., Levitsky V.G., Merkulova T.I. Application of alternative *de novo* motif recognition models for analysis of structural heterogeneity of transcription factor binding sites: a case study of FOXA2 binding sites. *Vavilov J. Genet. Breed*. 2021;25(1):7-17. doi 10.18699/VJ21.002

Tsukanov A.V., Mironova V.V., Levitsky V.G. Motif models proposing independent and interdependent impacts of nucleotides are related to high and low affinity transcription factor binding sites in Arabidopsis. *Front. Plant Sci*. 2022;13:938545. doi 10.3389/fpls.2022.938545

Vorontsov I.E., Eliseeva I.A., Zinkevich A., Nikonov M., Abramov S., Boytsov A., Kamenets V., Kasianova A., Kolmykov S., Yevshin I.S., Favorov A., Medvedeva Y.A., Jolma A., Kolpakov F., Makeev V.J., Kulakovskiy I.V. HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res*. 2024;52(D1):D154-D163. doi 10.1093/nar/gkad1077

Wasserman W.W., Sandelin A. Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet*. 2004;5(4):276-287. doi 10.1038/nrg1315

Weirauch M.T., Yang A., Albu M., Cote A.G., Montenegro-Monter A., Drewe P., Najafabadi H.S., Lambert S.A., Mann I., Cook K., Zheng H., Goity A., van Bakel H., Lozano J.C., Galli M., Lewsey M.G., Huang E., Mukherjee T., Chen X., Reece-Hoyes J.S., Govindarajan S., Shaulsky G., Walhout A.J.M., Bouget F.Y., Ratsch G., Larrondo L.F., Ecker J.R., Hughes T.R. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. 2014;158(6):1431-1443. doi 10.1016/j.cell.2014.08.009

Wingender E. Criteria for an updated classification of human transcription factor DNA-binding domains. *J. Bioinform. Comput. Biol*. 2013;11(1):1340007. doi 10.1142/S0219720013400076

Wingender E., Schoeps T., Dönitz J. TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Res*. 2013;41(D1):D165-D170. doi 10.1093/nar/gks1123

Wingender E., Schoeps T., Haubrock M., Dönitz J. TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res*. 2015;43(D1):D97-D102. doi 10.1093/nar/gku1064

Wingender E., Schoeps T., Haubrock M., Krull M., Dönitz J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res*. 2018;46(D1):D343-D347. doi 10.1093/nar/gkx987

Yang Y.A., Yu J. Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. *Genes Dis*. 2015;2(2):144-151. doi 10.1016/j.gendis.2015.01.003

Zeitlinger J. Seven myths of how transcription factors read the cis-regulatory code. *Curr. Opin. Syst. Biol*. 2020;23:22-31. doi 10.1016/j.coisb.2020.08.002

Zhang Y., Liu T., Meyer C.A., Eeckhoute J., Johnson D.S., Bernstein B.E., Nussbaum C., Myers R.M., Brown M., Li W., Liu X.S. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137. doi 10.1186/gb-2008-9-9-r137

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 19.10.2024. После доработки 20.11.2024. Принята к публикации 21.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):834-842

doi 10.18699/vjgb-24-91

Компьютерный анализ показывает отличия митохондриальных микроРНК от остальных микроРНК

П.С. Ворожейкин[1], И.И. Титов[1, 2, 3]

[1] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
pavel.vorozheykin@gmail.com

Аннотация. Одним из подклассов микроРНК с до сих пор неизвестными специальными функциями являются митомиРы (mitomiRs) – митохондриальные микроРНК, которые в основном происходят из ядерной ДНК и импортируются в митохондрии, при этом изменение уровня их экспрессии ассоциировано с рядом заболеваний. Для выявления характерных особенностей митохондриальных микроРНК, отличающих их от остальных микроРНК, мы провели классификацию этих последовательностей с помощью метода случайного леса. Проведенный анализ впервые выявил достоверные различия между митомиРами и микроРНК по следующим характеристикам (по убыванию степени их важности в классификации): митомиРы имеют достоверно больший эволюционный возраст (низкий индекс филостратиграфического возраста, PAI), большее количество мишеней и ассоциаций с болезнями, в том числе митохондриальными (двусторонний точный тест Фишера, средние *p*-значения 1.82×10^{-89}/1.13×10^{-96} для всех мРНК/болезней и 6.01×10^{-22}/1.09×10^{-9} для митохондриальных); принадлежат к классу «циркулирующих» (среднее *p*-значение 1.20×10^{-56}). Обнаруженные различия между митомиРами и остальными микроРНК могут помочь раскрыть способ доставки микроРНК в митохондрии, свидетельствуют об эволюционной консервативности и важности митомиРов в регулировании функций и метаболизма митохондрий, а в целом говорят о том, что митомиРы не являются случайными микроРНК. Информация о 1312 экспериментально подтвержденных последовательностях митомиРов для трех организмов (*Homo sapiens*, *Mus musculus* и *Rattus norvegicus*) собрана в базе mitomiRdb (https://mitomiRdb.org).

Ключевые слова: митомиР; митохондрия; микроРНК; эволюция; база данных.

Для цитирования: Ворожейкин П.С., Титов И.И. Компьютерный анализ показывает отличия митохондриальных микроРНК от остальных микроРНК. *Вавиловский журнал генетики и селекции*. 2024;28(8):834-842. doi 10.18699/vjgb-24-91

Финансирование. Исследование выполнено за счет средств бюджетного проекта № FWNR-2022-0020.

Computer analysis shows differences between mitochondrial miRNAs and other miRNAs

P.S. Vorozheykin[1], I.I. Titov[1, 2, 3]

[1] Novosibirsk State University, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
pavel.vorozheykin@gmail.com

Abstract. A subclass of miRNAs with as yet unknown specific functions is mitomiRs – mitochondrial miRNAs that are mainly derived from nuclear DNA and are imported into mitochondria; moreover, changes in the expression levels of mitomiRs are associated with some diseases. To identify the most pronounced characteristics of mitochondrial miRNAs that distinguish them from other miRNAs, we classified mitomiR sequences using the Random Forest algorithm. The analysis revealed, for the first time, a significant difference between mitomiRs and other microRNAs by the following criteria (in descending order of importance in the classification): mitomiRs are evolutionarily older (have a lower phylostratigraphic age index, PAI); have more targets and disease associations, including mitochondrial ones (two-sided Fisher's exact test, average *p*-values 1.82×10^{-89}/1.13×10^{-96} for all mRNA/diseases and 6.01×10^{-22}/1.09×10^{-9} for mitochondria, respectively); and are in the class of "circulating" miRNAs (average *p*-value 1.20×10^{-56}). The identified differences between mitomiRs and other miRNAs may help uncover the mode of miRNA delivery into mitochondria, indicate the evolutionary conservation and importance of mitomiRs in the regulation of mitochondrial function and metabolism, and generally show that mitomiRs are not randomly encountered

© Ворожейкин П.С., Титов И.И., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

miRNAs. Information on 1,312 experimentally validated mitomiR sequences for three organisms (*Homo sapiens*, *Mus musculus* and *Rattus norvegicus*) is collected in the mitomiRdb database (https://mitomiRdb.org).
Key words: mitomiR; mitochondria; miRNA; evolution; database.

For citation: Vorozheykin P.S., Titov I.I. Computer analysis shows differences between mitochondrial miRNAs and other miRNAs. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):834-842. doi 10.18699/vjgb-24-91

Введение

Митохондрии участвуют в обширных двунаправленных межклеточных взаимодействиях, регулируя как свой собственный протеом, так и общую клеточную приспособленность и здоровье организма. На сегодняшний день хорошо известно, что основные пути биогенеза микроРНК начинаются в ядре и заканчиваются в цитоплазме (Bartel, 2018; Salim et al., 2022; Ziętara et al., 2023). Однако есть наблюдения, что эти малые некодирующие РНК присутствуют в органеллах, в том числе в митохондриях (Lung et al., 2006; Kren et al., 2009). Во многих случаях обнаруженные митохондриальные микроРНК (так называемые митомиРы, mitomiRs) более представлены в митохондриях по сравнению с цитоплазмой. Все это предполагает наличие механизмов транспортировки ядерных микроРНК в митохондрии или существование собственного процесса созревания микроРНК непосредственно внутри самих митохондрий.

В пользу существования механизмов транспортировки свидетельствует обнаружение так называемых циркулирующих микроРНК (Pozniak et al., 2022). Несколько фактов говорят в пользу альтернативы: во-первых, в митохондриях обнаружены белковые комплексы (AGO2 и Dicer), участвующие в каноническом пути биогенеза микроРНК (Bandiera et al., 2011; Wang W.-X. et al., 2015); во-вторых, экспрессия митохондриальных генов может регулироваться с помощью митохондриальных микроРНК, и такая регуляция может проявляться в митохондриально-ассоциированных болезнях (Li et al., 2012; Tomasetti et al., 2014; Zhang et al., 2014; Lin, Chu, 2021; Erturk et al., 2022; Gohel, Singh, 2022), а состав miRISC-комплекса (комплекса молчания генов, индуцируемого микроРНК) различается на разных стадиях развития организма. Перечисленное допускает возможность присутствия специфичного происхождения и биогенеза митомиРов, а также выполнение пока неизвестных функций внутри митохондрий, что делает митохондриальные микроРНК новым подклассом микроРНК, важным для научных исследований. Однако специальные функции и особые пути биогенеза митомиРов до сих пор не установлены, и поэтому остается открытым вопрос, не являются ли митомиРы обычными микроРНК, которые наблюдались в митохондриях по случайным причинам.

С целью определения факторов, специфичных для этого нового класса микроРНК, нами выполнен анализ микроРНК с помощью метода случайного леса, в рамках которого выявлены критерии, играющие наибольшую роль в классификации митомиРов (по убыванию степени их важности в классификации): индекс филостратиграфического возраста (PAI) микроРНК, наличие у микроРНК мишеней и принадлежность микроРНК к классу «циркулирующих». На основе полученных данных сделаны выводы о возрасте митомиРов, их возможном появлении в митохондриях и об их значимости для функционирования организма.

Анализируемые митомиРы собраны в базу данных mitomiRdb (https://mitomiRdb.org), в которой объединены вручную выверенные экспериментально подтвержденные митохондриальные микроРНК. В базе представлена информация о митомиРах для трех млекопитающих: *Homo sapiens*, *Mus musculus* и *Rattus norvegicus*. Всего аннотировано 1312 последовательностей со сведениями об их идентификаторах, нуклеотидных последовательностях, вторичных структурах предшественников. Также в базу включены ссылки на публикации с подтверждающими экспериментами и отражены экспериментально установленные связи этого класса малых некодирующих РНК с мРНК и болезнями (в том числе с митохондриальными). Все данные представлены в свободном доступе и могут быть загружены для дальнейшей автоматизированной компьютерной обработки.

Материалы и методы

В работе используются последовательности микроРНК из базы данных miRBase (https://miRBase.org, релизы с 10 по 22.1) (Kozomara et al., 2019). В последнем релизе базы аннотированы 48885 последовательностей микроРНК для 285 организмов. Общее количество микроРНК для трех рассматриваемых организмов (*H. sapiens*, *M. musculus* и *R. norvegicus*) составляет 5398 последовательностей, из них 2274 отмечены создателями базы как микроРНК с высокой степенью достоверности (микроРНК, картированные риды которых соответствуют шаблону при процессинге пре-миРНК комплексами Drosha/Dicer в классическом пути биогенеза).

Для изучения связи митомиРов с мРНК используется база экспериментально установленных мишеней микроРНК miRTarBase (https://mirtarbase.cuhk.edu.cn, релиз 8.0) (Huang et al., 2020) – собранная вручную по научным статьям база взаимодействий микроРНК-мишень с экспериментальными доказательствами прямых взаимодействий. Общее количество записей о взаимодействиях микроРНК и мРНК для человека, мыши и крысы составляет 553118, из которых 13311 записей соответствуют убедительным («сильным») экспериментам, а остальные 539807 записей отмечены как имеющие «слабые» подтверждения.

Набор экспериментально подтвержденных связей микроРНК с заболеваниями получен из базы данных RNADisease (http://www.rnadisease.org, релиз 4.0, раздел Experimental Data, записи с информацией о микроРНК-ассоциированных заболеваниях) (Chen et al., 2022). Каждая такая связь получена путем ручного отбора данных из публикаций, с особым вниманием к экспериментальному

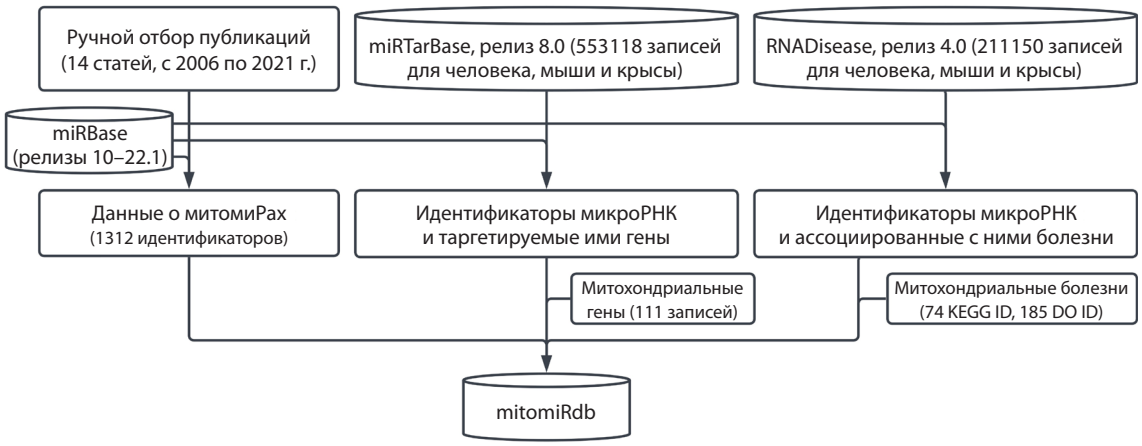


Рис. 1. Общая схема процесса формирования базы данных mitomiRdb.

доказательству роли микроРНК в регулировании и патогенезе болезней и к анализу комплементарного связывания мРНК, принимающих участие в течении болезни, с микроРНК. Общее количество записей в базе по трем рассматриваемым организмам (человек, мышь и крыса) составляет 211 150.

Для поиска локализации митомиРов в митохондриальных геномах и формирования списка митохондриальных генов использовались следующие эталонные последовательности мтДНК: *H. sapiens* (NC_012920.1), *M. musculus* (NC_005089.1) и *R. norvegicus* (NC_001665.2) (Sayers et al., 2022). Также для изучения эволюции сайтов связывания микроРНК let-7a-5p были взяты митохондриальные геномы приматов: *Gorilla gorilla* (NC_001645.1), *Pan paniscus* (NC_001644.1), *Pongo pygmaues* (NC_001646.1), *Pan troglodytes* (NC_001643.1) и *Symphalangus syndactylus* (NC_014047.1) (Sayers et al., 2022).

В расчете индекса филостратиграфического возраста (phylostratigraphic age index, PAI) применяются таксономические группы, полученные с NCBI-сервера (https://ftp.ncbi.nlm.nih.gov/pub/taxonomy/new_taxdump, данные на 12.07.2022) (Sayers et al., 2022). Для каждой последовательности микроРНК из 285 организмов вычислены все гомологичные микроРНК и тем самым определено распространение схожих микроРНК по организмам. Две микроРНК считаются гомологичными, если расстояние Хэмминга для их глобально выровненных нуклеотидных последовательностей составляет менее 10 % от длины полученного выравнивания. Параметры выравнивания: вес замены (совпадения) 5.0, вес замены (несовпадения) –4.0, вес для первой вставки/удаления –10.0, вес для последующих вставок/удалений –0.5. Чтобы рассчитать величину PAI каждой микроРНК, был задействован набор организмов, в которых обнаружены гомологи этой микроРНК. Для отобранных организмов устанавливается наиболее общий таксон согласно таксономическим группам. Порядковый номер этого наиболее общего таксона и будет значением PAI (нумерация с нуля) (Mustafin et al., 2019).

DO- и KEGG-идентификаторы, а также названия заболеваний получены из проекта «Онтология болезней» (Schriml et al., 2022) и Киотской энциклопедии генов и геномов (Kanehisa et al., 2017). С помощью этой информации был сформирован список митохондриально-ассоциированных заболеваний: в качестве таковых взяты названия и идентификаторы заболеваний, входящих в супергруппу (и ее подгруппы) KEGG H01427 (Mitochondrial disease), и заболевания, в названиях которых упоминаются митохондрии. Общее количество полученных заболеваний – 74 записи KEGG и 185 записей DO (данные представлены в Приложении 1)[1].

Информация о циркулирующих (circulating) микроРНК человека, мыши и крысы взята из баз данных miRandola (релиз 02.2017, 606 микроРНК) и plasmiR (релиз от 17.06.2021, 251 микроРНК) (Russo et al., 2018; Tastsoglou et al., 2021). Этому классу внеклеточных микроРНК характерно присутствие в отслеживаемых количествах в кровеносной системе и других жидкостях организма. Объединенное количество циркулирующих микроРНК из двух баз – 628 шт. (590 микроРНК человека, 18 – мыши, 20 – крысы).

Схематическое описание процесса получения информации о митомиРах, а также процесса формирования связей митомиРов с мишенями и заболеваниями представлено на рис. 1. На первом этапе были отобраны статьи, в которых исследовались локализованные в митохондриях микроРНК и/или упоминается термин «митомиР/mitomiR». Из этих публикаций выбраны 14 (с 2006 по 2021 г.), в которых экспериментально подтверждено присутствие микроРНК в митохондриях трех млекопитающих (*H. sapiens*, *M. musculus*, *R. norvegicus*) для разных типов клеток и тканей (Lung et al., 2006; Kren et al., 2009; Bian et al., 2010; Bandiera et al., 2011; Barrey et al., 2011; Mercer et al., 2011; Das et al., 2012; Sripada et al., 2012; Dasgupta et al., 2015; Jagannathan et al., 2015; Wang W.-X. et al., 2015; Wang X. et al., 2017; Fan et al., 2019; Zheng et al., 2021). В указанных работах микроРНК (митомиРы) идентифицируются по их названию (например, hsa-miR-1), которое с течением времени может изменяться в базе miRBase. На основании истории аннотаций miRBase для каждого названия микроРНК определен соответствующий ему уникальный идентификатор (MIMAT-идентификатор) в базе miRBase, который позволяет однозначно идентифициро-

[1] Приложения 1–4 см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx27.xlsx

вать последовательность микроРНК (и митомиРа) как в прошлых, так и в будущих релизах базы.

В процессе сопоставления данных мы обнаружили, что некоторые ранее аннотированные митомиРы исключены из базы miRBase в более поздних релизах. Чтобы получить более полную информацию, мы расширили нашу выборку митомиРов, добавив в нее 40 исключенных микроРНК и их 41 предшественника из ранних аннотаций в базе miRBase (Приложение 2). В результате собраны идентификаторы, нуклеотидные последовательности, вторичные структуры предшественников и другая информация о 1312 митомиРах.

Для изучения характеристик рассматриваемых митохондриально-ассоциированных микроРНК дополнительно сформирована выборка из 4126 последовательностей так называемых не-митомиРов, которая включает в себя все микроРНК человека, мыши и крысы из базы данных miRBase (релиз 22.1), за исключением идентифицированных выше митомиРов.

На втором этапе, используя историю аннотаций микроРНК в базе miRBase, для названий микроРНК в каждой записи баз miRTarBase и RNADisease мы установили соответствующий уникальный идентификатор базы miRBase, позволяющий в дальнейшем однозначно ассоциировать микроРНК с таргетируемыми мРНК и заболеваниями. У некоторых записей не удалось определить такой идентификатор из-за некорректности данных в базах (например, названия микроРНК hsa-miR-b5539, hsa-miRPlus-C1100 отсутствуют в базе miRBase) или из-за неполноты приведенных данных (например, запись в базе содержит название предшественника микроРНК hsa-let-7a-1, по которому невозможно однозначно определить последовательность микроРНК из микроРНК-дуплекса). Исключив из рассмотрения такие записи, в итоге мы получили информацию о надежных связях между собранными митомиРами и мРНК/заболеваниями.

Наконец, имея в наличии набор ассоциаций митомиР-мРНК и митомиР-заболевание, мы зафиксировали связи митомиРов с известными митохондриальными генами и заболеваниями. В базе mitomiRdb отмечена информация о связях обнаруженных митомиРов со 111 известными митохондриальными генами, кодирующими рРНК, тРНК и белки, для трех рассматриваемых митохондриальных ДНК человека, мыши и крысы. Каждая запись базы данных RNADisease содержит название заболевания и один или несколько идентификаторов заболевания: идентификатор DO (Schriml et al., 2022), идентификатор MeSH (Sayers et al., 2022) и идентификатор KEGG (Kanehisa et al., 2017; Schriml et al., 2022). Для каждой последовательности митомиРа мы дополнительно отразили ее связь (или отсутствие связи) с подготовленным общим списком заболеваний, а также отдельно с митохондриальными болезнями (Приложение 3).

Чтобы определить и ранжировать наиболее выраженные характеристики митомиРов в сравнении со всеми остальными микроРНК, мы провели анализ микроРНК с помощью метода случайного леса (алгоритма машинного обучения Random Forest) (Breiman, 2001). В качестве критериев классификации выбраны четыре бинарных и один числовой. Бинарные критерии: 1) является ли последовательность «циркулирующей» микроРНК; 2) является ли микроРНК «достоверной» согласно критериям базы miRBase; 3) есть ли у микроРНК мишени; 4) ассоциирована ли микроРНК с заболеваниями. Числовой критерий – значение величины PAI микроРНК.

Метод случайного леса применяется 100 раз на выборках микроРНК, подготовленных специальным образом: каждая выборка состоит из всех последовательностей митомиРов и не-митомиРов, из которых удалены микроРНК из семейств гомологичных последовательностей за исключением одного (случайно выбранного) представителя. В каждом тесте метода в качестве проверочной выборки выступает случайно сгенерированная подвыборка (1/3 от общего числа последовательностей), оставшаяся часть используется для обучения модели. Статистические оценки и уровни значимости критериев усредняются по всем тестам.

Результаты

Статистика. Мы собрали информацию о 1312 митохондриальных микроРНК (митомиРах) млекопитающих *Homo sapiens*, *Mus musculus* и *Rattus norvegicus*. В аннотированные митомиРы были добавлены те последовательности, которые к настоящему времени исключены из базы miRBase по разным причинам: например, микроРНК hsa-miR-1974, hsa-miR-1977 и hsa-miR-1978 располагаются в области митохондриальных тРНК; характер ридов из экспериментов RNAseq для микроРНК hsa-miR-6723-5p не соответствует правилам аннотации, принятым в miRBase; микроРНК mmu-miR-2145 представляет собой фрагмент 5S рРНК; и другие записи, предположительно представляющие собой транскрипционный шум или имеющие признаки неканонических путей биогенеза созревания. Около 66.6 % (874 шт.) обнаруженных митомиРов являются микроРНК человека, все остальные принадлежат мыши (30.6 %, 401 шт.) и крысе (2.8 %, 37 шт.).

Сравнивая нуклеотидные последовательности и названия митомиРов попарно друг с другом, мы обнаружили 16 (из 37 возможных по минимальному числу митомиРов в *R. norvegicus*) консервативных митомиРов, т. е. тех, которые наблюдаются в митохондриях каждого из трех рассмотренных организмов. 30.6 % всех митомиРов описаны более чем в одной из рассматриваемых статей, что может свидетельствовать об их более высокой достоверности в качестве митохондриальных. Только девять последовательностей митомиРов (hsa-miR-1973, hsa-miR-1974, hsa-miR-1977, hsa-miR-1978, hsa-miR-4461, hsa-miR-4463, hsa-miR-4284, hsa-miR-4485-3p, mmu-miR-805) полностью картированы на митохондриальную ДНК (три – на тРНК и рРНК, два – на белок-кодирующие области, один – на D-петлю), из них лишь пять (hsa-miR-1973, hsa-miR-1974, hsa-miR-1977, hsa-miR-1978, hsa-miR-4485-3p) дополнительно подтверждены RT-PCR/RT-qPCR/qRT-PCR экспериментами или наблюдаются в митохондриях в превалирующем количестве по сравнению с цитоплазмой.

Критерии. Анализ характеристик митомиРов в сравнении с остальными микроРНК не дает возможности однозначно классифицировать этот класс ни по одному из критериев (рис. 2, *а*). Классификация митомиРов с помощью алгоритма случайного леса (Random Forest) по-

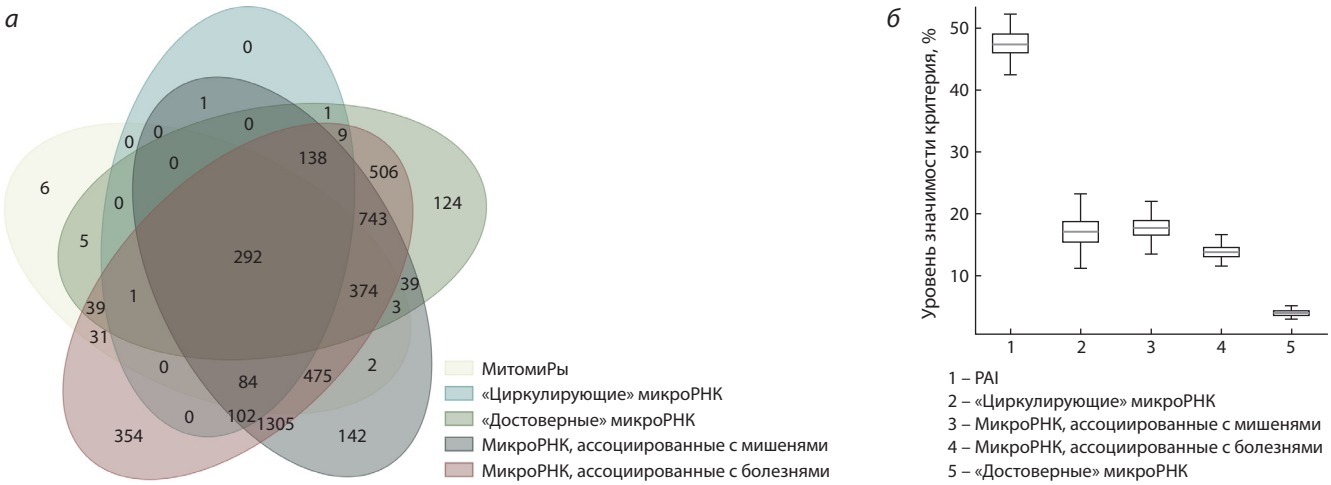


Рис. 2. Сравнение критериев классов микроРНК.

а – диаграмма Венна, охватывающая разные критерии последовательностей микроРНК: принадлежность классу митомиРов, «достоверность» микроРНК, «циркулируемость» микроРНК, наличие у микроРНК мишеней, ассоциация микроРНК с болезнями; *б* – уровни значимости критериев для предсказания последовательностей митомиРов с помощью алгоритма Random Forest в 100 тестах. Линией отмечено среднее значение уровня значимости критерия по выборке, длина «усов» – разность первого квартиля и полутора межквартильных расстояний и сумма третьего квартиля и полутора межквартильных расстояний.

зволяет ранжировать критерии по степени оказываемого влияния в классификации и тем самым определить наиболее важные характеристики митомиРов. С рассматриваемыми критериями метод демонстрирует усредненные ошибки предсказания на обучающей/тестовой выборках 0.20 ± 0.003 и 0.22 ± 0.006 соответственно. При этом наиболее значимыми критериями для классификации оказались (1) значение PAI, (2) наличие мишеней и (3) принадлежность микроРНК к классу циркулирующих (см. рис. 2, *б*). Менее значимые критерии – ассоциация микроРНК с болезнями и «достоверность» микроРНК; при этом «достоверность» микроРНК играет в классификации наименьшую роль.

Эволюционная характеристика PAI (phylostratigraphic age index – индекс филостратиграфического возраста), которая отражает возраст митомиРов, оказалась самым значимым критерием для классификации митомиРов. PAI принимает значения порядкового номера таксона (узла филостратиграфического дерева), наиболее удаленного от его корня и встречающегося в таксономических группах последовательности микроРНК и всех ее гомологов. Согласно значениям PAI, последовательности митомиРов в среднем демонстрируют большую эволюционную консервативность, чем не-митомиРы. Минимальное значение PAI у митомиРов равно 4 (рис. 3). Только четыре митомиРа (hsa-miR-99a-5p, mmu-miR-99a-5p, hsa-miR-100-5p и mmu-miR-100-5p) и два не-митомиРа (rno-miR-99a-5p, rno-miR-100-5p) имеют такое значение, поскольку их гомологи были обнаружены для организма *Nematostella vectensis* и свидетельствуют о древней истории происхождения этих микроРНК (Grimson et al., 2008).

Однако у эволюционно отдаленных видов гомология последовательностей микроРНК, даже при наличии соответствующей шпилечной структуры предшественника, не гарантирует существование настоящей микроРНК (Grimson et al., 2008). Все вышеперечисленные митомиРы относятся к большому семейству микроРНК mir-10. Это семейство содержит также митомиРы человека и мыши (miR-10a, miR-10b, miR-125b из 5p-ветви предшественников) с PAI = 5, тогда как микроРНК крысы из этого же семейства соответствуют не-митомиРам, что, вероятно, связано с недостаточной изученностью митохондрий крыс и небольшим количеством обнаруженных митомиРов для этого организма. Известно, что «семейные» микроРНК, такие как mir-10, обычно старше, более эффективны, нацелены на большее количество мРНК и с большей вероятностью связаны с заболеваниями. Все это указывает на важность митомиРов для функции и метаболизма митохондрий.

Следующие по важности критерии для классификации митомиРов – это факт обнаруженных мишеней и «циркулируемость» микроРНК. Количество записей о взаимодействии митомиРов с мРНК составляет 23 151 от общего числа для всех микроРНК, включая 3318 записей с «сильными» подтверждениями взаимодействий и 19 833 – со «слабыми» (Приложение 4). Надо отметить, что рассматриваемая часть miRTarBase не содержит записей с межвидовыми взаимодействиями, т. е. наблюдений, где организм микроРНК не совпадает с организмом таргетируемой мишени. Примечательно, что значительно меньшее количество митомиРов (по сравнению с количеством не-митомиРов) связано примерно с таким же количеством мРНК, что и у не-митомиРов. Двусторонний точный тест Фишера (среднее p-значение $1.82 \times 10^{-89} \pm 7.73 \times 10^{-89}$) демонстрирует статистически достоверную связь между видом микроРНК (митомиР/не-митомиР) и ассоциацией с мРНК (см. таблицу). Все перечисленное может свидетельствовать о важной регуляторной роли митохондриальных микроРНК.

Далее мы ограничились митохондриальными генами и их связью с митомиРами. Общее количество таких генов (кодирующих рРНК, тРНК и белки) для трех исследуемых мтДНК составляет 111. Особенностью базы данных miRTarBase является то, что она содержит информацию

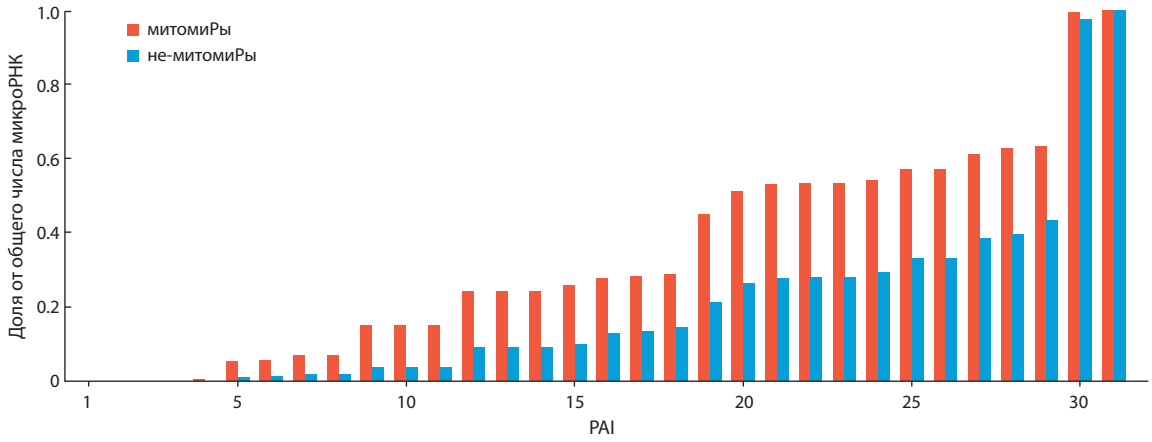


Рис. 3. Кумулятивное распределение последовательностей митомиРов и не-митомиРов по значениям PAI. Доля митомиРов с PAI < 16 превышает соответствующую долю остальных микроРНК с уровнем значимости $1.10 \times 10^{-41} \pm 5.97 \times 10^{-41}$.

Минимальное значение PAI, равное 4, соответствует четырем митомиРам (hsa-miR-99a-5p, mmu-miR-99a-5p, hsa-miR-100-5p, mmu-miR-100-5p) и двум не-митомиРам (rno-miR-99a-5p , rno-miR-100-5p) из широко распространенного семейства микроРНК mir-10.

Характеристики, по которым наблюдаются достоверные различия между микроРНК и митомиРами

Характеристика	митомиРы	Остальные микроРНК	*p*-значение по 100 тестам
Общее число	1312	4126	–
Доля микроРНК, которые таргетируют мРНК (miRTarBase)	0.94	0.60	$1.82 \times 10^{-89} \pm 7.73 \times 10^{-89}$
Доля микроРНК, которые таргетируют митохондриальные мРНК (miRTarBase)	0.05	0.0002	$1.20 \times 10^{-56} \pm 7.44 \times 10^{-56}$
Доля микроРНК, которые ассоциированы с болезнями (RNADisease)	0.99	0.77	$1.13 \times 10^{-96} \pm 3.89 \times 10^{-96}$
Доля микроРНК, которые ассоциированы с митохондриальными болезнями (RNADisease)	0.03	0.002	$1.09 \times 10^{-9} \pm 2.14 \times 10^{-9}$
Доля микроРНК, которые являются циркулирующими	0.29	0.06	$1.20 \times 10^{-56} \pm 7.44 \times 10^{-56}$
Доля микроРНК с PAI < 16	0.26	0.01	$1.10 \times 10^{-41} \pm 5.97 \times 10^{-41}$

Примечание. Оценка значимости различия характеристик проводилась при усреднении по 100 реализациям при случайном выборе одной последовательности из гомологичных.

только о белок-кодирующих митохондриальных генах и не содержит об РНК-кодирующих. Установлено, что всего 65 митомиРов таргетируют митохондриальные мРНК, а 1247 – нет. При этом последовательности всех таргетирующих митомиРов не картируются на митохондриальную ДНК, т. е. эти митомиРы – внешние по отношению к митохондриям. МитомиРы таргетируют 12 митохондриальных мРНК: *ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5*, *ND6*, *COX1*, *COX2*, *COX3*, *CYTB*, *ATP6*. Наибольшее количество митомиРов (>15) нацелено только на две мРНК человека (*ATP6*, *COX1*), наименьшее количество митомиРов (<5) связано с мРНК человека *ND3*/*ND4L* и мРНК мыши и крысы *COX1*. В то же время всего один не-митомиР hsa-miR-15a-3p из 4126 не-митомиРов таргетирует митохондриальную мРНК *ND4L*. Двусторонний точный тест Фишера (среднее *p*-значение $6.01 \times 10^{-22} \pm 2.30 \times 10^{-21}$) демонстрирует статистически достоверную связь между видом микроРНК (митомиР/не-митомиР) и ассоциацией с митохондриальными мРНК.

Значимость критерия ассоциации митомиРов с классом «циркулирующих» микроРНК может отражать особый способ транспортировки митомиРов в митохондрии. Сравнение выборок митомиРов и не-митомиРов (см. рис. 2, *а*) показывает, что среди циркулирующих микроРНК митомиРы более представлены, чем не-митомиРы (всего 377 митомиРов против 251 не-митомиРа). Точный тест Фишера на наборах микроРНК, «очищенных» от гомологичных последовательностей, демонстрирует среднее *p*-значение $1.20 \times 10^{-56} \pm 7.44 \times 10^{-56}$.

Хотя ассоциация с заболеваниями демонстрирует более низкую значимость критерия для классификации, чем упомянутые выше (из-за обнаруженного сходства с критерием наличия мишеней, см. таблицу и Приложение 4), этот критерий может быть важным для общего понимания функций митомиРов. Для каждой последовательности митомиРа и не-митомиРа мы установили связь (или отсутствие связи) с заболеваниями, в том числе митохондриальными. Используя данные базы RNADisease мы обнаружили всего 36 (из 1312) митомиРов, связанных с митохондриальными заболеваниями (по их названиям или идентификаторам). С другой стороны, только 9 (из 4126) не-митомиРов имеют такую же ассоциацию, возможно, обусловленную таргети-

рованием не-митомиРами ядерных мРНК с митохондриальной локализацией продукта. В обоих случаях выявлены ассоциации с группой заболеваний “Mitochondrial disease” и с MNGIE-синдромом (см. Приложение 3). Двусторонний точный тест Фишера (средние p-значения $1.13 \times 10^{-96} \pm \pm 3.89 \times 10^{-96} / 1.09 \times 10^{-9} \pm 2.14 \times 10^{-9}$ для всех и для митохондриальных заболеваний соответственно) показывает, что митомиРы более тесно связаны с заболеваниями, чем не-митомиРы. Ассоциации с митохондриальными мРНК и болезнями могут свидетельствовать о важной роли митомиРов в активности митохондрий.

Данные. База mitomiRdb (https://mitomiRdb.org) предоставляет пользовательский веб-интерфейс для доступа к данным митомиРов и выполнения поиска информации. mitomiRdb включает следующие сведения (в соответствии с базой miRBase): уникальный идентификатор (MIMAT), название, нуклеотидную последовательность и организм наблюдения микроРНК в митохондриях. Кроме того, для записи отображается флаг достоверности митомиРа в случае, если он ассоциирован с митохондриальными мРНК/заболеваниями или картирован на митохондриальный геном. В качестве дополнительной информации представлена вторичная структура предшественников микроРНК, ссылки на подтверждающие публикации, список ассоциированных болезней и мРНК. Для «достоверных» митомиРов приведен список ассоциированных митохондриальных мРНК/заболеваний и отмечен комментарием факт присутствия последовательности митомиРа в мтДНК. Все размещенные на сайте данные доступны для скачивания в формате SQLite для дальнейшей компьютерной обработки (doi.org/10.6084/m9.figshare.22592380).

Обсуждение

К настоящему времени в митохондриях обнаружено большое число микроРНК. При этом до сих пор неизвестно, обусловлено ли присутствие микроРНК в митохондриях выполняемыми ими функциями или же наблюдались случайные микроРНК. В первом случае такие митохондриальные микроРНК могут иметь особенности своего биогенеза и регуляции экспрессии генов, в частности митохондриальных. В рамках данной работы мы проанализировали характеристики микроРНК с целью определения факторов, отличающих митомиРы от остальных микроРНК, и подтверждения факта неслучайности наблюдения этого класса микроРНК.

Самым значимым для митомиРов оказалось значение индекса филостратиграфического возраста – величины, характеризующей эволюционный возраст последовательности. Меньшие значения PAI у митомиРов свидетельствуют о том, что в среднем митомиРы более старые и, как большинство старых микроРНК, чаще задействованы в большем числе важных регуляторных процессов, в том числе включающих участие митохондрий.

Выявлена также значимая связь митомиРов с мРНК (включая митохондриальные) на основе информации об экспериментально установленных взаимодействиях микроРНК с мишенями. Это может свидетельствовать о неслучайности наблюдения митомиРов внутри митохондрий и о важности митомиРов для функционирования организма и митохондрий в частности. Взаимодействие микроРНК-мРНК не обязательно влечет за собой молчание генов. Около половины экспериментально идентифицированных взаимодействий микроРНК-мРНК приходятся на те, в которых отсутствует непрерывная комплементарность нуклеотидов в сид-регионе (позиции 2–5) микроРНК (Grosswendt et al., 2014) и которые наиболее важны для ассоциации с мишенью (Chandradoss et al., 2015; Salomon et al., 2015). Неканонические сайты связывания, хотя и идентифицируются CLIP-методами, не всегда влияют на экспрессию генов (Agarwal et al., 2015). Полезным свидетельством функциональности сайта будет его эволюционное сохранение. Чтобы проверить предложенный аргумент на одном примере митомиРа, мы выбрали единственный сайт связывания, для которого экспериментальные предсказания соответствуют компьютерному (Khorsandi et al., 2018). Сайт находится в позиции 13418–13439 гена *mt-ND5* мтДНК человека, недалеко от нее (1081 п. н. от начала *ND5*) у других приматов и связывает микроРНК hsa-let-7a. Этот сайт возникает у человека благодаря синонимичной замене C>T в позиции, которая соответствует второму нуклеотиду сид-региона микроРНК let-7a. Между тем в рассматриваемой позиции наблюдается обратный SNP T>C (rs386829181, частота аллеля 7×10^{-4}) (Sherry, 2001), который тесно связан с краниальными менингиомами у китайских пациентов.

Следующим фактором, выделяющим митомиРы среди остальных микроРНК, является их принадлежность к классу циркулирующих. Циркулирующие микроРНК относятся к классу внеклеточных РНК, наблюдаемых в достаточных количествах в разных жидкостях организма. Важность данного фактора для классификации митомиРов может свидетельствовать о возможном сходстве между механизмами свободного перемещения микроРНК из клетки и транспортировкой митомиРов внутрь митохондрий.

Хотя критерий, основанный на связи митомиРов с болезнями, выглядит менее важным для классификации вследствие совпадения ассоциаций как митомиРов, так и остальных микроРНК с мишенями и болезнями (см. Приложение 4), статистические критерии показывают значимую связь между болезнями и митомиРами, что может говорить о важной роли митомиРов в регулировании процессов организма и митохондрий в частности.

Наименее важный из рассмотренных факторов определения митомиРов – критерий «достоверности» (согласно авторам базы miRBase). Это может указывать на наличие у митомиРов неизвестного пути созревания, при котором формируется дуплекс микроРНК-микроРНК с неканоническими свисающими концами, в отличие от канонических двунуклеотидных (которые используются в качестве одного из критериев «достоверности») при редактировании предшественника микроРНК комплексами Drosha и Dicer.

Несмотря на присутствие митохондрий во всех типах клеток исследуемых млекопитающих, вклад факторов тканеспецифичности и уровня экспрессии микроРНК в различие митомиРов и не-митомиРов пока невозможно оценить ввиду того, что существующие экспериментальные наблюдения митомиРов охватывают незначительное количество тканей и содержат недостаточно информации об экспрессии митомиРов.

Заключение

В рамках работы выделены следующие характеристики митохондриальных микроРНК, которые отличают их от остальных микроРНК, по убыванию степени их важности: индекс филостратиграфического возраста (PAI), наличие у микроРНК мишеней и принадлежность микроРНК к классу «циркулирующих». Выявленные характеристики могут помочь пролить свет на особенности происхождения, процессинга и функции митомиРов.

Все экспериментально исследованные митомиРы собраны в базе данных mitomiRdb (https://mitomiRdb.org). Представленная база данных может быть полезна для более полного изучения микроРНК и их подкласса – митомиРов.

Список литературы / References

Agarwal V., Bell G.W., Nam J.-W., Bartel D.P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*. 2015;4: e05005. doi 10.7554/eLife.05005

Bandiera S., Rüberg S., Girard M., Cagnard N., Hanein S., Chrétien D., Munnich A., Lyonnet S., Henrion-Caude A. Nuclear outsourcing of RNA interference components to human mitochondria. *PLoS One*. 2011;6(6):e20746. doi 10.1371/journal.pone.0020746

Barrey E., Saint-Auret G., Bonnamy B., Damas D., Boyer O., Gidrol X. Pre-microRNA and mature microRNA in human mitochondria. *PLoS One*. 2011;6(5):e20220. doi 10.1371/journal.pone.0020220

Bartel D.P. Metazoan microRNAs. *Cell*. 2018;173(1):20-51. doi 10.1016/j.cell.2018.03.006

Bian Z., Li L.-M., Tang R., Hou D.-X., Chen X., Zhang C.-Y., Zen K. Identification of mouse liver mitochondria-associated miRNAs and their potential biological functions. *Cell Res.* 2010;20(9):1076-1078. doi 10.1038/cr.2010.119

Breiman L. Random forests. *Mach. Learn.* 2001;45(1):5-32. doi 10.1023/A:1010933404324

Chandradoss S.D., Schirle N.T., Szczepaniak M., MacRae I.J., Joo C. A dynamic search process underlies microRNA targeting. *Cell*. 2015;162(1):96-107. doi 10.1016/j.cell.2015.06.032

Chen J., Lin J., Hu Y., Ye M., Yao L., Wu L., Zhang W., Wang M., Deng T., Guo F., Huang Y., Zhu B., Wang D. RNADisease v4.0: an updated resource of RNA-associated diseases, providing RNA-disease analysis, enrichment and prediction. *Nucleic Acids Res.* 2022;51(D1):D1397-D1404. doi 10.1093/nar/gkac814

Das S., Ferlito M., Kent O.A., Fox-Talbot K., Wang R., Liu D., Raghavachari N., Yang Y., Wheelan S.J., Murphy E., Steenbergen C. Nuclear miRNA regulates the mitochondrial genome in the heart. *Circ. Res.* 2012;110(12):1596-1603. doi 10.1161/CIRCRESAHA.112.267732

Dasgupta N., Peng Y., Tan Z., Ciraolo G., Wang D., Li R. miRNAs in mtDNA-less cell mitochondria. *Cell Death Discov.* 2015;1(1): 15004. doi 10.1038/cddiscovery.2015.4

Erturk E., Enes Onur O., Akgun O., Tuna G., Yildiz Y., Ari F. Mitochondrial miRNAs (mitomiRs): their potential roles in breast and other cancers. *Mitochondrion*. 2022;66:74-81. doi 10.1016/j.mito.2022.08.002

Fan S., Tian T., Chen W., Lv X., Lei X., Zhang H., Sun S., Cai L., Pan G., He L., Ou Z., Lin X., Wang X., Perez M.F., Tu Z., Ferrone S., Tannous B.A., Li J. Mitochondrial miRNA determines chemoresistance by reprogramming metabolism and regulating mitochondrial transcription. *Cancer Res.* 2019;79(6):1069-1084. doi 10.1158/0008-5472.CAN-18-2505

Gohel D., Singh R. Different platforms for mitomiRs in mitochondria: emerging facets in regulation of mitochondrial functions. *Mitochondrion*. 2022;66:67-73. doi 10.1016/j.mito.2022.08.003

Grimson A., Srivastava M., Fahey B., Woodcroft B.J., Chiang H.R., King N., Degnan B.M., Rokhsar D.S., Bartel D.P. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*. 2008;455(7217):1193-1197. doi 10.1038/nature07415

Grosswendt S., Filipchyk A., Manzano M., Klironomos F., Schilling M., Herzog M., Gottwein E., Rajewsky N. Unambiguous identification of miRNA: target site interactions by different types of ligation reactions. *Mol. Cell*. 2014;54(6):1042-1054. doi 10.1016/j.molcel.2014.03.049

Huang H.-Y., Lin Y.-C.-D., Li J., Huang K.-Y., Shrestha S., Hong H.-C., Tang Y., Chen Y.-G., Jin C.-N., Yu Y., Xu J.-T., Li Y.-M., Cai X.-X., Zhou Z.-Y., Chen X.-H., Pei Y.-Y., Hu L., Su J.-J., Cui S.-D., Wang F., Xie Y.-Y., Ding S.-Y., Luo M.-F., Chou C.-H., Chang N.-W., Chen K.-W., Cheng Y.-H., Wan X.-H., Hsu W.-L., Lee T.-Y., Wei F.-X., Huang H.-D. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* 2020;48(D1):D148-D154. doi 10.1093/nar/gkz896

Jagannathan R., Thapa D., Nichols C.E., Shepherd D.L., Stricker J.C., Croston T.L., Baseler W.A., Lewis S.E., Martinez I., Hollander J.M. Translational regulation of the mitochondrial genome following redistribution of mitochondrial microRNA in the diabetic heart. *Circ. Cardiovasc. Genet.* 2015;8(6):785-802. doi 10.1161/CIRCGENETICS.115.001067

Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353-D361. doi 10.1093/nar/gkw1092

Khorsandi S.E., Salehi S., Cortes M., Vilca-Melendez H., Menon K., Srinivasan P., Prachalias A., Jassem W., Heaton N. An in silico argument for mitochondrial microRNA as a determinant of primary non function in liver transplantation. *Sci. Rep.* 2018;8(1):3105. doi 10.1038/s41598-018-21091-9

Kozomara A., Birgaoanu M., Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019;47(D1): D155-D162. doi 10.1093/nar/gky1141

Kren B.T., Wong P.Y.-P., Sarver A., Zhang X., Zeng Y., Steer C.J. MicroRNAs identified in highly purified liver-derived mitochondria may play a role in apoptosis. *RNA Biol.* 2009;6(1):65-72. doi 10.4161/rna.6.1.7534

Li P., Jiao J., Gao G., Prabhakar B.S. Control of mitochondrial activity by miRNAs. *J. Cell. Biochem.* 2012;113(4):1104-1110. doi 10.1002/jcb.24004

Lin H.-Y., Chu P.-Y. Advances in understanding mitochondrial microRNAs (mitomiRs) on the pathogenesis of triple-negative breast cancer (TNBC). *Oxid. Med. Cell. Longev.* 2021;2021:5517777. doi 10.1155/2021/5517777

Lung B., Zemann A., Madej M.J., Schuelke M., Techritz S., Ruf S., Bock R., Hüttenhofer A. Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nucleic Acids Res.* 2006; 34(14):3842-3852. doi 10.1093/nar/gkl448

Mercer T.R., Neph S., Dinger M.E., Crawford J., Smith M.A., Shearwood A.-M.J., Haugen E., Bracken C.P., Rackham O., Stamatoyannopoulos J.A., Filipovska A., Mattick J.S. The human mitochondrial transcriptome. *Cell*. 2011;146(4):645-658. doi 10.1016/j.cell.2011.06.051

Mustafin Z.S., Zamyatin V.I., Konstantinov D.K., Doroshkov A.V., Lashin S.A., Afonnikov D.A. Phylostratigraphic analysis shows the earliest origination of the abiotic stress associated genes in *A. thaliana*. *Genes*. 2019;10(12):963. doi 10.3390/genes10120963

Pozniak T., Shcharbin D., Bryszewska M. Circulating microRNAs in medicine. *Int. J. Mol. Sci.* 2022;23(7):3996. doi 10.3390/ijms23073996

Russo F., Di Bella S., Vannini F., Berti G., Scoyni F., Cook H.V., Santos A., Nigita G., Bonnici V., Laganà A., Geraci F., Pulvirenti A., Giugno R., De Masi F., Belling K., Jensen L.J., Brunak S., Pellegrini M., Ferro A. miRandola 2017: a curated knowledge base of non-invasive biomarkers. *Nucleic Acids Res.* 2018;46(D1):D354-D359. doi 10.1093/nar/gkx854

Salim U., Kumar A., Kulshreshtha R., Vivekanandan P. Biogenesis, characterization, and functions of mirtrons. *WIREs RNA*. 2022; 13(1):e1680. doi 10.1002/wrna.1680

Salomon W.E., Jolly S.M., Moore M.J., Zamore P.D., Serebrov V. Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. *Cell*. 2015;162(1):84-95. doi 10.1016/j.cell.2015.06.029

Sayers E.W., Bolton E.E., Brister J.R., Canese K., Chan J., Comeau D.C., Connor R., Funk K., Kelly C., Kim S., Madej T., Marchler-Bauer A., Lanczycki C., Lathrop S., Lu Z., Thibaud-Nissen F., Murphy T., Phan L., Skripchenko Y., Tse T., Wang J., Williams R., Trawick B.W., Pruitt K.D., Sherry S.T. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2022;50(D1):D20-D26. doi 10.1093/nar/gkab1112

Schriml L.M., Munro J.B., Schor M., Olley D., McCracken C., Felix V., Baron J.A., Jackson R., Bello S.M., Bearer C., Lichenstein R., Bisordi K., Dialo N.C., Giglio M., Greene C. The Human Disease Ontology 2022 update. *Nucleic Acids Res.* 2022;50(D1):D1255-D1261. doi 10.1093/nar/gkab1063

Sherry S.T. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308-311. doi 10.1093/nar/29.1.308

Sripada L., Tomar D., Prajapati P., Singh Rochika, Singh A.K., Singh Rajesh. Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. *PLoS One*. 2012;7(9):e44873. doi 10.1371/journal.pone.0044873

Tastsoglou S., Miliotis M., Kavakiotis I., Alexiou A., Gkotsi E.C., Lambropoulou A., Lygnos V., Kotsira V., Maroulis V., Zisis D., Skoufos G., Hatzigeorgiou A.G. PlasmiR: a manual collection of circulating microRNAs of prognostic and diagnostic value. *Cancers*. 2021;13(15):3680. doi 10.3390/cancers13153680

Tomasetti M., Santarelli L., Neuzil J., Dong L. MicroRNA regulation of cancer metabolism: role in tumour suppression. *Mitochondrion*. 2014;19:29-38. doi 10.1016/j.mito.2014.06.004

Wang W.-X., Visavadiya N.P., Pandya J.D., Nelson P.T., Sullivan P.G., Springer J.E. Mitochondria-associated microRNAs in rat hippocampus following traumatic brain injury. *Exp. Neurol.* 2015;265:84-93. doi 10.1016/j.expneurol.2014.12.018

Wang X., Song C., Zhou X., Han X., Li J., Wang Z., Shang H., Liu Y., Cao H. Mitochondria associated microRNA expression profiling of heart failure. *BioMed Res. Int.* 2017;2017:4042509. doi 10.1155/2017/4042509

Zhang X., Zuo X., Yang B., Li Z., Xue Y., Zhou Y., Huang J., Zhao X., Zhou J., Yan Y., Zhang H., Guo P., Sun H., Guo L., Zhang Y., Fu X.-D. MicroRNA directly enhances mitochondrial translation during muscle differentiation. *Cell*. 2014;158(3):607-619. doi 10.1016/j.cell.2014.05.047

Zheng H., Liu J., Yu J., McAlinden A. Expression profiling of mitochondria-associated microRNAs during osteogenic differentiation of human MSCs. *Bone*. 2021;151:116058. doi 10.1016/j.bone.2021.116058

Ziętara K.J., Lejman J., Wojciechowska K., Lejman M. The importance of selected dysregulated microRNAs in diagnosis and prognosis of childhood B-cell precursor acute lymphoblastic leukemia. *Cancers*. 2023;15(2):428. doi 10.3390/cancers15020428

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 18.09.2024. После доработки 06.11.2024. Принята к публикации 07.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):843-853

doi 10.18699/vjgb-24-92

Новый подход к анализу эволюции SARS-CoV-2, основанный на визуализации и кластеризации больших объемов генетических данных, компактно представленных в оперативной памяти

А.Ю. Пальянов [1, 2, 3], Н.В. Пальянова [2]

[1] Институт систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Научно-исследовательский институт вирусологии, Федеральный исследовательский центр фундаментальной и трансляционной медицины, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
palyanov@iis.nsk.su

Аннотация. Коронавирус SARS-CoV-2 – это вирус, для которого было собрано, секвенировано и сохранено рекордное количество вариантов генома из источников по всему миру. Нуклеотидные последовательности в формате FASTA включают 16.8 млн геномов, каждый длиной ≈29 900 нт (нуклеотидов), общим размером $\approx 500 \cdot 10^9$ нт, или 466 Гб. Мы предлагаем способ представления данных, позволяющий разместить без потерь всю эту информацию в оперативной памяти (RAM) обычного персонального компьютера. Более того, будет достаточно всего ≈330 Мб. Выравнивание их всех относительно исходной референсной последовательности Wunah-Hu-1 позволяет представить каждый геном как структуру данных, содержащую списки точечных мутаций, делеций и вставок. Наша реализация такого представления данных привела к коэффициенту сжатия 1:1500 (для сравнения, упаковка данных с помощью популярного архиватора WinRAR дает степень сжатия только 1:62) и обеспечила возможность быстрого вычисления редакционного расстояния между различными вариантами генома. С помощью этого подхода, реализованного в виде программы на C++, мы провели анализ различных свойств набора геномов SARS-CoV-2, содержащихся в NCBI Genbank, собранных за 4.5 года (с 24.12.2019 по 24.06.2024). Были рассчитаны распределение числа геномов от числа неопределенных нуклеотидов "N" в них, число уникальных геномов и кластеров из идентичных геномов, а также распределение кластеров по размеру (числу идентичных геномов) и продолжительности (длине временного интервала между первым и последним геномом каждого кластера). Наконец, эволюция распределений числа изменений (редакционное расстояние между каждым геномом и референсной последовательностью), вызванных заменами, делециями и вставками, была визуализирована в виде 3D поверхностей, наглядно изображающих процесс вирусной эволюции в течение 4.5 лет, с интервалом в одну неделю. Такая визуализация хорошо соотносится с филогенетическими деревьями (обычно рассчитываемыми по 3–4 тыс. представителей вариантов генома), но строится на основе миллионов геномов, отображает больше деталей и не зависит от типа классификации линий/клад.

Ключевые слова: коронавирус; SARS-CoV-2; геном; варианты; эволюция; программная система; большие данные; компактизация; анализ; визуализация.

Для цитирования: Пальянов А.Ю., Пальянова Н.В. Новый подход к анализу эволюции SARS-CoV-2, основанный на визуализации и кластеризации больших объемов генетических данных, компактно представленных в оперативной памяти. *Вавиловский журнал генетики и селекции*. 2024;28(8):843-853. doi 10.18699/vjgb-24-92

Финансирование. Исследование выполнено за счет гранта Российского научного фонда (проект № 23-64-00005).

A novel approach to analyzing the evolution of SARS-CoV-2 based on visualization and clustering of large genetic data compactly represented in operative memory

A.Yu. Palyanov [1, 2, 3], N.V. Palyanova [2]

[1] A.P. Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Research Institute of Virology, Federal Research Center of Fundamental and Translational Medicine, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
palyanov@iis.nsk.su

Abstract. SARS-CoV-2 is a virus for which an outstanding number of genome variants were collected, sequenced and stored from sources all around the world. Raw data in FASTA format include 16.8 million genomes, each ≈29,900 nt (nucleotides), with a total size of $\approx 500 \cdot 10^9$ nt, or 465 Gb. We suggest an approach to data representation and organization,

© Пальянов А.Ю., Пальянова Н.В., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

with which all this can be stored losslessly in the operative memory (RAM) of a common PC. Moreover, just ≈330 Mb will be enough. Aligning all genomes versus the initial Wuhan-Hu-1 reference sequence allows each to be represented as a data structure containing lists of point mutations, deletions and insertions. Our implementation of such data representation resulted in a 1:1500 compression ratio (for comparison, compression of the same data with the popular WinRAR archiver gives only 1:62) and fast access to genomes (and their metadata) and comparisons between different genome variants. With this approach implemented as a C++ program, we performed an analysis of various properties of the set of SARS-CoV-2 genomes available in NCBI Genbank (within a period from 24.12.2019 to 24.06.2024). We calculated the distribution of the number of genomes with undetermined nucleotides, 'N's, vs the number of such nucleotides in them, the number of unique genomes and clusters of identical genomes, and the distribution of clusters by size (the number of identical genomes) and duration (the time interval between each cluster's first and last genome). Finally, the evolution of distributions of the number of changes (editing distance between each genome and reference sequence) caused by substitutions, deletions and insertions was visualized as 3D surfaces, which clearly show the process of viral evolution over 4.5 years, with a time step = 1 week. It is in good correspondence with phylogenetic trees (usually based on 3–4 thousand of genome variant representatives), but is built over millions of genomes, shows more details and is independent of the type of lineage/clade classification.

Key words: coronavirus; SARS-CoV-2; genome; variants; evolution; software system; big data; compact representation of data; analysis; visualization.

For citation: Palyanov A.Yu., Palyanova N.V. A novel approach to analyzing the evolution of SARS-CoV-2 based on visualization and clustering of large genetic data compactly represented in operative memory. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):843-853. doi 10.18699/vjgb-24-92

Введение

Коронавирус SARS-CoV-2 (самый первый образец которого, названный Wuhan/Hu-1/2019, был получен 24 декабря 2019 г.) (Wu et al., 2020) вызвал крупнейшую пандемию за последние 100 лет (со времен испанского гриппа 1918–1920 гг.). Спустя 4.5 года он по-прежнему продолжает свое существование, эволюционирует и выявляется у людей по всему миру, хотя и уже далеко не в тех объемах, которые были во время пика пандемии, и не с такими тяжелыми последствиями. Впрочем, как правило, с приходом осени показатели числа заражений снова возрастают, и 2024 г. не исключение. По данным Всемирной ассоциации здравоохранения (https://data.who.int/dashboards/covid19/cases, раздел "COVID-19 cases, country level trends"), к середине сентября 2024 г. во многих странах уже начался рост заболеваемости. Так, например, в России за июль 2024 г. было зарегистрировано 26.7 тыс. случаев заражения SARS-CoV-2, за август – 24.7 тыс., а за первую половину сентября – уже 62.2 тыс. В разных странах наблюдаются те или иные особенности динамики количества заражений, зависящие от множества факторов, анализом взаимосвязей между которыми, в частности, занимаемся и мы (Palyanova et al., 2022, 2023).

Получаемые по всему миру образцы вируса SARS-CoV-2 секвенируют и вносят в базы данных, крупнейшими из которых являются GISAID (gisaid.org) и NCBI Genbank (www.ncbi.nlm.nih.gov/sars-cov-2/): по состоянию на 06.2024 в них содержится более $16.7 \cdot 10^6$ и более $8.6 \cdot 10^6$ образцов геномов SARS-CoV-2 соответственно. Для сравнения, вирус гриппа человека, самые ранние образцы которого в GISAID датируются 1905 г., за более чем столетний период представлен в ней примерно $5.22 \cdot 10^5$ геномами. Типичный размер генома SARS-CoV-2 составляет 29.9 тыс. нт, поэтому полный объем геномов этого вируса, содержащихся в GISAID, составляет около $500 \cdot 10^9$ нт (или 465 Гб), а в Genbank – около $258 \cdot 10^9$ нт (241 Гб). В объем оперативной памяти среднего современного ПК (16…64 Гб) все эти данные одновременно не поместятся, тогда как работа с ними непосредственно из файлов, расположенных на жестком диске (HDD) или твердотельном накопителе (SSD), будет происходить существенно медленнее, чем из оперативной памяти (RAM). Скорости чтения данных с современных HDD/SDD/RAM имеют характерные значения порядка 0.2, 3 и 50 Гб/с соответственно, так что при значительных объемах данных и вычислительных нагрузках работа именно с оперативной памятью крайне желательна.

Несмотря на вакцинацию и медикаментозное лечение, в настоящее время не существует способа полностью устранить SARS-CoV-2 (Cui et al., 2023), так что, по-видимому, он надолго останется с человечеством, пополнив многочисленный перечень возбудителей ОРВИ, насчитывающий более 200 наименований, включая грипп, респираторно-синцитиальную, риновирусную, коронавирусную, аденовирусную и другие инфекции, вызывающие катаральные воспаления дыхательных путей.

Чем дольше существует вирус, тем больше изменений накапливается в его геноме, каждая новая генерация получается в результате репликации вирусов предыдущего поколения, в процессе которой возможно возникновение ошибок/изменений. В результате генных мутаций могут происходить замены, делеции и вставки одного или нескольких нуклеотидов, транслокации, дупликации и инверсии различных частей гена. Так, например, точечные мутации возникают самопроизвольно с частотами 10^{-8}–10^{-6} для ДНК-вирусов и 10^{-6}–10^{-4} – для РНК-вирусов (Sanjuán, Domingo-Calap, 2016), у которых собственная молекулярная машина для репликации (РНК-полимераза) лишена корректирующего ошибки механизма (экзонуклеазы). Исключение составляют коронавирусы и торовирусы, у которых она все-таки имеется (Campagnola et al., 2022), поскольку они обладают одними из наибольших для РНК-вирусов геномами, слишком быстрое накопление ошибок в которых, по-видимому, не является желательным и не способствует выживанию вируса.

Частота возникновения ошибок при репликации SARS-CoV-2 составляет, согласно (Amicone et al., 2022), $1.3 \cdot 10^{-6} \pm 0.2 \cdot 10^{-6}$ замен на позицию за один инфекцион-

ный цикл заражения клетки (т. е. от входа вируса в нее до выхода новых вирионов наружу). При этом скорость эволюционных изменений в геноме SARS-CoV-2 оценивается как $8.9 \cdot 10^{-4}$ замен на позицию в год (Sonnleitner et al., 2022).

Помимо упомянутых механизмов, которые могут воздействовать на отдельный (одиночный) геном, имеются и такие, благодаря которым могут возникать новые комбинации на основе генетического материала разных вариантов генома. Если два разных варианта одного и того же вируса заражают один и тот же организм одновременно (например, заражение штаммами SARS-CoV-2 «Дельта» и «Омикрон» (Bolze et al., 2022)), у них появляется возможность взаимодействовать во время репликации (Simon-Loriere, Holmes, 2011), в результате чего могут возникать рекомбинанты или реассортанты (для вирусов с сегментированным геномом).

Безотносительно того, какой именно механизм вызвал то или иное изменение, для любой пары геномов рассматриваемого вируса может быть рассчитано расстояние Левенштейна (также называемое редакционным расстоянием или дистанцией редактирования), определяемое как минимальное количество односимвольных операций (замены, делеции, вставки), которые нужно внести в первый геном, чтобы получить из него второй (или во второй, чтобы получить первый, – результат получится тот же самый). Другими словами, расстояние Левенштейна задает метрику, определяющую разность между двумя последовательностями символов. Таким образом, каждый вариант генома SARS-CoV-2 из имеющихся миллионов можно сравнить с исходным референсным геномом Wuhan-Hu-1. Для этого необходимо осуществить глобальное выравнивание всех последовательностей относительно референса, что было выполнено нами с помощью программы NextAlign/NextClade (https://github.com/nextstrain/nextclade) (Aksamentov et al., 2021). В результате для каждой последовательности был рассчитан перечень изменений (делеций, вставок или точечных замен), отличающих ее от референса.

Для вируса с размером генома в 30000 нт одна точечная замена может произойти в каждой из 30 тыс. позиций и привести к изменению имеющегося нуклеотида (А, Т, Г или Ц) на один из трех других, что порождает $30000 \cdot 3 = 90000$ различных вариантов. Одиночная вставка может быть сделана в 30001 позиции – добавлена как в начало или конец последовательности, так и в любой из 29999 промежутков между имеющимися нуклеотидами. Она может содержать любую из четырех букв алфавита, т. е. имеется 120004 различных варианта таких вставок. И наконец, делеция может произойти в любой из 30000 позиций, порождая число вариантов, равное числу позиций. Впрочем, делеции и вставки, оставляющие вирус жизнеспособным, чаще всего происходят блоками, размер которых кратен трем, поскольку иначе такое изменение привело бы к сдвигу рамки считывания, что в подавляющем большинстве случаев делает геном нежизнеспособным. Таким образом, даже одно одиночное изменение может быть осуществлено более чем 240 тыс. различных способов, хотя значительная их часть (особенно те, что соответствуют делециям и вставкам) сделает геном нежизнеспособным.

Комбинация из двух произвольных точечных замен – это уже $(240000)^2 = 5.8 \cdot 10^{10}$, а трех – $(240000)^3 = 1.4 \cdot 10^{16}$ вариантов, причем на этот раз среди них будут и те, у которых с рамкой считывания все будет в порядке (результат изменений – удаление или вставка одного триплета, т. е. трех нуклеотидов подряд). При этом число различий между некоторыми современными вариантами SARS-CoV-2 и референсным геномом уже превышает 200, а (для масштаба) редакционное расстояние между SARS-CoV-2 и ближайшим к нему геномом другого вируса – коронавируса летучих мышей, RaTG13, – составляет 1136 (совпадают 96.1 % нуклеотидов) (Zhou et al., 2020; Temmam et al., 2022). Подробнее ряд вопросов о пространстве вариантов генетических последовательностей SARS-CoV-2 рассмотрен в работе (Palyanov, Palyanova, 2023), в которой, в частности, показано, что количество уже реализованных вариантов вируса составляет ничтожно малую долю относительно тех, что являются потенциально возможными. Таким образом, как продолжение мониторинга новых вариантов SARS-CoV-2, так и анализ уже накопленных за прошедшие 4.5 года миллионов геномов представляют интерес как с практической точки зрения, так и для получения новых фундаментальных знаний в области вирусологии и эпидемиологии.

Материалы и методы

Приведенные в работе результаты получены с помощью программного комплекса, созданного нами для осуществления анализа эволюции вирусов. Для разработки использован язык программирования C++, среда разработки – Microsoft Visual Studio Community 2019. Задействован один сторонний программный модуль, необходимый для осуществления глобального выравнивания вирусных геномов, – NextAlign от NextClade (https://github.com/nextstrain/nextclade/releases). Аппаратное обеспечение – компьютер на базе процессора Intel Core i7-10700K, 3.8 ГГц, 8 ядер, 32 Гб оперативной памяти.

Данные для анализа – генетические последовательности SARS-CoV-2. Данные, использованные в настоящей работе, – это полный набор геномов SARS-CoV-2, из БД Genbank (www.ncbi.nlm.nih.gov/sars-cov-2/) за период с 24.12.2019 по 24.06.2024 (4.5 года с момента сбора первого образца этого вируса, Wuhan-Hu-1, 24.12.2019). Их количество составило 8 641 740 шт., объем – 242 Гб. Референсный геном SARS-CoV-2, имеющий длину 29903 нт, состоит из 5′ UTR длиной 265 нт, CDS (в котором закодированы 29 белков (Bai et al., 2022)) длиной 29409 нт и 3′ UTR длиной 229 нт (UTR – нетранслируемая область, CDS – кодирующая последовательность). Этот набор данных, который продолжает пополняться с течением времени, и есть тот фундамент, на котором строятся исследования в области эволюционных изменений SARS-CoV-2. Еще одним источником данных является БД GISAID (в которую, вероятно, данные из Genbank входят практически полностью), геномы из которой еще предстоит проанализировать и сравнить с результатами для геномов из Genbank.

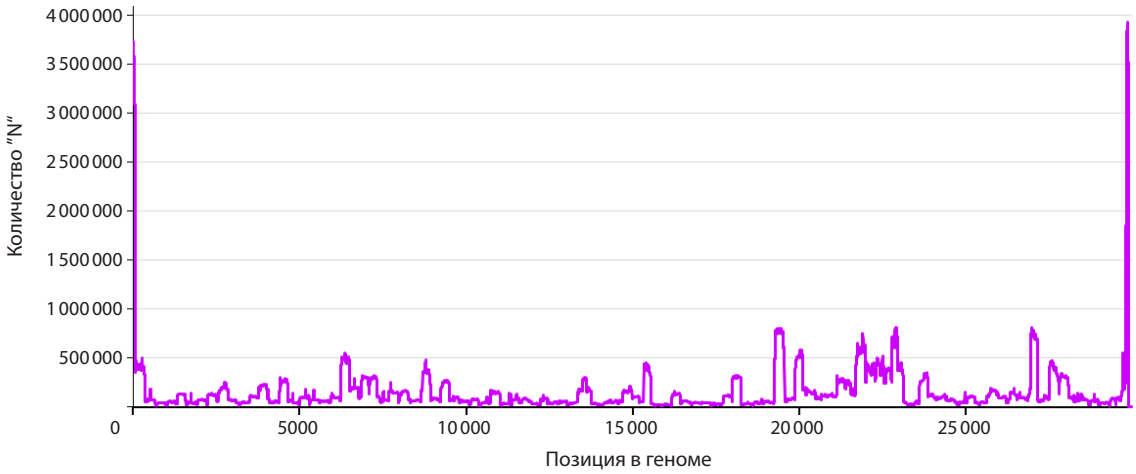


Рис. 1. Зависимость количества "N" от позиции в геноме (по оси абсцисс), полученная в результате суммирования по полному набору выровненных генетических последовательностей SARS-CoV-2 из Genbank в интервале с 24.12.2019 по 24.06.2024.

Качество данных, их предварительный анализ и фильтрация. Один из первых вопросов, возникающих при работе с набором нуклеотидных последовательностей вирусных геномов, – это их качество. В частности, в последовательностях могут присутствовать не только буквы, кодирующие нуклеотиды (A, T, G, C), но и "N", обозначающие неидентифицированные, неизвестные нуклеотиды в соответствующих позициях. Чем больше "N", тем больше неопределенности, и тем хуже для результатов анализа и их достоверности. В связи с этим, конечно, полезно представлять, насколько много таких последовательностей в исследуемом наборе данных, и насколько много "N" может встречаться в том или ином секвенированном геноме.

Расчеты показали, что из полного набора последовательностей (8641740) неидентифицированные нуклеотиды "N" встречаются в 6609933 геномах (76.5 %), а отсутствуют только в оставшихся 2031807 (23.5 %). Однако если рассматривать только CDS, то число геномов без "N" возрастает почти вдвое – до 3742117 (43.3 %). Помимо этого, мы построили зависимость частоты встречаемости "N" от позиции в геноме – на основе полного набора последовательностей, для которых было произведено глобальное выравнивание (рис. 1).

Как видно, имеются два наиболее значительных пика, в начале и в конце генома, соответствующие некодирующим участкам 5′ UTR и 3′ UTR, суммарная длина которых составляет 1.65 % от длины всего генома. Известно также, что в генетических последовательностях SARS-CoV-2 из GISAID и Genbank нетранслируемые области имеют высокий разброс по длинам 5′ UTR и 3′ UTR (Palyanov, Palyanova, 2023). С учетом того, что число геномов, в которых "N" встречаются в UTR и не встречаются в CDS, составляет 22.5 % от числа всех геномов, исключение UTR-участков из рассмотрения увеличит набор пригодных для анализа данных практически вдвое (к 23.5 % последовательностей, в которых "N" вообще не встречаются, ни в CDS, ни в UTR, добавятся еще 22.5 %, в которых "N" есть только в UTR).

В зависимости от того, каким является распределение геномов по числу "N", содержащихся в их CDS, можно либо использовать те из них, где всего несколько "N" (на фоне различий порядка 100 точечных замен это незначительная величина, хотя их наличие и вносит некоторую неопределенность), либо использовать только те геномы, в которых "N" в CDS отсутствуют. Построив упомянутое распределение (рис. 2), мы выяснили, что геномов с одной "N" – 1.8 %, с двумя и тремя – 0.8 и 0.9 % соответственно, а с "N" в пределах от 1 до 10 шт. на геном – 5.4 %. В результате на данном этапе было принято решение работать только с геномами, в которых "N" отсутствуют в CDS, и использовать в расчетах только CDS, исключив 5′ UTR и 3′ UTR.

Методы, алгоритмы и структуры данных. Для построения глобальных выравниваний всех геномов (с геномом Wuhan-Hu-1 в качестве референсной последовательности) мы использовали консольную версию NextAlign (работающую в многопоточном режиме), вызываемую с необходимыми параметрами из нашей программной системы. Это происходит при ее первом запуске или при необходимости пересчета выравниваний (например, в случае задействования другой выборки геномов). На полной выборке, упомянутой выше, состоящей из 8.6 млн геномов SARS-CoV-2, расчет выравниваний занимает около суток на рабочей станции с процессором Intel Core i7-10700K @ 3.8 ГГц (8 ядер, 16 потоков) и 32 Гб оперативной памяти (DDR4, 3600 ГГц). Результатом работы программы на данном этапе является сохранение на жестком диске в рабочей директории программы файлов с результатами всех рассчитанных выравниваний, которые затем используются нашей системой в качестве входных данных, взятых для анализа. Файлы представляют собой таблицы с несколькими десятками колонок, включающих различные характеристики геномов, метаданные, а также списки мутаций, делеций и вставок, отличающих рассматриваемый геном от референсного.

По мере считывания данных в оперативной памяти компьютера динамически формируется список структур,

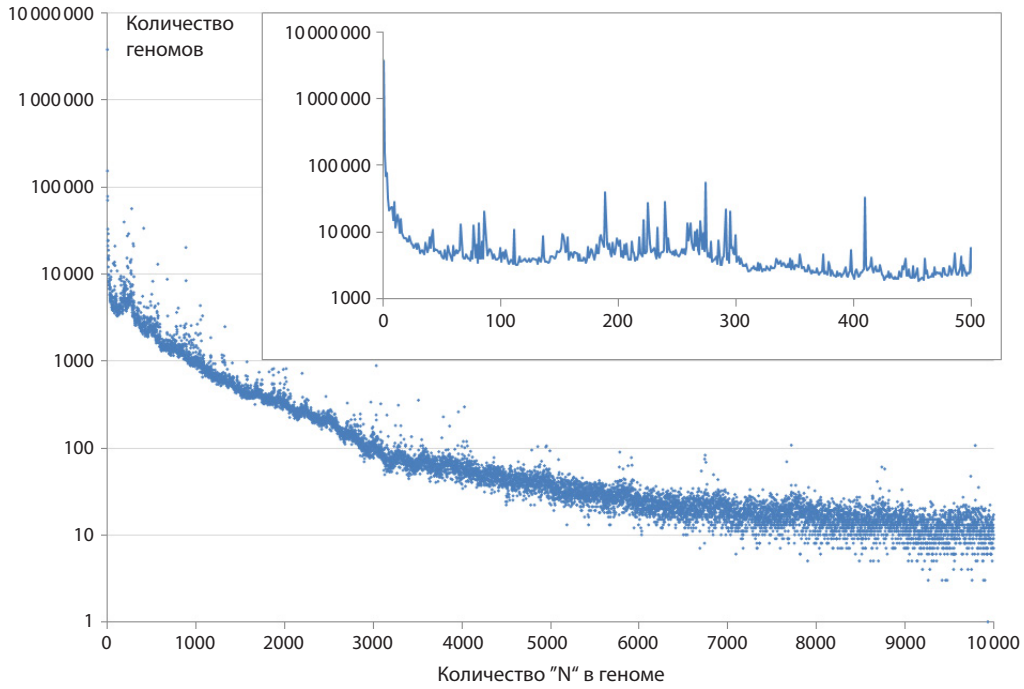


Рис. 2. Распределение геномов по количеству неидентифицированных нуклеотидов "N" в них.

Рассчитано по полному набору генетических последовательностей SARS-CoV-2 из Genbank в интервале с 24.12.2019 по 24.06.2024. Врезка показывает ту же зависимость, но с большим разрешением, для количества "N" в геноме в пределах от 0 до 500.

каждая из которых включает имя вируса, дату получения образца (collection date), географические данные, а также полный набор изменений, отличающих данный вариант от референсного генома:

- список точечных мутаций (однопозиционных замен), каждый элемент которого содержит номер позиции в геноме, соответствующий данной мутации, и букву, кодирующую нуклеотид, появившийся в данной позиции в результате замены (предыдущий нуклеотид, который был до мутации, не храним, при необходимости его всегда можно прочитать из соответствующей позиции референсного генома);
- список делеций, каждая из которых определяется двумя числами – позициями начала и конца делеции;
- список вставок, каждая из которых задается позицией в геноме, сразу после которой произошла вставка, а также вставленной последовательностью.

Такая организация данных позволяет легко и быстро сравнивать два произвольных генома. Особенно быстро получается определить, идентичны они или нет. Вместо сравнения каждой из 29 409 позиций первого и второго геномов достаточно просто сравнить число элементов в их списках точечных мутаций, делеций и вставок, хотя бы при одном отличии уже понятно, что геномы различаются. Впрочем, таким образом можно не только получить результат сравнения геномов, но и вычислить редакционное расстояние между ними. Совпадающие элементы списков различий не дают вклада в разницу между геномами, тогда как каждый элемент отличия от референса, присутствующий в одном геноме и отсутствующий в другом, добавляет соответствующее количество отличий. Каждая замена, произошедшая в одной и той же позиции обоих геномов, но приведшая к заменам на разные нуклеотиды, тоже, конечно, добавляет +1 к редакционному расстоянию. С учетом того что размеры списков небольшие, сравнение осуществляется значительно быстрее, чем сравнение двух геномов без предварительного выравнивания.

Предложенный нами способ компактного представления нуклеотидных последовательностей родственных геномов в оперативной памяти компьютера имеет много общего со способом сжатия, представляющим последовательности в виде филогенетического дерева с заменами на ребрах. Более того, само представление каждого генома как совокупности изменений, которые необходимо осуществить, чтобы перейти от референса к рассматриваемому геному, основано на тех же данных о структуре филогенетического дерева, выстраиваемого на базе множественного выравнивания рассматриваемых последовательностей.

Особенность нашей реализации состоит в том, что структура данных в оперативной памяти, представляющая множество рассматриваемых последовательностей, не является филогенетическим деревом, а вместо этого представляет собой список его «листьев», сортированный в хронологическом порядке, по дате получения образцов. Для таких задач, как анализ не просто имеющегося спектра вариантов вируса, а его эволюционных изменений с учетом времени их возникновения, наш подход обеспечивает значительное преимущество в скорости доступа к данным. Дело в том, что он позволяет двигаться по оси времени, просто увеличивая или уменьшая индекс элемента массива, состоящего из упорядоченных по времени геномов. В древовидном же представлении поиск всех геномов, соответствующих определенному году, месяцу

и дню, в общем случае сведится к обходу всего дерева, и так для каждого нового варианта. При этом каждый «лист» в нашем подходе содержит всю информацию о своей «ветви» дерева, что дает возможность легко и быстро вычислять редакционное расстояние для любой пары вариантов геномов.

Результаты

Кластерная структура множества геномов

В процессе исследования мы обратили внимание на то, что среди рассматриваемых геномов довольно часто оказываются такие, у которых CDS полностью, на 100 %, идентичны между собой, при том что дата сбора образца, географические данные и прочие метаданные чаще всего различаются. Добавив в нашу программную систему функцию выявления всех геномов с идентичными CDS (и объединения их в «кластеры»), мы разбили весь имеющийся у нас набор геномов на группы. Статистика по ним приведена в таблице.

Также мы рассчитали зависимость между размером кластера и числом кластеров того или иного размера (рис. 3). При этом между размерами кластеров и их временем жизни какой-либо явно выраженной зависимости не прослеживается, распределение представляет собой облако точек, большая часть которого сосредоточена в области от 1 до 1000 по оси «размер кластера» и от 1 до 500 – по оси «время существования кластера» (рис. 4).

Статистические данные о кластерах, объединяющих геномы с полностью одинаковыми CDS, включая размер и протяженность во времени

Полное число геномов SARS-CoV-2 из БД Genbank за период с 24.12.2019 по 24.06.2024	8 641 740
Число геномов из полного набора, в которых CDS не содержит неидентифицированных нуклеотидов, “N”	3 742 117
В выборке с CDS без “N” – число геномов с уникальной, больше нигде не встречающейся последовательностью нуклеотидов в CDS	1 690 699
В выборке с CDS без “N” – число геномов, образующих кластеры размером 2 и более	2 051 448
Число кластеров с размером ≥ 2	461 511
Число кластеров, для которых время существования (интервал между самой ранней и самой поздней датой сбора образца генома из числа входящих в кластер) составляет более одного дня	366 427
Максимальный размер кластера (число входящих в него геномов)	12 824
Максимальное время существования кластера (lineage 19A)	1539 дней
Средний размер кластера	4.4
Среднее время существования кластера	14.8 дня
Среднее время существования кластера (исключая те, у которых время существования один день)	18.4 дня

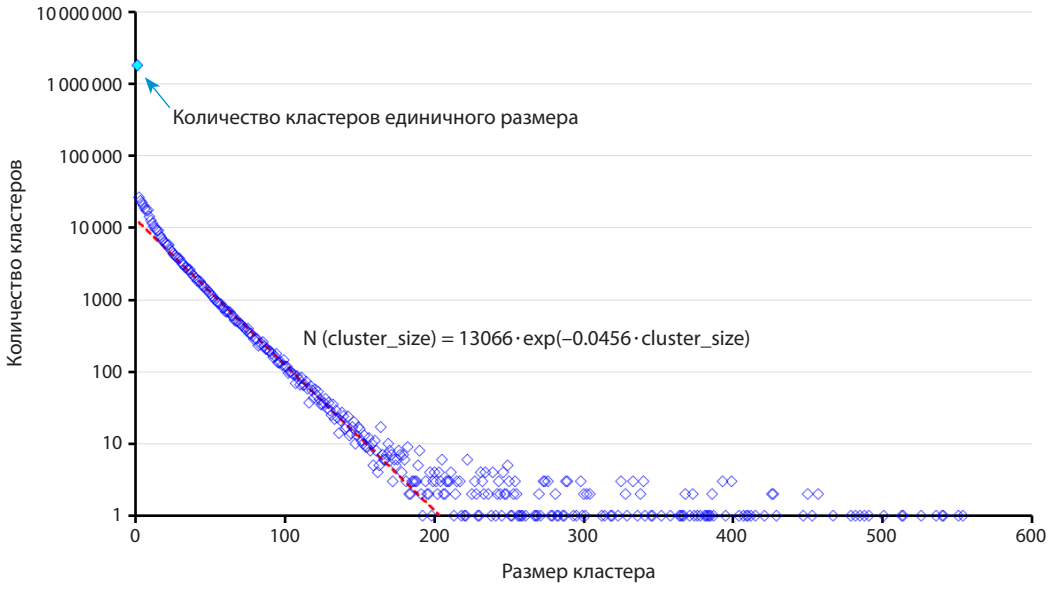


Рис. 3. Зависимость между числом кластеров и их размером для множества геномов SARS-CoV-2 из Genbank в интервале с 24.12.2019 по 24.06.2024.

В интервале значений размера кластера от 20 до 200 она хорошо аппроксимируется экспонентой с параметрами, указанными на рисунке.

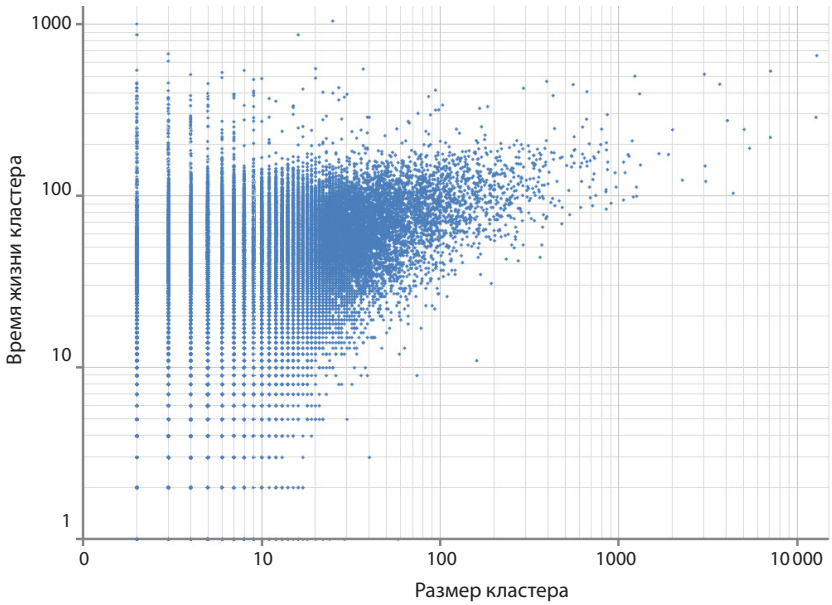


Рис. 4. Облако точек, представляющее множество геномов SARS-CoV-2 из Genbank (в интервале с 24.12.2019 по 24.06.2024) при использовании характеристик «размер кластера» и «время жизни кластера».

Мы отобразили все кластеры размером ≥ 200 на временной оси в виде линий, начало и конец которых соответствуют интервалам существования кластеров. Помимо этого, добавлены все кластеры размером 100–199, конец интервала существования у которых имеет значение ≥ 1000 дней с момента получения первого генома SARS-CoV-2. Это множество кластеров охватывает всю временную шкалу, хотя в интервале между 3 и 3.5 года явно есть и другие кластеры, но все они меньшего размера, чем представленные на рис. 5. Отдельная линия 19A, самая длинная на рис. 5, соответствует кластеру с самым долгим временем жизни (1539 дней или 4.2 года), упомянутому в таблице. В связи с этим генетическая линия 19A, просуществовавшая так долго, представляется довольно интересной. Данный вариант генома обнаруживался достаточно стабильно как в начале пандемии, так и в 2023–2024 гг.

Новый подход к визуализации эволюционных изменений SARS-CoV-2

Получив возможность быстро вычислять редакционное расстояние между парой любых вариантов нуклеотидных последовательностей, мы сделали это для всего множества геномов SARS-CoV-2 из Genbank за 4.5 года. Таким образом, для каждого генома имеется пара чисел – дата получения образца генома (collection date) и редакционное расстояние между ним и референсом. Бывает, что разные варианты геномов оказываются обладателями одинаковой пары значений «дата + редакционное расстояние», поскольку одно и то же значение редакционного расстояния может быть обеспечено как отличием, к примеру, на делецию длиной 30, вставку такой же

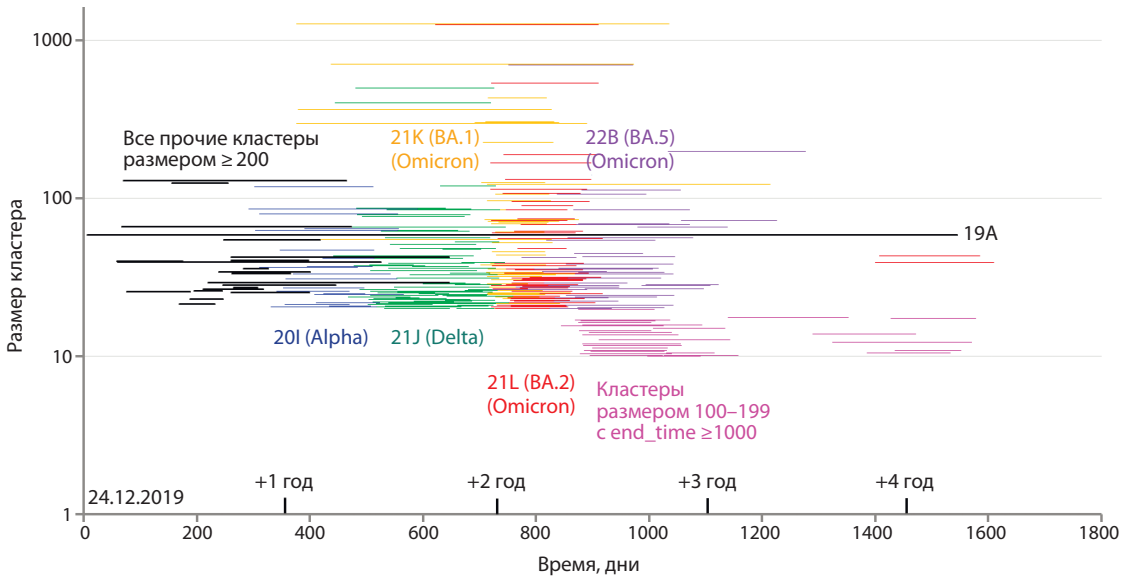


Рис. 5. Наиболее крупные кластеры (размером ≥ 200) и их интервалы существования (линии, соединяющие день первого появления последовательности нуклеотидов, представляющих данный кластер, и день, когда был взят последний образец).

Все кластеры размером ≥ 200, обозначенные одним и тем же цветом, имеют разные последовательности, но относятся к одной и той же генетической линии, название которой показано соответствующим цветом. Исключение составляют «все прочие кластеры размером ≥ 200», обозначенные черным цветом, которые представляют собой набор различных генетических линий (19A, 20A, 20B, 20C, 20E и 20F), а также кластеры размером 100–199, отмеченные пурпурным цветом (время завершения существования у которых составляет ≥ 1000 дней с даты получения первого генома SARS-CoV-2, 24.12.2019 г.).

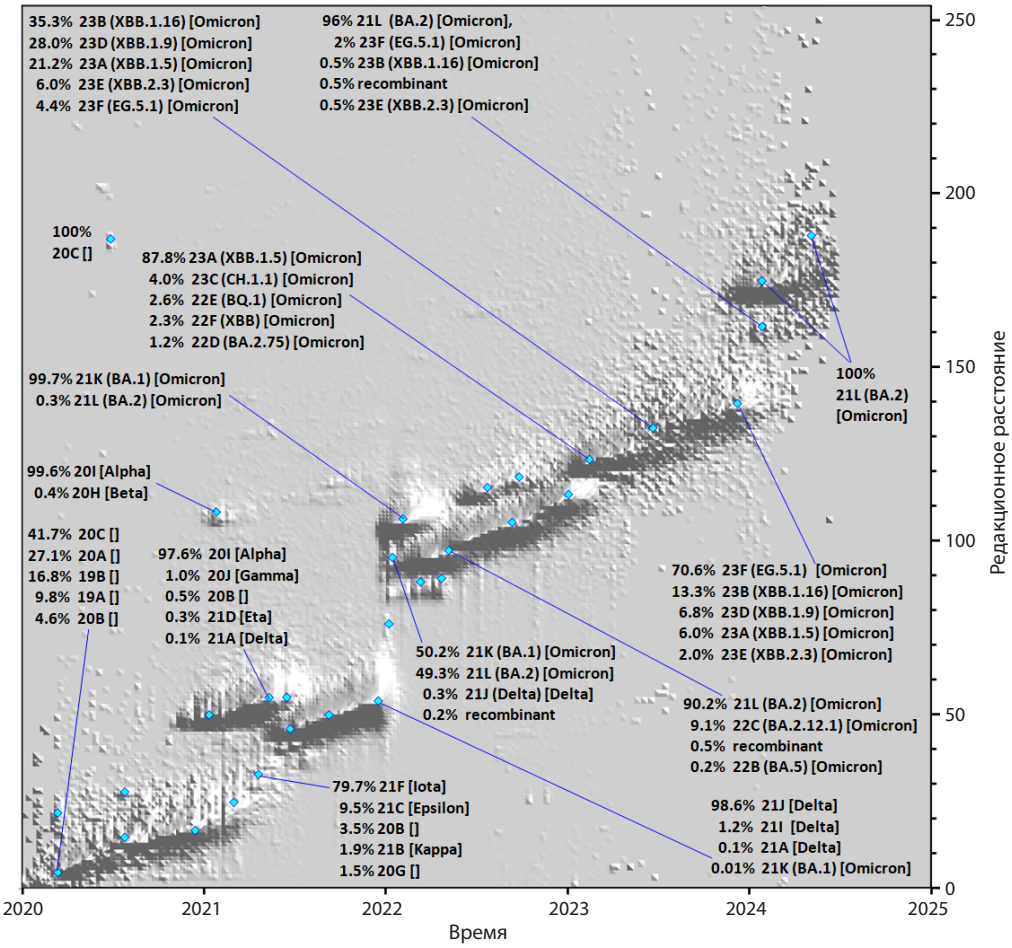


Рис. 6. Ландшафт пространства вариантов SARS-CoV-2, «посещенных» вирусом за период с 24.12.2019 до 24.06.2024, спроецированный на три оси: OX – время (с дискретизацией в 1 неделю), OY – различие (редакционное расстояние) между точкой на ландшафте и референсным геномом, OZ – доля геномов, соответствующих точке с определенными значениями X и Y, отнесенная к общему числу геномов, попавших в базу в неделю X. Такая нормировка необходима в связи с тем, что зависимость числа образцов геномов, собранных в ту или иную неделю по всему миру, имеет существенные изменения с течением времени, и без предложенной нормировки участки, для которых за день было собрано, к примеру, не 10 000, а всего 100 геномов, будут совершенно незаметны на рисунке, тогда как даже с этим числом геномов они вполне информативны.

длины, или же на 30 точечных мутаций, рассредоточенных по всему геному. Если ввести третью величину – количество случаев, когда вариант генома имеет определенное редакционное расстояние до референса и определенную дату получения образца, то можно рассчитать тройки этих величин на основе полного набора геномов SARS-CoV-2 и отобразить их в виде поверхности, что мы и сделали (рис. 6).

На рис. 6 мы отметили голубыми точками ряд интересных элементов ландшафта, для каждого из которых был рассчитан спектр соответствующих им вариантов. Для многих из них эту информацию удалось разместить на рисунке. На ландшафте видны области с различными особенностями – узкие протяженные «горные хребты», имеющие начало, конец и характерный угол наклона с близкими значениями для большинства их них, по-видимому, связанный со скоростью накопления изменений в геноме, появляющихся за счет точечных замен нуклеотидов.

Есть также области, в которых редакционное расстояние для всего множества вариантов, существующих в определенный момент времени, быстро и значительно изменяет среднее значение или испытывает ветвление, разделяясь на несколько параллельных, визуально различимых путей. Предположив, что такие резкие и значительные изменения могут быть связаны с делециями, вставками или рекомбинационными событиями, мы построили еще три рисунка, подобных рис. 6, для которых в качестве редакционного расстояния использовали не полную его величину, а три раздельных вклада – от совокупности точечных замен (рис. 7), делеций (рис. 8) и вставок (рис. 9). При этом из полного числа геномов (в CDS которых нет “N”), 3 742 117 шт., число геномов с мутациями относительно референса составило 3 741 518 шт., число геномов с делециями – 3 520 077 шт., а число геномов со вставками – 528 414 шт.

Как видно на рис. 7–9, особенности динамики эволюционных изменений, вносимых в геномы SARS-CoV-2

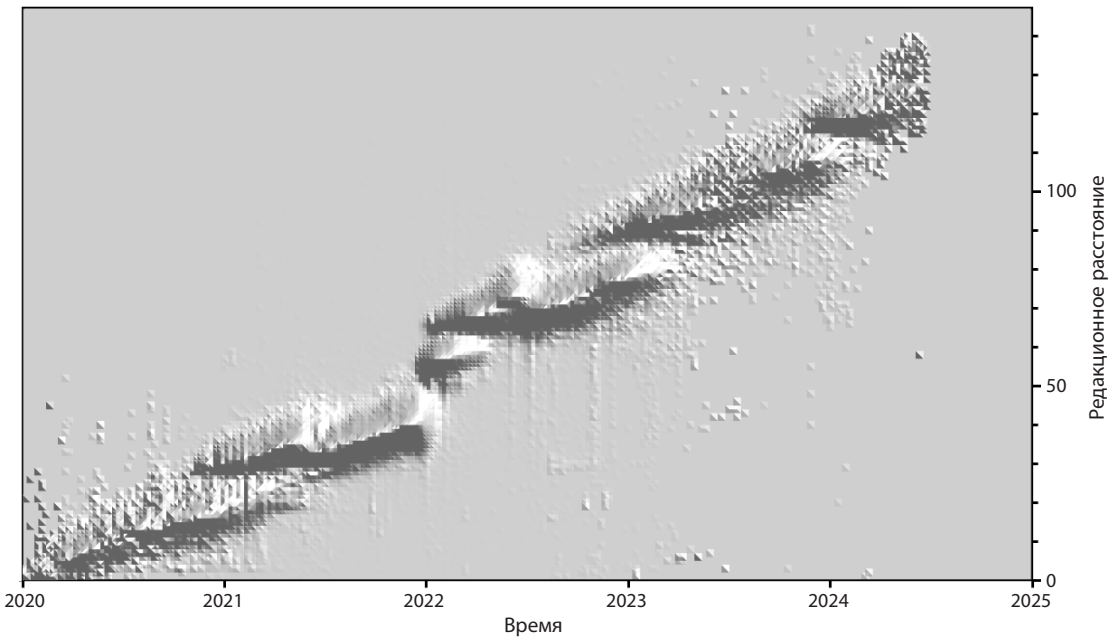


Рис. 7. Ландшафт эволюционных изменений на основе вклада только от точечных замен.

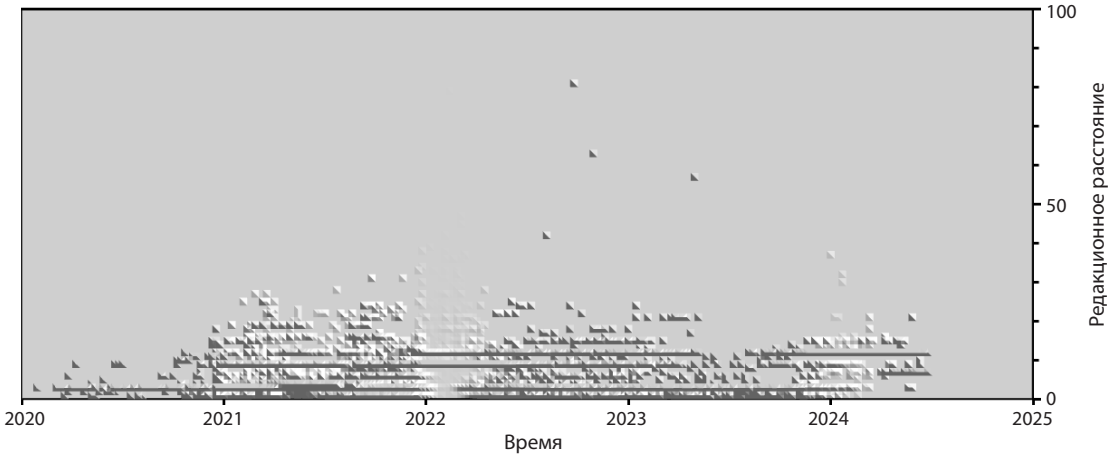


Рис. 8. Ландшафт эволюционных изменений на основе вклада только от вставок.

различными эволюционными механизмами, существенно различаются для точечных замен, делеций и вставок. Из рис. 7 можно заключить, что накопление количества мутаций на большом масштабе времени возрастает линейно (особенно если судить по верхней границе области эволюционного пути). За 4.5 года накопилось около 130 точечных мутаций, т. е. скорость накопления составила примерно 29 шт. в год (2.4 в месяц или 0.08 в день). Влияние делеций тоже ощутимо, и они тоже накапливаются по закону, близкому к линейному, но с меньшей скоростью – около 50 шт. (по совокупной длине) за 4.5 года, т. е. около 11 в год или чуть менее 1 в месяц. Вставки же, как видно из рис. 8, дают заметно меньший вклад, чем делеции, величина которого, за исключением первого года, практически не растет со временем, – он держится на уровне 20 шт. относительно референса (хотя содержание этих вставок, в принципе, может меняться на протяжении рассматриваемого интервала времени).

Обсуждение

Мы выполнили ряд оценок и расчетов, в основном с помощью разработанных нами программных средств, для улучшения понимания того, какие особенности и закономерности обнаруживаются у эволюции генетических последовательностей коронавируса SARS-CoV-2. Предложенный метод визуализации ландшафтов эволюционных изменений позволил наглядно увидеть множество деталей и особенностей, которые не видны, например, на филогенетическом древе. При этом быстрые изменения эволюционной траектории, сопровождающиеся ступенчатыми изменениями величины редакционного расстояния, как видно из рис. 6, как правило, сопровождаются сменой доминирующего в популяции варианта вируса. Так, например, для одного из таких «скачков» на эволюционном ландшафте удалось зарегистрировать генетические линии «Йота», «Эпсилон» и «Каппа» (примерно в первой четверти 2022 г.).

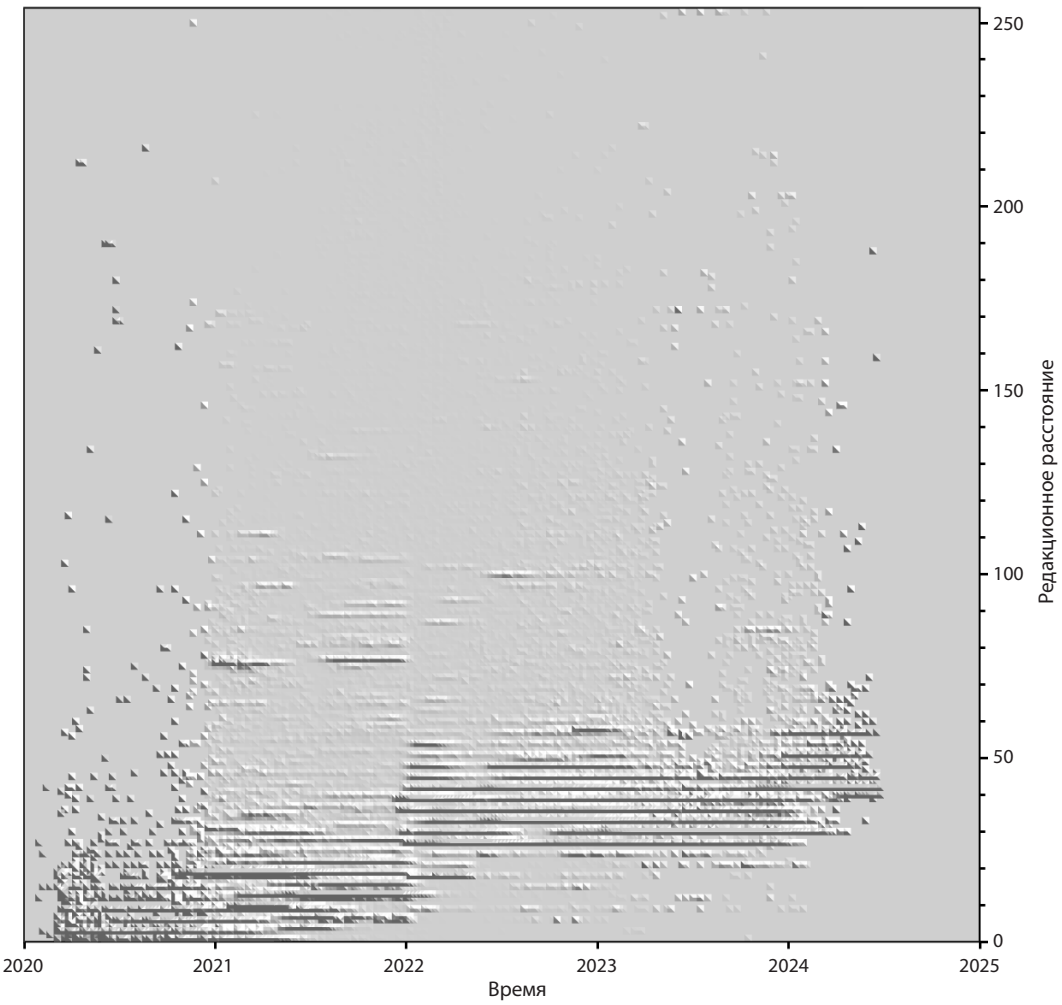


Рис. 9. Ландшафт эволюционных изменений на основе вклада только от делеций.

Наблюдаемое отсутствие роста величины вклада в редакционное расстояние, обеспечиваемого вставками, ранее упомянутое для результатов, связанных с рис. 8, может быть обусловлено тем, что слишком большое число вставок способно нарушать стабильность вируса. Увеличение числа вставок увеличивает физический размер генома и тем самым может ухудшать его способность помещаться внутрь белковой оболочки, которая, предположительно, рассчитана на нахождение внутри нее объекта вполне определенного объема. Таким образом, в ходе эволюции число вставок относительно референсного генома может расти, но не превышать 20–30 нт. Если какие-то из новых вариантов оказываются более приспособленными, чем их предшественники, то очень скоро они могут их вытеснить. Видно, что только в начале 2024 г. исчезают варианты, в которых вообще нет вставок, вероятно, в связи с тем, что за четыре года эволюции все-таки нашлись такие вставки, которые оказались заметно более приспособленными, чем варианты вовсе без вставок, и закрепились в популяции. Также из рис. 8 и 9 понятно, что большинство делеций и вставок имеют длину, кратную 3, что имеет вполне очевидное объяснение, связанное с тем, что иные варианты длин будут приводить к сдвигу рамки считывания при синтезе закодированных в геноме белков и в большинстве случаев – к нежизнеспособным экземплярам геномов, вирионы которых не могут быть собраны.

Заключение

В результате проделанной работы получены следующие основные результаты:

- предложен и программно реализован способ представления нуклеотидных последовательностей геномов вирусов, обеспечивающий исключительно компактное размещение их в оперативной памяти компьютера. На примере коронавируса SARS-CoV-2 показано, что обеспечивается сжатие в 1500 раз. Использование его для пересылки генетических данных по сети могло бы снизить нагрузку на серверы и снизить сетевой трафик в соответствующее число раз (особенно при передаче больших массивов данных);
- для полного набора геномов SARS-CoV-2 (без “N” в CDS) исследовано наличие кластеров полностью идентичных геномов. Выяснено, что их размер может превышать 10 000 шт., а интервал времени, в пределах которого они встречаются, может охватывать до нескольких сотен дней и более;

- предложен новый способ отображения эволюционной динамики вирусов в виде ландшафта, визуализирующего проекцию пространства вариантов геномов вируса на три оси – время (T), редакционное расстояние до референсного генома (D) и долю геномов (P) в каждой точке (T, D) в полном количестве геномов, соответствующих данному T;
- показано, что ландшафт, построенный для D (вычисляемого как сумма вкладов от точечных мутаций, делеций и вставок), можно разделить на три отдельных ландшафта, рассчитываемых отдельно для каждого из вкладов. Каждый из них имеет различный характер, позволяющий оценить вклад и влияние каждого из упомянутых механизмов на эволюцию вируса. Рассчитаны константы, характеризующие каждый из механизмов и скорость приобретаемых благодаря ему изменений;
- обнаруженный нами факт того, что линия 19A существовала наиболее долго по сравнению с остальными кластерами, охватывая весь период пандемии, позволяет нам предложить создавать новые вакцины против SARS-CoV-2 на основе именно этой линии как сохраняющей наибольшую конкурентоспособность по сравнению с остальными вариантами, а значит, содержащей наиболее характерные для данного вируса особенности, которые могут быть распознаны иммунной системой.

В будущем мы планируем исследовать возможности предложенного метода визуализации эволюции более детально, однако уже сейчас можно сказать, что он представляется полезным, имеет потенциал для дальнейшего использования и развития и может быть применен не только к SARS-CoV-2, но и к другим вирусам. Это же можно сказать о предложенном методе компактного представления вирусных геномов, применение которого во всех областях, связанных с хранением, передачей по сети, обработкой и анализом большого количества вариантов родственных геномов (как вирусов, так и живых организмов), обеспечит значительные преимущества.

Список литературы / References

Aksamentov I., Roemer C., Hodcroft B., Neher R.A. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Software*. 2021;6(67):3773. doi 10.21105/joss.03773

Amicone M., Borges V., Alves M.J., Isidro J., Zé-Zé L., Duarte S., Vieira L., Guiomar R., Gomes J.P., Gordo I. Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *Evol. Med. Public Health.* 2022;10(1):142-155. doi 10.1093/emph/eoac010

Bai C., Zhong Q., Gao G.F. Overview of SARS-CoV-2 genome-encoded proteins. *Sci. China Life Sci*. 2022;65(2):280-294. doi 10.1007/s11427-021-1964-4

Bolze A., Basler T., White S., Rossi A.D., Wyman D., Dai H., Roychoudhury P., Greninger A.L., Hayashibara K., Beatty M., Shah S., Stous S., McCrone J.T., Kil E., Cassens T., Tsan K., Nguyen J., Ramirez J., Carter S., Cirulli E.T., Barrett K.S., Washington N.L., Belda-Ferre P., Jacobs S., Sandoval E., Becker D., Lu J.T., Isaksson M., Lee W., Luo S. Evidence for SARS-CoV-2 Delta and Omicron co-infections and recombination. *Med.* 2022;3(12):848-859. doi 10.1016/j.medj.2022.10.002

Campagnola G., Govindarajan V., Pelletier A., Canard B., Peersen O.B. The SARS-CoV-2 nsp12 polymerase active site is tuned for large-genome replication. *J. Virol.* 2022;96(16):e0067122. doi 10.1128/jvi.00671-22

Cui X., Wang Y., Zhai J., Xue M., Zheng C., Yu L. Future trajectory of SARS-CoV-2: Constant spillover back and forth between humans and animals. *Virus Res.* 2023;328:199075. doi 10.1016/j.virusres.2023.199075

Palyanov A.Yu., Palyanova N.V. On the space of SARS-CoV-2 genetic sequence variants. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):839-850. doi 10.18699/VJGB-23-97

Palyanova N.V., Sobolev I.A., Alekseev A., Glushenko A., Kazachkova E., Markhaev A., Kononova Y., Gulyaeva M., Adamenko L., Kurskaya O., Bi Y., Xin Y., Sharshov K., Shestopalov A. Genomic and epidemiological features of COVID-19 in the Novosibirsk region during the beginning of the pandemic. *Viruses*. 2022;14(9):2036. doi 10.3390/v14092036

Palyanova N.V., Sobolev I.A., Palyanov A.Yu., Kurskaya O.G., Komissarov A.B., Danilenko D.M., Fadeev A.V., Shestopalov A.M. The development of the SARS-CoV-2 epidemic in different regions of Siberia in the 2020–2022 period. *Viruses*. 2023;15(10):2014. doi 10.3390/v15102014

Sanjuán R., Domingo-Calap P. Mechanisms of viral mutation. *Cell. Mol. Life Sci*. 2016;73(23):4433-4448. doi 10.1007/s00018-016-2299-6

Simon-Loriere E., Holmes E.C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* 2011;9(8):617-626. doi 10.1038/nrmicro2614

Sonnleitner S.T., Prelog M., Sonnleitner S., Hinterbichler E., Halbfurter H., Kopecky D.B.C., Almanzar G., Koblmüller S., Sturmbauer C., Feist L., Horres R., Posch W., Walde G. Cumulative SARS-CoV-2 mutations and corresponding changes in immunity in an immunocompromised patient indicate viral evolution within the host. *Nat. Commun.* 2022;13(1):2560. doi 10.1038/s41467-022-30163-4

Temmam S., Vongphayloth K., Baquero E., Munier S., Bonomi M., Regnault B., Douangboubpha B., Karami Y., Chrétien D., Sanamxay D., Xayaphet V., Paphaphanh P., Lacoste V., Somlor S., Lakeomany K., Phommavanh N., Pérot P., Dehan O., Amara F., Donati F., Bigot T., Nilges M., Rey F.A., van der Werf S., Brey P.T., Eloit M. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature.* 2022;604(7905):330-336. doi 10.1038/s41586-022-04532-4

Wu F., Zhao S., Yu B., Chen Y.M., Wang W., Song Z.-G., Hu Y., Tao Z.-W., Tian J.-H., Pei Y.-Y., Yuan M.-L., Zhang Y.-L., Dai F.-H., Liu Y., Wang Q.-M., Zheng J.-J., Xu L., Holmes E.C., Zhang Y.-Z. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. doi 10.1038/s41586-020-2008-3

Zhou P., Yang X.L., Wang X.G., Hu B., Zhang L., Zhang W., Si H.-R., Zhu Y., Li B., Huang C.-L., Chen H.-D., Chen J., Luo Y., Guo H., Jiang R.-D., Liu M.-Q., Chen Y., Shen X.-R., Wang X., Zheng X.-S., Zhao K., Chen Q.-J., Deng F., Liu L.-L., Shi Z.-L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273. doi 10.1038/s41586-020-2012-7

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 15.09.2024. После доработки 23.10.2024. Принята к публикации 24.10.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):854-863

doi 10.18699/vjgb-24-93

Поиск и функциональная аннотация многодоменных белков семейства ФА2 у плоских червей

М.Е. Бочарникова [1, 2, 3] ✉, И.И. Турнаев [1, 3], Д.А. Афонников [1, 2, 3]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
✉ bocharnikova@bionet.nsc.ru

Аннотация. Фосфолипазы типа А2 (ФА2) – это семейство гидролаз, которые катализируют процесс гидролиза фосфолипидов, играя ключевую роль во многих молекулярных процессах при функционировании клеток и организма в целом. Данное семейство подразделяется на 16 групп, объединенных в шесть основных типов. Впервые ФА2 были выделены как цитотоксины яда у змей и ферменты панкреатического сока у свиней. Изучение этих ферментов в настоящее время вызывает большой интерес, поскольку было показано, что ряд ФА2 участвует в процессах канцерогенеза. Наиболее хорошо изучены ферменты ФА2 у модельных организмов и человека. Однако их наличие и функциональная роль у немодельных организмов изучены слабо. К таким малоизученным таксонам относятся плоские черви, ряд видов которых является паразитами человека. У паразитических плоских червей ранее было охарактеризовано несколько генов ФА2 и показана их возможная роль во взаимодействии «паразит–хозяин». Но систематической идентификации генов ФА2 у этого таксона не проведено. В работе осуществлены поиск и сравнительный анализ последовательностей ФА2, кодируемых в геномах плоских червей. Исследовано 44 вида, представленных 2 свободноживущими и 42 паразитическими организмами. Анализ выполнен на основе поиска ортологических групп белок-кодирующих генов с учетом доменной структуры белков. У плоских червей обнаружено 12 из 13 известных типов фосфолипаз А2, имеющихся в 11 ортологических группах. Часть фосфолипаз нескольких типов попала в одну ортологическую группу, часть типов распалась на несколько ортогрупп в соответствии с особенностями доменной структуры. Показано, что ФА2 кальций-независимого типа, ФА2 тромбоцитарно-активирующего типа групп G8 и лизосомальные ФА2 группы G15 представлены во всех крупных таксонах плоских червей и в большинстве изученных нами видов. Для генов, кодирующих ферменты у свободноживущих червей, наблюдается множественное число копий. У паразитических плоских червей, наоборот, происходит потеря основной части генов специфически по отношению как к отдельным таксонам, так и к отдельным группам/подсемействам ФА2. Обнаружена ортологическая группа секретируемых фосфолипаз, которая среди паразитов имеется только у дигенетических сосальщиков, при этом в геномах описторхид это семейство подверглось дупликациям. Интересно, что ранее в ряде экспериментальных работ показано влияние белков *Clonorchis sinensis* этой ортогруппы на раковую трансформацию клеток организма-хозяина. Наши результаты дали возможность впервые систематически идентифицировать последовательности ФА2 у плоских червей и продемонстрировали, что их эволюция подвержена процессам потерь генов, характерных в целом для геномов паразитов. Кроме того, наш анализ позволил выявить таксон-специфические процессы дупликаций и потерь генов ФА2 у паразитических организмов, которые могут быть связаны с процессами их взаимодействия с организмом хозяина.

Ключевые слова: фосфолипаза А2; плоские черви; многодоменные белки; паразитизм; филогения; структура доменов.

Для цитирования: Бочарникова М.Е., Турнаев И.И., Афонников Д.А. Поиск и функциональная аннотация многодоменных белков семейства ФА2 у плоских червей. *Вавиловский журнал генетики и селекции*. 2024;28(8):854-863. doi 10.18699/vjgb-24-93

Финансирование. Работа поддержана бюджетным проектом № FWNR-2022-0020. Обработка данных проводилась с использованием вычислительных ресурсов ЦКП «Биоинформатика» ИЦиГ СО РАН.

Search for and functional annotation of multi-domain PLA2 family proteins in flatworms

M.E. Bocharnikova [1, 2, 3] ✉, I.I. Turnaev [1, 3], D.A. Afonnikov [1, 2, 3]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
✉ bocharnikova@bionet.nsc.ru

© Бочарникова М.Е., Турнаев И.И., Афонников Д.А., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

Abstract. The phospholipase A2 (PLA2) is a superfamily of hydrolases that catalyze the hydrolysis of phospholipids and play a key role in many molecular processes in the cells and the organism as a whole. This family consists of 16 groups divided into six main types. PLA2 were first isolated from venom toxins and porcine pancreatic juice. The study of these enzymes is currently of great interest, since it has been shown that a number of PLA2 are involved in the processes of carcinogenesis. PLA2 enzymes were characterized in detail in model organisms and humans. However, their presence and functional role in non-model organisms is poorly understood. Such poorly studied taxa include flatworms, a number of species of which are human parasites. Several PLA2 genes have previously been characterized in parasitic flatworms and their possible role in parasite-host interaction has been shown. However, no systematic identification of the PLA2 genes in this taxon has been carried out. The paper provides a search for and a comparative analysis of PLA2 sequences encoded in the genomes of flatworms. 44 species represented by two free-living and 42 parasitic organisms were studied. The analysis was based on identification of orthologous groups of protein-coding genes, taking into account the domain structure of proteins. In flatworms, 12 of the 13 known types of animal A2 phospholipases were found, represented by 11 orthologous groups. Some phospholipases of several types fell into one orthologous group, some types split into several orthogroups in accordance with their domain structure. It has been shown that phospholipases A2 of the calcium-independent type, platelet-activating phospholipases from group G8 and lysosomal phospholipases from group G15 are represented in all large taxa of flatworms and the vast majority of the species studied by us. In free-living flatworms PLA2 genes have multiple copies. In parasitic flatworms, on the contrary, loss of genes occur specifically in individual taxa specifically for groups or subfamilies of PLAs. An orthologous group of secreted phospholipases has been identified, which is represented only in Digenea and this family has undergone duplications in the genomes of opisthorchids. Interestingly, a number of experimental studies have previously shown the effect of *Clonorchis sinensis* proteins of this orthogroup on the cancer transformation of host cells. Our results made it possible for the first time to systematically identify PLA2 sequences in flatworms, and demonstrated that their evolution is subject to gene loss processes characteristic of parasite genomes in general. In addition, our analysis allowed us to identify taxon-specific processes of duplication and loss of PLA2 genes in parasitic organisms, which may be associated with the processes of their interaction with the host organism.

Key words: phospholipase A2; flatworms; multi-domain proteins; parasitism; phylogeny; domain structure.

For citation: Bocharnikova M.E., Turnaev I.I., Afonnikov D.A. Search for and functional annotation of multi-domain PLA2 family proteins in flatworms. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):854-863. doi 10.18699/vjgb-24-93

Введение

Белковое семейство фосфолипаз А2 (ФА2) представляет собой группу гидролаз, которые катализируют процесс гидролиза фосфолипидов, играя ключевую роль в функционировании клеток и организма в целом (Филькин и др., 2020; Murakami et al., 2020). Фосфолипазы А2 известны тем, что являются основными компонентами токсинов ядов змей (Bitar et al., 2021), насекомых (Bitar et al., 2021), хищных беспозвоночных, например паукообразных (Salabi, Jafari, 2024) или моллюсков (McIntosh et al., 1995). Фосфолипазы А2 из яда змей гидролизуют фосфолипиды клеточных мембран, что приводит к разрушению клеток, высвобождению арахидоновой кислоты и активации воспалительных процессов. Их воздействие может обуславливать и более серьезные системные эффекты, включая поражение нервной системы (Bitar et al., 2021), это демонстрирует множественность их функций (Gutiérrez, Lomonte, 2013).

Семейство ФА2 подразделяется на 16 групп (Dennis et al., 2011), объединенных в шесть основных типов: секретируемые, цитозольные, кальций-независимые, тромбоцитарно-активирующие факторы, лизосомальные и адипоспецифические (Murakami et al., 2020). Среди основных молекулярных функций ФА2 выделяются: расщепление липидов, ремоделинг жирных кислот, взаимодействие с фосфолипидами лизосом и жировой ткани (Mouchlis, Dennis, 2022). Однако эти ферменты у животных вовлечены в большое количество таких важнейших функций, как антибактериальная, антивирусная, иммунная и противовоспалительная (Dennis et al., 2011).

Известны и антипаразитические свойства ФА2 (Teixeira et al., 2022). В настоящее время эти белки вызывают большой интерес в связи с тем, что нарушение липидного метаболизма, регулируемого ФА2, часто приводит к различным заболеваниям, в том числе и к канцерогенезу (Турнаев и др., 2022). Показано их участие в процессах канцерогенеза. Секретируемые ФА2 демонстрируют повышенную экспрессию в злокачественных опухолях таких органов, как желудок (Scott et al., 2010), легкие (Park et al., 2012), кишечник (Murase et al., 2017) и печень (Shang et al., 2017).

Фосфолипазы А2 – древние гены, они встречаются у всех таксонов живых организмов: бактерий, протистов, архей, животных, грибов и растений (Nevalainen et al., 2012). Их эволюционный анализ позволяет более детально рассмотреть функциональные особенности этих белков, уточнить их роль в важнейших биологических процессах (Murakami et al., 2020; Турнаев и др., 2022). Наиболее хорошо изучены ферменты ФА2 у модельных организмов и человека. Однако их наличие и функциональная роль у немодельных организмов изучены слабо. К таким малоизученным таксонам относятся плоские черви, ряд видов которых являются паразитами человека.

Плоские черви (Platyhelminthes) – одна из древнейших групп многоклеточных животных. Их происхождение восходит к ранним этапам эволюции многоклеточных организмов. Исследования B. Egger с коллегами (2015) демонстрируют, что плоские черви появились более 500 млн лет назад в кембрийский период, что делает их одними из первых животных с организованной тканевой структурой. Наравне с моллюсками (Mollusca) и кольча-

тыми червями (Annelida) они входят в более широкую группу Lophotrochozoa (Egger et al., 2015; Laumer et al., 2015). При этом плоские черви часто рассматриваются как сестринская группа к моллюскам (Laumer et al., 2015), что подчеркивает их близкие эволюционные связи. Важность изучения биологии плоских червей связана с тем, что большинство их видов – паразиты, основные агенты гельминтозных заболеваний, передающихся через зараженную рыбу, поражающих значительное количество людей[1]. Многочисленные исследования показали, что продолжительные инфекции, такие как описторхоз, шистосомоз и подобные гельминтозы, могут привести к серьезным последствиям для здоровья организма-хозяина (Ogorodova et al., 2015; Pakharukova et al., 2019a; Carbonell et al., 2021), в том числе к развитию рака (Pakharukova et al., 2019b; Mordvinov et al., 2021).

У паразитических плоских червей ФА2 широко представлена в экскреторно-секреторных продуктах (ESP), которые выделяются для воздействия на хозяина (Wang et al., 2014), что указывает на потенциальную возможность патогенного воздействия этих ферментов на организм хозяина. Например, в ряде работ экспериментально показано, что фосфолипазы A, C, D плоского паразитического червя *Clonorchis sinensis* связаны с фиброзом у хозяина (Hu et al., 2009). Также выявлено, что фосфолипазы А2 группы G3 *C. sinensis* участвуют в процессах канцерогенеза у клеток хозяина (Shang et al., 2017). Однако в настоящее время существует лишь разрозненная информация о ФА2 у плоских червей, их представленности в геномах. Их функции у паразитов описаны недостаточно. Это подчеркивает необходимость более глубокого анализа и аннотации функций ФА2 у плоских червей, включая паразитических, для более полного понимания их роли в патогенезе и разработки эффективных методов борьбы с гельминтозными инфекциями.

Цель настоящей работы – изучение структуры, функций и эволюции фосфолипаз А2 у плоских червей. Для достижения этой цели была поставлена задача: идентифицировать белковые последовательности ФА2 у плоских червей, реконструировать филогению на основе этих последовательностей, провести анализ первичной структуры белков и предсказать их функции.

Материалы и методы

Вычислительный конвейер OrthoDom для идентификации ортологических групп белков с учетом доменной структуры. Для идентификации ортологических групп ФА2 у плоских червей с учетом доменной структуры мы используем информацию об эталонных последовательностях хорошо аннотированных ФА2 у модельных животных и вычислительный конвейер OrthoDom. Его схема приведена на рис. 1.

Конвейер OrthoDom позволяет выполнять поиск последовательностей семейств многодоменных белков на основе ортологии и анализа доменов среди аминокислотных последовательностей, кодируемых в геномах исследуемых организмов. В качестве входных данных (маркер 1 на рис. 1) использованы последовательности исследуемого семейства многодоменных белковых последовательностей с аннотацией (как правило, выявленных и аннотированных у модельных организмов). Для эталонных последовательностей указываются списки функциональных доменов, которые они включают. Для этих доменов из базы данных Pfam 33.1 (Mistry et al., 2021) извлекаются соответствующие профили HMM (маркер 2). Далее с помощью программы hmmsearch пакета HMMer 3.3.2 (Eddy, 2011) проводится валидация эталонных белков по наличию отмеченных доменов (маркер 3), так как для некоторых из них они могут быть фрагментированы или отсутствовать.

Еще один набор входных данных – аминокислотные последовательности (протеомы) исследуемых организмов (как правило, немодельных), в которых требуется установить ортологи эталонных белков (маркер 4). Ортологические группы для аминокислотных последовательностей эталонных белков и белков исследуемых организмов определялись программой OrthoFinder v. 2.5.4 (Emms, Kelly, 2019). Найденные с помощью нее ортологические группы (маркер 5), включающие эталонные последовательности, дополнительно проверялись на наличие доменов. Выявленные таким образом последовательности ортологов эталонных белков (маркер 6) обрабатывались далее для реконструкции филогении программой IQ-TREE (Nguyen et al., 2015). Филогенетические деревья визуализировались при помощи веб-версии программы iTOL (Letunic, Bork, 2024).

Референсные последовательности фосфолипаз А2 и их функциональные домены. Для идентификации ФА2 у плоских червей мы использовали фосфолипазы позвоночных, последовательности которых классифицированы по типам в ряде предыдущих работ и хорошо аннотированы. Эти белки рассматривались как эталонные и использовались для определения типа фосфолипаз в ортологических группах белков плоских червей. Выборка эталонных белков основана на последовательностях PLA2 из работы (Huang et al., 2015) (9 типов фосфолипаз человека и некоторых позвоночных животных). Эти последовательности были дополнены последовательностями из базы данных NCBI, идентифицированными на основе гомологии с помощью BLASTP (Турнаев и др., 2022). По классификации фосфолипаз А2, предложенной M. Murakami с коллегами (2020), из общего числа 16 групп фосфолипаз живых организмов в эталонную выборку вошли фосфолипазы 13 групп, поскольку группы фосфолипаз А2 11, 13 и 14 присутствуют только у растений (Murakami et al., 2020). В итоге эталонная выборка фосфолипаз А2 включала 13 групп ФА2 из 15 таксонов позвоночных животных. Список эталонных последовательностей из статей (Huang et al., 2015; Турнаев и др., 2022), тип фосфолипазы, видовое название организма, идентификатор, используемый в настоящей работе приведены в Приложении 1[2]. Список ключевых доменов этих белков и их моделей HMM представлен в Приложении 2.

[1] О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2014 году: Государственный доклад. М.: Роспотребнадзор, 2015

[2] Приложения 1–6 см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx28.pdf

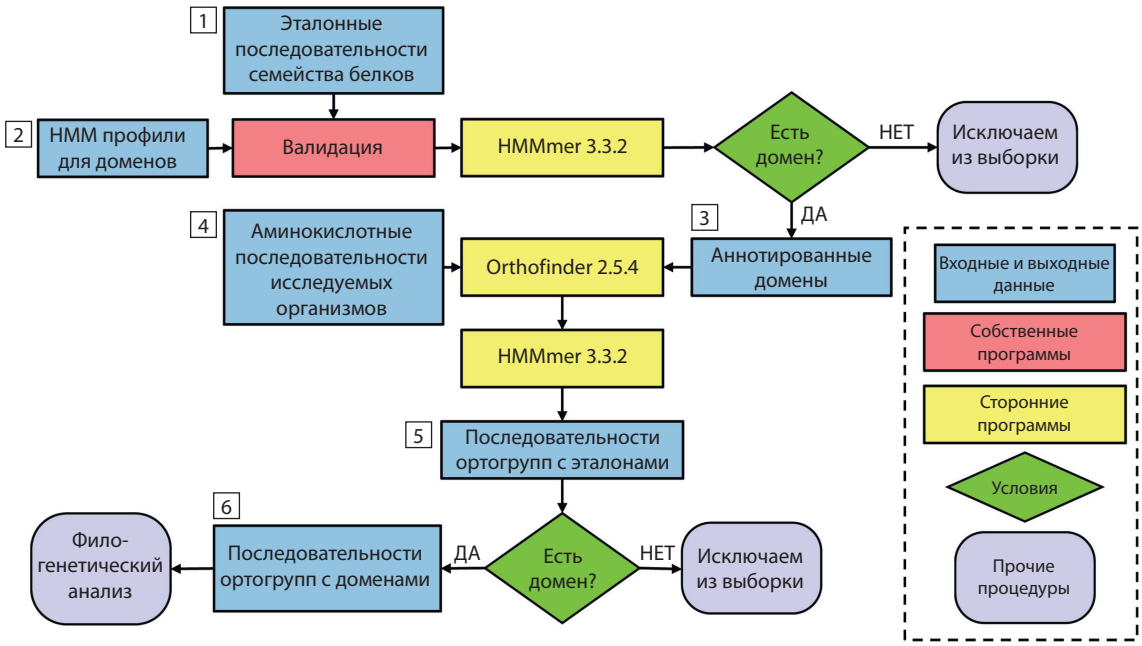


Рис. 1. Блок-схема вычислительного конвейера OrthoDom.

Для основных входных и выходных данных в прямоугольных рамках приведены цифровые маркеры (см. текст).

Поскольку ФА2 содержат не только фосфолипазные, но и другие характерные домены (Dennis et al., 2011), они тоже были определены уже после идентификации ортологов. Их список приведен в Приложении 3.

Источники геномных данных. Мы исследовали последовательности белок-кодирующих генов из геномов плоских червей 44 видов, представленных 2 свободноживущими и 42 паразитическими организмами. Анализировали аминокислотные последовательности, кодируемые мРНК соответствующих генов, имеющихся в базе данных Wormbase Parasite 18.0 (Howe et al., 2017). Эти виды включают основные таксоны плоских червей: класс дигенетических сосальщиков (Digenea), класс ленточных червей (Cestoda), класс моногенетических сосальщиков (Monogenea) и класс ресничных червей (Turbellaria) (Brusa et al., 2020). Среди перечисленных классов последний является классом свободноживущих червей, представители всех остальных классов – облигатные паразиты, причем моногенетические сосальщики – это энтопаразиты, а дигенетические сосальщики и цестоды – эндопаразиты. В качестве внешней группы при анализе мы использовали последовательности моллюсков из геномов тихоокеанской устрицы (*Crassostrea gigas*), морского блюдечка (*Lottia gigantea*) и филиппинской мидии (*Modiolus philippinarum*), поскольку известно, что моллюски – сестринская группа по отношению к плоским червям (Bernhard et al., 2015; Laumer et al., 2015). Аминокислотные последовательности моллюсков были взяты из базы данных MolluskDB 2.0 (Caurcel et al., 2021). Идентификаторы геномов плоских червей и моллюсков, видовые названия организмов и их типы, образ жизни представлены в Приложении 4.

Статистическая обработка результатов. Для оценки распространенности фосфолипаз различных ортологических групп у плоских червей, крупных таксонов Digenea, Cestoda, Monogenea и Turbellaria мы оценивали среднее количество последовательностей фосфолипаз для ортогруппы в геноме (n) и стандартное отклонение (σ). Среднее количество последовательностей в каждой ортогруппе по таксонам показывает, насколько часто встречаются последовательности фосфолипаз в исследуемых организмах. Стандартные отклонения демонстрируют разброс значений количеств последовательностей вокруг среднего. Чем больше стандартное отклонение, тем большее разнообразие в количестве последовательностей по таксонам. Дополнительно мы оценивали параметр f (представленность, %), долю организмов в крупном таксоне, которые содержат хотя бы одну из последовательностей ортогруппы. Если он равен 100 %, то все организмы таксона имеют по крайней мере одну последовательность из ортогруппы. Если какие-то организмы не содержат ни одной последовательности из ортогруппы фосфолипаз, то значение f меньше 100 %.

Результаты

В результате анализа, проведенного при помощи конвейера OrthoDom, у плоских червей было выделено 11 ортогрупп, включающих эталонные последовательности фосфолипаз А2. Отметим, что из всех групп ФА2, последовательности которых были использованы в качестве эталона, только последовательности группы 9 не обнаружили гомологии в протеомах моллюсков и плоских червей (они не вошли ни в одну из ортогрупп, определенных для этих организмов). Таким образом, согласно классификации M. Murakami с колегами (2020), в ортогруппы ФА2 моллюсков и плоских червей попали 12 из 13 известных типов фосфолипаз А2.

Характеристики встречаемости генов ортогрупп ФА2 у моллюсков и крупных таксонов плоских червей

Orthogroup ID	PLA2 type	PLA2 group	Mollusca (3)			Tricladida (2)			Monogenea (1)			Digenea (22)			Cestoda (19)		
			n	σ	*f*, %	*n*	σ	*f*, %	*n*	σ	*f*, %	*n*	σ	*f*, %	*n*	σ	*f*, %
OG0003047	Secreted	G1, 2, 5,10	2.0	1.0	100	5.0	2.8	100	1.0	0.0	100	0.0	0.0	0	0.0	0.0	0
OG0003722	Secreted	G3	4.7	3.8	100	3.0	1.4	100	0.0	0.0	0	1.0	1.9	41	0.0	0.0	0
OG0007610	Secreted	G12	1.0	0.0	100	0.5	0.7	50	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0
OG0000019*	Calcium-independent	G6	2.3	0.6	100	2.0	1.4	100	2.0	0.0	100	1.0	0.4	91	1.3	0.9	95
OG0000217*	Calcium-independent	G6	1.0	0.0	100	6.5	7.8	100	2.0	0.0	100	1.1	0.4	95	1.2	0.6	95
OG0000961*	Calcium-independent	G6	1.3	0.6	100	1.5	0.7	100	1.0	0.0	100	1.0	0.0	100	0.5	0.6	47
OG0007914	PAF	G7	1.0	0.0	100	4.5	6.4	50	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0
OG0004972*	PAF	G8	1.0	0.0	100	1.0	0.0	100	1.0	0.0	100	1.0	0.5	86	1.1	0.7	95
OG0000127	Cytosolic	G4	2.7	2.1	100	0.5	0.7	50	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0
OG0000135*	Lysosomal	G15	1.7	1.2	100	**8.5**	0.7	100	**3.0**	0.0	100	**4.7**	2.0	100	**2.1**	2.5	89
OG0007915	Adipo-specific	G16	3.0	1.0	100	1.0	1.4	50	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0

Примечание. Строки соответствуют различным ортогруппам фосфолипаз. Столбцы включают: индекс ортогруппы, тип и группы фосфолипаз, представленных в ней, статистику для исследованных крупных таксонов: средние значения (*n*), стандартные отклонения (σ) количества последовательностей в ортогруппах по таксонам, представленность (*f*) последовательностей у разных видов. В скобках рядом с названием таксона приведено количество видов. Максимальные средние значения *n* числа последовательностей в ортогруппах по таксонам показаны жирным шрифтом, минимальные подчеркнуты. Максимальные средние значения *n* числа последовательностей в каждой ортогруппе выделены серым фоном. Расширенная версия таблицы, содержащей информацию о представленности ортогрупп в разных таксонах, приведена в Приложении 5.
*Ортогруппы, последовательности которых представлены во всех крупных таксонах плоских червей.

В таблице показано распределение установленных ортогрупп, содержащих последовательности фосфолипаз плоских червей и моллюсков, и для них ряд статистических характеристик по представленности в пяти основных таксонах. Видно, что соответствие между ортогруппами и известными типами фосфолипаз не является однозначным. Также продемонстрировано, что часть ортогрупп включает несколько типов фосфолипаз. Например, ортогруппа OG0003047 содержит последовательности групп фосфолипаз 1, 2, 5, 10. С другой стороны, некоторые типы фосфолипаз были представлены несколькими ортогруппами. Например, последовательности ФА2 типа 6 распались на ортогруппы OG0000019, OG0000217, OG0000961. В остальных случаях каждая ортогруппа соответствовала одному типу и группе ФА2.

Следует отметить, что ортогруппы отличаются по числу последовательностей, которыми они представлены. Так, в ортогруппе OG0000135 – единственной группе лизосомальных фосфолипаз G15 – среднее число ортологов на протеом в каждом таксоне плоских червей максимальное, по сравнению с другими ортогруппами (от 2.1 у цестод до 8.5 у трикладид). У моллюсков эта группа фосфолипаз не относится по среднему числу последовательностей на протеом (1.7) к наиболее многочисленной; у этого таксона наиболее многочисленная группа OG00003722 – среднее число последовательностей 4.7 (секретируемые ФА2 G3).

Из таблицы также видно, что высокое среднее число последовательностей протеома, отнесенных к различным фосфолипазам, характерно для таксона Tricladida – свободноживущих, в отличие от остальных таксонов, являющихся паразитическими. Только в случае ортогруппы OG00004972 (тип PAF, группа G8) среднее число последовательностей на протеом у свободноживущих червей (1) меньше, чем у цестод (1.11), но, по крайней мере, не меньше, чем у остальных таксонов.

Выявленные нами ортогруппы в различных таксонах представлены неравномерно (см. таблицу). Во-первых, группы ФА2, которые встречаются во всех крупных таксонах плоских червей: это ФА2 кальций-независимого типа, а именно ортогруппы OG0000019, OG0000217, OG0000961 (шестая группа ФА2). При этом белки первых двух ортогрупп имеются в большинстве видов из крупных таксонов (более 90 %). Для ортогруппы OG0000961 характерно отсутствие ортологов для половины видов цестод. Для одной из этих групп, OG0000217, при этом среднее число белков у свободноживущих червей (6.5) в несколько раз превышает таковое у паразитических червей (1.1–2). Белки этой ортогруппы у цестод есть лишь в половине из изученных видов (среднее количество ФА2 на протеом – 0.5, стандартное отклонение – 0.6).

Другая ортогруппа, представители которой обнаруживаются во всех таксонах плоских червей, – OG0004972 (восьмая группа ФА2 тромбоцитарно-активирующего типа). Во всех крупных таксонах эти белки имеются у более чем 95 % видов, кроме дигенетических сосальщиков, у которых эта доля составляет 87 %. Эти гены встречаются в числе одной-двух копий на протеом.

Еще одна ортогруппа, входящая в состав всех крупных таксонов, OG0000135, включает лизосомальные ФА2 группы G15. Последовательности этой ортогруппы пред-

ставлены более чем одной копией на протеом, по сравнению с другими характеризуются наибольшим числом копий (см. выше).

Во-вторых, в таблице можно выделить ортогруппы, специфические для свободноживущих червей, гены которых полностью отсутствуют у паразитических червей. Эти ортогруппы разделились на четыре типа: секретируемые, ТАФ, цитозольные и адипо-специфичные: OG0007610, OG0007914, OG0000127, OG0007915 соответственно. Белки у всех этих ортогрупп имеются по крайней мере у одного из двух свободноживущих видов, исследованных нами.

В-третьих, в таблице показано наличие ортогрупп, специфических для отдельных паразитических таксонов. Например, ортогруппа OG0003047 (фосфолипазы групп G1,2,5,10) есть только у Monogenea (у всех видов); ортогруппа OG0003722 – только у дигенетических сосальщиков (примерно у половины видов). При этом у цестод представлено наименьшее количество ортогрупп фосфолипаз, в частности, отсутствуют все секретируемые фосфолипазы.

Таким образом, результаты позволяют заключить, что у свободноживущих червей обнаруживаются все найденные группы ФА2, причем большинство из них – со значительным количеством копий. Число генов в ортогруппах и число ортогрупп у паразитических червей редуцировано по отношению к свободноживущим. У моногеней присутствуют: одна ортогруппа, включающая секретируемые белки ФА2, все кальций-независимые ФА2; одна ортогруппа, включающая ТАФ, и одна ортогруппа, включающая лизосомальные фосфолипазы А2. У дигенетических сосальщиков из секретируемых фосфолипаз имеются белки, из ортогруппы, отличной от моногеней, и ортогруппы, включающей ТАФ и лизосомальные фосфолипазы. У цестод присутствуют все кальций-независимые ФА2, ТАФ и лизосомальные фосфолипазы А2, однако секретируемые ФА2 полностью отсутствуют. У различных таксонов паразитических червей есть как общие для всех них, так и специфические фосфолипазы.

Структурное разнообразие фосфолипаз

Проведенный анализ позволил охарактеризовать доменную структуру фосфолипаз разного типа у плоских червей. Для ряда фосфолипаз результаты показаны на рис. 2 и 3. На рис. 2 приведена доменная структура фосфолипаз из ортогруппы OG0003047, в которую попали эталонные белки групп ФА2, G1, G2, G5 и G10.

Из рис. 2 видно, что у последовательностей секретируемых ФА2 ортогруппы OG0003047 длина составляет примерно 200–250 аминокислот. Фосфолипазный домен занимает более 80 % от всего белка. Таким образом, первичная структура секретируемых ФА2 у плоских червей демонстрирует высокое сходство со структурами ФА2 человека соответствующих типов (Турнаев и др., 2022). Примечательно, что данная ортогруппа присутствует только у свободноживущих организмов.

Доменная организация последовательностей ортогрупп OG0000019, OG0000217 и OG0000961 продемонстрирована на рис. 3. Это ферменты, которые относятся к типу 6. Несмотря на то что ФА2 группы 6 разделилась на три

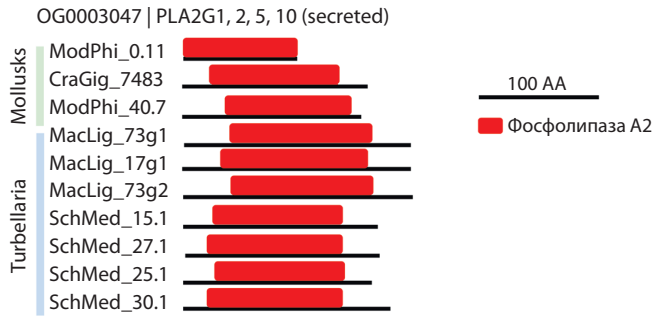


Рис. 2. Доменная структура последовательностей ортогруппы OG0003047, фосфолипазы А2 секретируемого типа.

Масштабная линейка, соответствующая 100 аминокислотам, изображена справа, фосфолипазный домен отмечен красным цветом. На рисунке представлено 10 последовательностей, выбранных случайным образом среди всех последовательностей ортогруппы OG0003047.

указанных ортогруппы, все они содержат ключевой для этой группы пататиновый домен (см. рис. 3). Доменная структура последовательностей ортогруппы OG0000019 соответствует типичной для ФА2 группы 6 подгруппе *A*, для которой характерны пататиновый домен и семь анкириновых доменов. Состав доменов белков ортогруппы OG0000217 соответствует типичной для ФА2 группы 6 подгруппе *C*, у которой в дополнение к пататиновому домену обнаруживаются три cNMP домена. Состав доменов последовательностей ортогруппы OG0000961 схож с типичными для ФА2 группы 6 подгруппами *D* и *E*, для которых характерен один лишь пататино-подобный фосфолипазный домен, расположенный на N-конце последовательности (Турнаев и др., 2022).

Таким образом, анализ функциональных доменов фосфолипаз показывает, что белки, относящиеся к фосфолипазам разного типа, но имеющие сходный доменный состав, формируют общую ортогруппу, а последовательности с различным доменным составом фосфолипаз даже одного типа распадаются на разные ортогруппы.

Филогенетический анализ фосфолипаз плоских червей

Для ортогрупп, доменная структура которых была представлена на рис. 2 и 3, мы реконструировали филогенетические деревья.

Последовательности ортогруппы OG0003047 были обнаружены у свободноживущих червей и одного представителя моногеней (см. рис. 4). У видов свободноживущих червей число последовательностей, относящихся к этой ортогруппе, высоко (см. таблицу), у представителя моногеней *Protopolystoma xenopodis* (краткое обозначение ProXen) наблюдается лишь один ген, кодирующий фосфолипазу данного типа.

Филогенетические деревья ортогрупп, содержащих ФА2 группы 6, приведены в Приложении 6. На рисунках кальций-независимой ФА2 группы 6 (см. Приложение 6, рис. 1–3) можно наблюдать сходные закономерности. Следует отметить, что гены паразитов отличаются высокой консервативностью. Если вернуться к рис. 3, то видно, что доменная структура последовательностей схожа у пред-

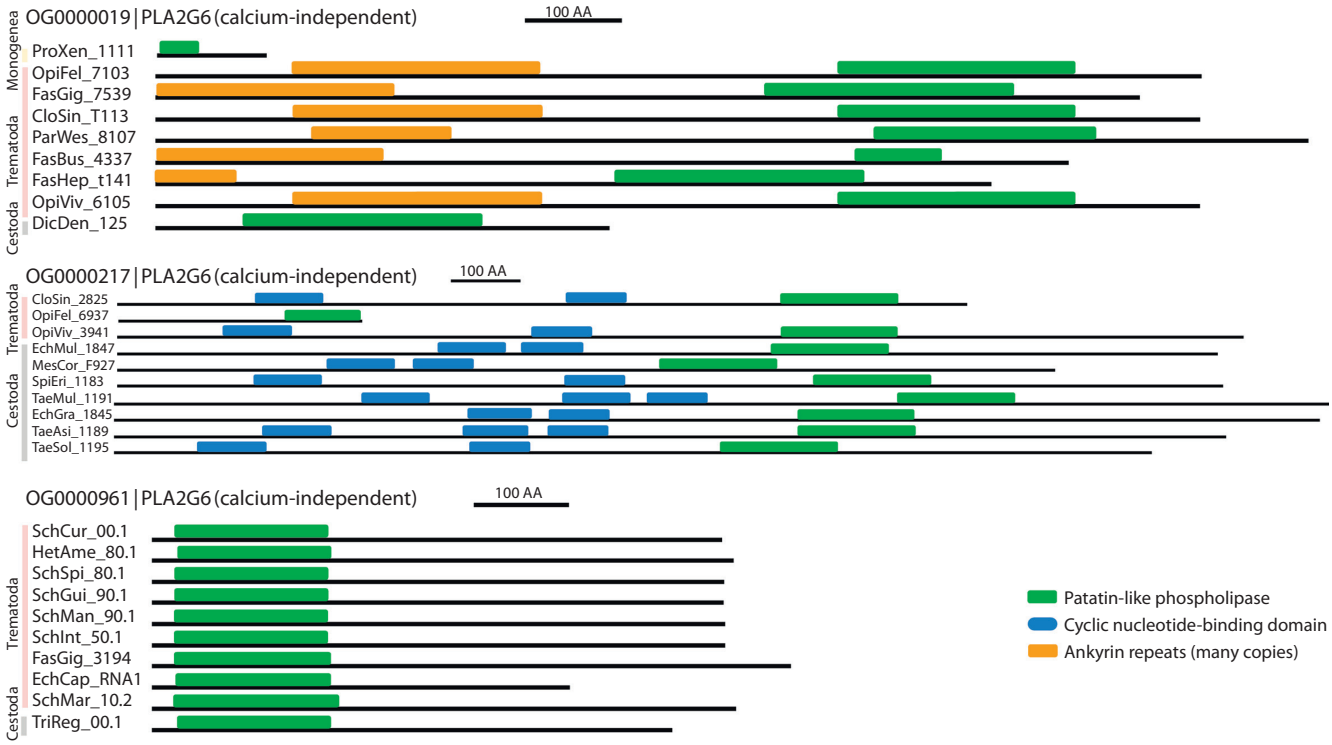


Рис. 3. Доменная структура последовательностей ортогрупп OG0000019, OG0000217, OG0000961 фосфолипазы А2 кальций-независимого типа.

Масштабная линейка соответствует 100 аминокислотам, пататино-подобный фосфолипазный домен отмечен зеленым цветом, домен cNMP – синим, анкириновые повторы – оранжевым. На рисунке представлено по 10 последовательностей (всего 30), выбранных случайным образом среди всех последовательностей ортогрупп OG0000019, OG0000217, OG0000961.

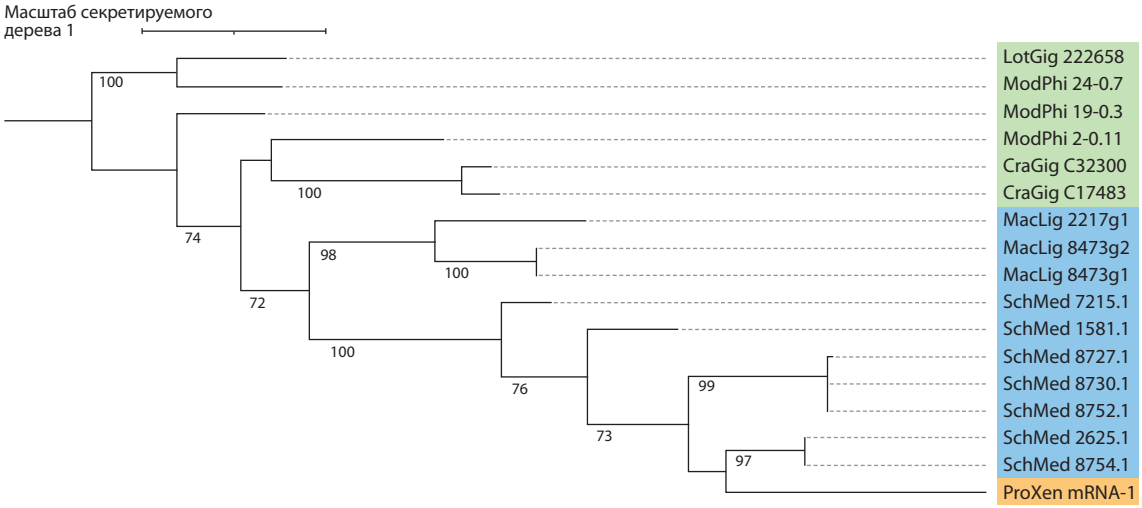


Рис. 4. Филогенетическое дерево последовательностей ФА2 ортогруппы OG0003047 (ФА2 G1, 2, 5, 10 секретируемого типа).

Выделены: зеленым цветом – последовательности моллюсков (Mollusca), синим – свободноживущих червей (Turbellaria), оранжевым – моногеней (Monogenea).

ставителей разных таксонов паразитов. Это позволяет сделать вывод, что ФА2 группы 6 является консервативным белком, играющим ключевую роль в базовых процессах жизнедеятельности паразитов плоских червей.

Необходимо отметить секретируемые фосфолипазы А2, которые относятся к ортогруппе OG0003722. Эта ортогруппа характерна тем, что у паразитических червей только таксон дигенетических сосальщиков содержит входящие в нее последовательности. Филогенетические дерево последовательностей, относящихся к этой ортогруппе, показано на рис. 5.

Из рис. 5 видно, что у свободноживущих червей обнаруживаются несколько копий гена ФА2 этой ортогруппы. У дигенетических сосальщиков также имеется несколько копий этого гена, которые распределяются по разным кладам. Это говорит о том, что для данной группы организмов характерны дупликации гена ФА2 группы G3, при этом копийные гены имеют различную структуру. Как

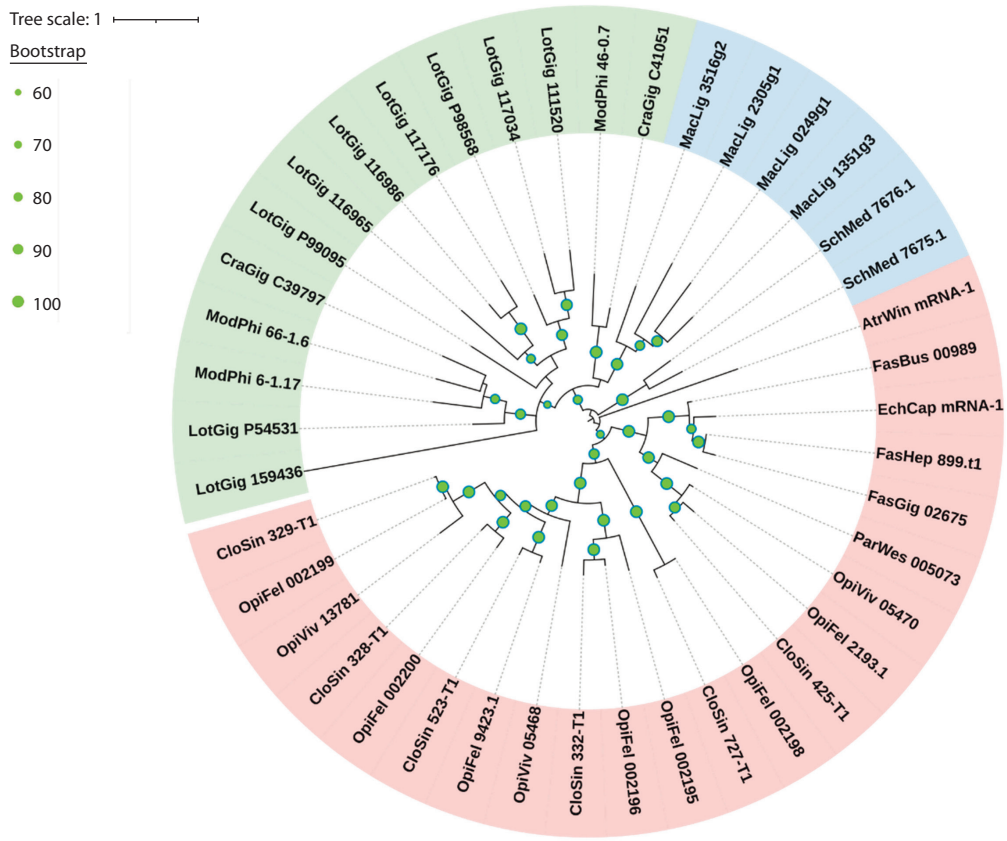


Рис. 5. Филогенетическое дерево круговой формы последовательностей ФА2 ортогруппы OG0003722 (ФА2 G3 секретируемого типа).

Выделены: зеленым цветом – моллюски (Mollusca), синим – свободноживущие черви (Turbellaria), красным – дигенетические сосальщики (Digenea).

правило, молекулярная эволюция у паразитов протекает значительно быстрее, по сравнению с представителями свободноживущих организмов (Trouvé et al., 1998). Анализ филогенетических деревьев подтверждает данное утверждение и для ФА2, где наблюдаются более длинные ветви у паразитов, что свидетельствует о высокой скорости эволюции данных молекул.

Обсуждение

Несмотря на то что фосфолипазы разных типов, и среди них ФА2, являются компонентами ESP паразитических плоских червей (Wang et al., 2014) и для ряда из них была продемонстрирована ассоциация с канцерогенезом у хозяина (Hu et al., 2009; Shang et al., 2017), они до сих пор недостаточно изучены для таксона Platyhelminthes (Dennis et al., 2011). При анализе обнаружены почти все известные группы фосфолипаз у плоских червей. Наш метод классификации дал возможность распределить их по подгруппам и объединить в ортогруппы с учетом доменной структуры. Эти результаты согласуются с классификацией фосфолипаз А2 и их доменной организацией, приведенными в работах E.A. Dennis с коллегами (2011). Метод идентификации ортологов, основанный на доменной структуре, показал свою эффективность в выделении ортогрупп белков с учетом различий в составе их доменов.

Наш анализ позволил идентифицировать ФА2 у плоских червей и показал, что в их эволюции можно выделить закономерности, которые характерны для эволюции геномов паразитов, например потери генов в связи с паразитическим образом жизни (Langleib et al., 2024). Действительно, в нашем исследовании выявлено, что у паразитических плоских червей редуцированы некоторые группы ФА2, а большинство генов представлено в одной копии (и не у всех видов крупных таксонов).

Сравнительный анализ ортогрупп генов ФА2 демонстрирует, что среди свободноживущих червей наблюдается относительно высокая степень дупликаций со средним числом паралогов на вид, достигающим пяти. Это явление предполагает наличие значительных адаптивных возможностей, которые могут быть обусловлены разнообразием экологических факторов. Свободноживущие организмы, подвергающиеся более высоким уровням экологической конкуренции, могут использовать это многообразие генов для повышения жизнеспособности и устойчивости к изменениям в окружающей среде. Однако в ортогруппе OG0000135, содержащей ФА2 лизосомального типа, дуплицированы гены даже у паразитических плоских червей. Чем вызвана эта аномалия, еще предстоит выяснить.

В ряде экспериментальных исследований установлено, что некоторые ФА2 могут участвовать в канцерогенезе, способствуя активации ряда клеточных сигнальных путей

и взаимодействию с иммунной системой организма-хозяина. Например, хроническая инфекция, вызванная *C. sinensis*, приводит к фиброзу печени и холангиокарциноме (Shang et al., 2017). Причем *C. sinensis* использует ФА2 группы 3 в качестве ESP, который играет важную роль в патогенезе печени (Wu et al., 2021). В результате проведенного исследования установлено, что среди секретируемых фосфолипаз дигенетических сосальщиков присутствует только ФА2 группы 3, тогда как у цестод секреция ФА2 отсутствует. Учитывая, что паразитические плоские черви способны манипулировать метаболизмом хозяев, используя фосфолипазы для извлечения необходимых ресурсов, можно предположить, что аналогичные механизмы могут работать и в раковых клетках.

Заключение

Фосфолипазы А2 – это семейство гидролаз, которые катализируют процесс гидролиза фосфолипидов, играя ключевую роль во многих молекулярных процессах при функционировании клеток и организма в целом. Их разнообразие у плоских червей было слабо изучено, и в нашей работе впервые проведен такой анализ. Мы установили, что у свободноживущих червей представлены 12 из 13 известных типов фосфолипаз А2. У этих организмов наблюдается повышенное число копий генов по сравнению с паразитическими червями. Были выявлены уникальные особенности некоторых ортогрупп, которые, вероятно, могут быть связаны с канцерогенезом у хозяина, вызываемым паразитической инфекцией.

Список литературы / References

Турнаев И.И., Бочарникова М.Е., Афонников Д.А. Фосфолипазы А2 человека: функциональный и эволюционный анализ. *Вавиловский журнал генетики и селекции*. 2022;26(8):787-797. doi 10.18699/VJGB-22-95
[Turnaev I.I., Bocharnikova M.E., Afonnikov D.A. Human phospholipases A2: a functional and evolutionary analysis. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2022;26(8):787-797. doi 10.18699/VJGB-22-95]

Филькин С.Ю., Липкин А.В., Федоров А.Н. Суперсемейство фосфолипаз: структура, функции и применение в биотехнологии. *Успехи биол. химии*. 2020;60:369-410
[Filkin S.Yu., Lipkin A.V., Fedorov A.N. Phospholipase superfamily: structure, functions, and biotechnological applications. *Uspekhi Biologicheskoi Khimii = Biochemistry* (*Moscow*). 2020;85(Suppl.1): S177S195. DOI 10.1134/S0006297920140096]

Bitar L., Jundi D., Rima M., Al Alam J., Sabatier J.M., Fajloun Z. Bee venom PLA2 *versus* snake venom PLA2: Evaluation of structural and functional properties. *Venoms Toxins.* 2021;2(1):22-33. doi 10.2174/2666121701999210101225032

Brusa F., Leal-Zanchet A.M., Noreña C., Damborenea C. Phylum Platyhelminthes. In: Thorp and Covich's Freshwater Invertebrates. Ch. 5. Academic Press, 2020;101-120. doi 10.1016/B978-0-12-804225-0.00005-8

Carbonell C., Rodríguez-Alonso B., López-Bernús A., Almeida H., Galindo-Pérez I., Velasco-Tirado V., Belhassen-García M. Clinical spectrum of schistosomiasis: an update. *J. Clin. Med.* 2021;10(23): 5521. doi 10.3390/jcm10235521

Caurcel C., Laetsch D.R., Challis R., Kumar S., Gharbi K., Blaxter M. MolluscDB: a genome and transcriptome database for molluscs. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2021;376(1825):20200157. doi 10.1098/rstb.2020.0157

Dennis E.A., Cao J., Hsu Y.-H., Magrioti V., Kokotos G. Phospholipase A_2 enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention. *Chem. Rev.* 2011;111(10):6130-6185. doi 10.1021/cr200085w

Eddy S.R. Accelerated profile HMM searches. *PLoS Comput. Biol.* 2011;7(10):e1002195. doi 10.1371/journal.pcbi.1002195

Egger B., Lapraz F., Tomiczek B., Müller S., Dessimoz C., Girstmair J., Telford M.J. A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. *Curr. Biol.* 2015;25(10):1347-1353. doi 10.1016/j.cub.2015.03.034

Emms D.M., Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. doi 10.1186/s13059-019-1832-y

Gutiérrez J.M., Lomonte B. Phospholipases A_2: unveiling the secrets of a functionally versatile group of snake venom toxins. *Toxicon.* 2013; 62:27-39. doi 10.1016/j.toxicon.2012.09.006

Howe K.L., Bolt B.J., Shafie M., Kersey P., Berriman M. WormBase ParaSite − a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 2017;215:2-10. doi 10.1016/j.molbiopara.2016.11.005

Hu F., Hu X., Ma C., Zhao J., Xu J., Yu X. Molecular characterization of a novel *Clonorchis sinensis* secretory phospholipase A_2 and investigation of its potential contribution to hepatic fibrosis. *Mol. Biochem. Parasitol.* 2009;167(2):127-134. doi 10.1016/j.molbiopara.2009.05.003

Huang Q., Wu Y., Qin C., He W., Wei X. Phylogenetic and structural analysis of the phospholipase A_2 gene family in vertebrates. *Int. J. Mol. Med.* 2015;35(3):587-596. doi 10.3892/ijmm.2014.2047

Langleib M., Calvelo J., Costábile A., Castillo E., Tort J.F., Hoffmann F.G., Iriarte A. Evolutionary analysis of species-specific duplications in flatworm genomes. *Mol. Phylogenet. Evol.* 2024;199: 108141. doi 10.1016/j.ympev.2024.108141

Laumer C.E., Hejnol A., Giribet G. Nuclear genomic signals of the 'microturbellarian' roots of platyhelminth evolutionary innovation. *eLife.* 2015;4:e05503. doi 10.7554/eLife.05503

Letunic I., Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024;52(W1):W78-W82. doi 10.1093/nar/gkae268

McIntosh J.M., Ghomashchi F., Gelb M.H., Dooley D.J., Stoehr S.J., Giordani A.B., Olivera B.M. Conodipine-M, a novel phospholipase A_2 isolated from the venom of the marine snail *Conus magus*. *J. Biol. Chem.* 1995;270(8):3518-3526. doi 10.1074/jbc.270.8.3518

Mistry J., Chuguransky S., Williams L., Qureshi M., Salazar G.A., Sonnhammer E.L., Bateman A. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412-D419. doi 10.1093/nar/gkaa913

Mordvinov V.A., Minkova G.A., Kovner A.V., Ponomarev D.V., Lvova M.N., Zaparina O., Pakharukova M.Y. A tumorigenic cell line derived from a hamster cholangiocarcinoma associated with *Opisthorchis felineus* liver fluke infection. *Life Sci.* 2021;277:119494. doi 10.1016/j.lfs.2021.119494

Mouchlis V.D., Dennis E.A. Membrane association allosterically regulates phospholipase A_2 enzymes and their specificity. *Acc. Chem. Res.* 2022;55(23):3303-3311. doi 10.1021/acs.accounts.2c00497

Murakami M., Sato H., Taketomi Y. Updating phospholipase A_2 biology. *Biomolecules.* 2020;10(10):1457. doi 10.3390/biom10101457

Murase R., Taketomi Y., Miki Y., Nishito Y., Saito M., Fukami K., Murakami M. Group III phospholipase A_2 promotes colitis and colorectal cancer. *Sci. Rep.* 2017;7(1):12261. doi 10.1038/s41598-017-12434-z

Nevalainen T.J., Cardoso J.C., Riikonen P.T. Conserved domains and evolution of secreted phospholipases A_2. *FEBS J.* 2012;279(4): 636-649. doi 10.1111/j.1742-4658.2011.08453.x

Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 2015;32(1):268-274. doi 10.1093/molbev/msu300

Ogorodova L.M., Fedorova O.S., Sripa B., Mordvinov V.A., Katokhin A.V., Keiser J.; TOPIC Consortium. Opisthorchiasis: an overlooked danger. *PLoS Negl. Trop. Dis*. 2015:9(4):e0003563. doi 10.1371/journal.pntd.0003563

Pakharukova M.Y., Zaparina O.G., Kapushchak Y.K., Baginskaya N.V., Mordvinov V.A. *Opisthorchis felineus* infection provokes time-dependent accumulation of oxidative hepatobiliary lesions in the injured hamster liver. *PLoS One.* 2019a;14(5):e0216757. doi 10.1371/journal.pone.0216757

Pakharukova M.Y., da Costa J.M.C., Mordvinov V.A. The liver fluke *Opisthorchis felineus* as a group III or group I carcinogen. *4open.* 2019b;2:23. doi 10.1051/fopen/2019016

Park J.B., Lee C.S., Jang J.H., Ghim J., Kim Y.J., You S., Ryu S.H. Phospholipase signalling networks in cancer. *Nat. Rev. Cancer.* 2012;12(11):782-792. doi 10.1038/nrc3379

Salabi F., Jafari H. Whole transcriptome sequencing reveals the activity of the PLA_2 family members in *Androctonus crassicauda* (Scorpionida: Buthidae) venom gland. *FASEB J.* 2024;38(10):e23658. doi 10.1096/fj.202400178RR

Scott K.F., Sajinovic M., Hein J., Nixdorf S., Galettis P., Liauw W., Russell P.J. Emerging roles for phospholipase A_2 enzymes in cancer. *Biochimie.* 2010;92(6):601-610. doi 10.1016/j.biochi.2010.03.019

Shang M., Xie Z., Tang Z., He L., Wang X., Wang C., Li X. Expression of *Clonorchis sinensis* GIIIsPLA$_2$ protein in baculovirus-infected insect cells and its overexpression facilitating epithelial-mesenchymal transition in Huh7 cells via AKT pathway. *Parasitol. Res.* 2017;116:1307-1316. doi 10.1007/s00436-017-5409-y

Teixeira S.C., da Silva M.S., Gomes A.A.S., Moretti N.S., Lopes D.S., Ferro E.A.V., de Melo Rodrigues V. Panacea within a Pandora's box: the antiparasitic effects of phospholipases A_2 (PLA_2s) from snake venoms. *Trends Parasitol.* 2022;38(1):80-94. doi 10.1016/j.pt.2021.07.004

Trouvé S., Sasal P., Jourdane J., Renau F., Morand S. The evolution of life-history traits in parasitic and free-living platyhelminthes: a new perspective. *Oecologia*. 1998;115:370-378. doi 10.1007/s004420050530

Wang X., Hu F., Hu X., Chen W., Huang Y., Yu X. Proteomic identification of potential *Clonorchis sinensis* excretory/secretory products capable of binding and activating human hepatic stellate cells. *Parasitol. Res.* 2014;113:3063-3071. doi 10.1007/s00436-014-3972-z

Wu Y.J., He Q., Shang M., Yin Y.X., Li Y., Du X., Li X.R. The NF-κB signalling pathway and TM7SF3 contribute to liver fibrosis caused by secreted phospholipase A2 of *Clonorchis sinensis*. *Parasit. Vectors.* 2021;14:1-9. doi 10.1186/s13071-021-04663-z

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 21.10.2024. После доработки 25.11.2024. Принята к публикации 26.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):864-873

doi 10.18699/vjgb-24-94

Реконструкция и компьютерный анализ структурно-функциональной организации генной сети регуляции биосинтеза холестерина у человека и эволюционная характеристика участвующих в ней генов

А.Д. Михайлова[1], С.А. Лашин[1, 2, 3], В.А. Иванисенко[1, 2, 3], П.С. Деменков[1, 2, 3], Е.В. Игнатьева[1, 2]

[1] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

eignat@bionet.nsc.ru

Аннотация. Холестерин – это незаменимая структурная компонента клеточных мембран, предшественник витамина D и стероидных гормонов. У человека и других видов животных холестерин поступает в организм с пищей, а также синтезируется в клетках многих тканей *de novo*. Ранее нами была реконструирована генная сеть регуляции внутриклеточного уровня холестерина, включавшая регуляторные контуры, функционирующие при участии транскрипционных факторов подсемейства SREBP (sterol regulatory element-binding proteins). Активность транскрипционных факторов подсемейства SREBP регулируется в обратной зависимости от уровня холестерина в клетке. Этот механизм реализуется при участии белков «холестеринового сенсора», включающего белки SCAP, INSIG1, INSIG2, MBTPS1/S1P, MBTPS2/S2P и транскрипционные факторы подсемейства SREBP. Повышенный уровень холестерина является фактором риска сердечно-сосудистых заболеваний, а также сопутствующим фактором многих патологических состояний. Систематизация сведений о молекулярных механизмах, контролирующих активность факторов подсемейства SREBP и биосинтез холестерина, в формате генной сети и получение новых знаний о генной сети как едином объекте чрезвычайно важны в контексте понимания молекулярных механизмов развития заболеваний. Средствами компьютерной системы ANDSystem нами построена генная сеть регуляции биосинтеза холестерина в клетке. Генная сеть включает данные: (1) о ферментах, осуществляющих биосинтез холестерина; (2) белках, функционирующих в составе «холестеринового сенсора»; (3) белках, регулирующих активность белков «холестеринового сенсора»; (4) генах, кодирующих белки этих групп; (5) генах, транскрипция которых регулируется при участии транскрипционных факторов подсемейства SREBP (генах-мишенях). Проведен анализ генной сети и выявлены замкнутые регуляторные контуры, контролирующие активность транскрипционных факторов подсемейства SREBP. Эти контуры реализуются с участием генов *PPARG*, *NR0B2*/*SHP1*, *LPIN1*, *AR* и кодируемых ими белков. Исследование филостратиграфического возраста генов показало, что предковые формы большинства генов человека, кодирующих ферменты биосинтеза холестерина и белки «холестеринового сенсора», могли возникнуть на достаточно ранних эволюционных этапах (*Cellular organisms* (корень филостратиграфического дерева) и этапах дивергенции *Eukaryota* и *Metazoa*). Однако механизм регуляции транскрипции генов в ответ на изменение уровня холестерина мог сформироваться только на более поздних эволюционных этапах, поскольку филостратиграфический возраст генов транскрипционных факторов подсемейства SREBP соответствует более позднему этапу эволюции (стадии дивергенции *Vertebrata*).

Ключевые слова: биосинтез холестерина; транскрипционные факторы; SREBP; генные сети; регуляторые обратные связи; эволюция; филостратиграфия; возраст гена.

Для цитирования: Михайлова А.Д., Лашин С.А., Иванисенко В.А., Деменков П.С., Игнатьева Е.В. Реконструкция и компьютерный анализ структурно-функциональной организации генной сети регуляции биосинтеза холестерина у человека и эволюционная характеристика участвующих в ней генов. *Вавиловский журнал генетики и селекции*. 2024;28(8):864-873. doi 10.18699/vjgb-24-94

Финансирование. Исследование выполнено за счет средств бюджетного проекта ФИЦ ИЦиГ СО РАН «Системная биология и биоинформатика: реконструкция, анализ и моделирование структурно-функциональной организации и эволюции генных сетей человека, животных, растений и микроорганизмов», № FWNR-2022-0020.

Благодарности. Авторы выражают благодарность З.С. Мустафину, предоставившему данные по индексам PAI для белок-кодирующих генов человека.

© Михайлова А.Д., Лашин С.А., Иванисенко В.А., Деменков П.С., Игнатьева Е.В., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

Reconstruction and computer analysis of the structural and functional organization of the gene network regulating cholesterol biosynthesis in humans and the evolutionary characteristics of the genes involved in the network

A.D. Mikhailova[1], S.A. Lashin [1, 2, 3], V.A. Ivanisenko [1, 2, 3], P.S. Demenkov [1, 2, 3], E.V. Ignatieva [1, 2]

[1] Novosibirsk State University, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
eignat@bionet.nsc.ru

Abstract. Cholesterol is an essential structural component of cell membranes and a precursor of vitamin D, as well as steroid hormones. Humans and other animal species can absorb cholesterol from food. Cholesterol is also synthesized *de novo* in the cells of many tissues. We have previously reconstructed the gene network regulating intracellular cholesterol levels, which included regulatory circuits involving transcription factors from the SREBP (sterol regulatory element-binding proteins) subfamily. The activity of SREBP transcription factors is regulated inversely depending on the intracellular cholesterol level. This mechanism is implemented with the participation of proteins SCAP, INSIG1, INSIG2, MBTPS1/S1P and MBTPS2/S2P. This group of proteins, together with the SREBP factors, is designated as "cholesterol sensor". An elevated cholesterol level is a risk factor for the development of cardiovascular diseases and may also be observed in obesity, diabetes and other pathological conditions. Systematization of information on the molecular mechanisms controlling the activity of SREBP factors and cholesterol biosynthesis in the form of a gene network and building new knowledge about the gene network as a single object is extremely important for understanding the molecular mechanisms underlying the predisposition to diseases. With a computer tool, ANDSystem, we have built a gene network regulating cholesterol biosynthesis. The gene network included data on: (1) the complete set of enzymes involved in cholesterol biosynthesis; (2) proteins that function as part of the "cholesterol sensor"; (3) proteins that regulate the activity of the "cholesterol sensor"; (4) genes encoding proteins of these groups; (5) genes whose transcription is regulated by SREBP factors (SREBP target genes). The gene network was analyzed and feedback loops that control the activity of SREBP factors were identified. These feedback loops involved the *PPARG*, *NR0B2/SHP1*, *LPIN1*, and *AR* genes and the proteins they encode. Analysis of the phylostratigraphic age of the genes showed that the ancestral forms of most human genes encoding the enzymes of cholesterol biosynthesis and the proteins of the "cholesterol sensor" may have arisen at early evolutionary stages (*Cellular organisms* (the root of the phylostratigraphic tree) and the stages of *Eukaryota* and *Metazoa* divergence). However, the mechanism of gene transcription regulation in response to changes in cholesterol levels may only have formed at later evolutionary stages, since the phylostratigraphic age of the genes encoding the transcription factors SREBP1 and SREBP2 corresponds to the stage of *Vertebrata* divergence.

Key words: cholesterol biosynthesis; transcription factors; SREBP; gene networks; feedback loops; evolution; phylostratigraphy; gene age.

For citation: Mikhailova A.D., Lashin S.A., Ivanisenko V.A., Demenkov P.S., Ignatieva E.V. Reconstruction and computer analysis of the structural and functional organization of the gene network regulating cholesterol biosynthesis in humans and the evolutionary characteristics of the genes involved in the network. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):864-873. doi 10.18699/vjgb-24-94

Введение

Холестерин является одним из важнейших соединений в организме животных. Он присутствует во всех тканях организма животных, поскольку входит в состав липидного слоя клеточных мембран, стабилизируя структуру мембран (Koolman, Roehm, 2005). При повышении содержания холестерина мембрана становится более плотно упакованной, содержит меньше полостей, за счет чего ее проницаемость для малых молекул, включая кислород, снижается. Данный механизм способствовал адаптации организмов к атмосфере, содержащей большое количество кислорода, и, как следствие, защите клетки от окислительного стресса (Zuniga-Hertz, Patel, 2019). Интересно отметить, что у грибов и растений холестерин не синтезируется, а клеточная мембрана этих организмов содержит похожие по структуре соединения – эргостерол (у грибов) и β-ситостерин и стигмастерол (у растений) (Desmond, Gribaldo, 2009; Ferrer et al., 2017; Choy et al., 2023).

У животных холестерин выполняет и другие важные функции. Это вещество является предшественником желчных кислот и стероидных гормонов: прогестерона, эстрадиола, тестостерона, кальцитриола, кортизола (Luo et al., 2020; Schade et al., 2020).

У человека и других видов животных холестерин поступает в организм с пищей, а также синтезируется в клетках многих тканей *de novo* (Luo et al., 2020). Исходными метаболитами для синтеза холестерина служат ацетил-КоА и ацетоацетил-КоА, всего в процессе биосинтеза участвует более 20 ферментов (Desmond, Gribaldo, 2009; Nes, 2011). Промежуточные продукты пути биосинтеза холестерина также выполняют важные функции в клетках животных. Например, геранил-геранилпирофосфат и фарнезилпи-

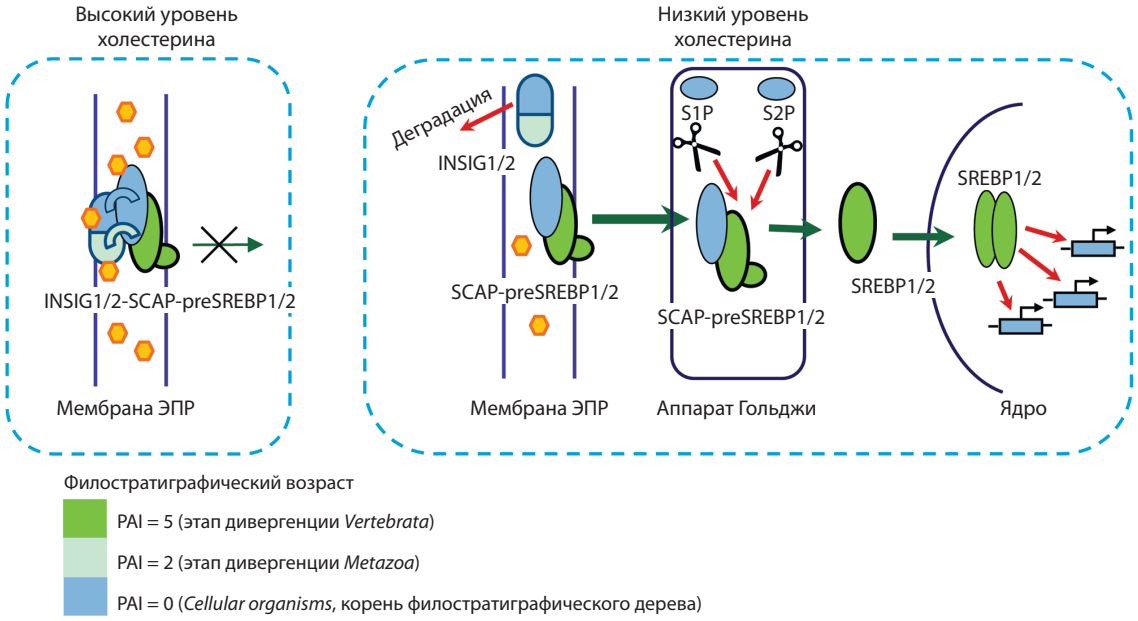


Рис. 1. Функционирование «холестеринового сенсора».

Желтыми шестиугольниками показаны молекулы холестерина; INSIG1/2 – якорные белки INSIG1 и INSIG2; SREBP1/2 – транскрипционные факторы SREBP1 и SREBP2; preSREBP1/2 – белки preSREBP1 и preSREBP2, которые являются неактивными предшественниками транскрипционных факторов SREBP1 и SREBP2; SCAP – белок-шаперон, взаимодействующий с белками preSREBP1 и preSREBP2; S1P и S2P – белки-протеазы, которые кодируются генами *MBTPS1* и *MBTPS2*. Цвет объектов соответствует филостратиграфическому возрасту генов, оценка которого производилась на основе индекса PAI (процедура расчета PAI описана в разделе «Материалы и методы»). При высоком уровне холестерина (левая часть рисунка) холестерин стабилизирует стуктуру белков INSIG1 и INSIG2 (обозначены как INSIG1/2), укрепляются связи между белками INSIG1/2 и SCAP. Якорные белки INSIG1 и INSIG2 удерживают комплекс SCAP-preSREBP1/2 в мембране эндоплазматического ретикулума (ЭПР). При низком уровне холестерина (правая часть рисунка) происходят убиквитинирование, а затем и деградация белков INSIG1 и INSIG2. Нарушается связь между INSIG1/2 и белком SCAP. Комплекс SCAP-preSREBP1/2 перемещается в аппарат Гольджи, где идет расщепление белков preSREBP1/2 протеазами S1P и S2P. В результате расщепления белков preSREBP1 и preSREBP2 образуются активные транскрипционные факторы SREBP1 и SREBP2 (обозначены как SREBP1/2). Описание рисунка составлено по (DeBose-Boyd, Ye, 2018; Jiang et al., 2020).

рофосфат – субстраты в реакциях пренилирования. Этой посттрансляционной модификации подвергаются многие белки (например, малые ГТФ-связывающие белки). Такое посттрансляционное пренилирование различных белков является важным для правильной локализации и активации этих белков (Waller et al., 2019).

Ранее в Институте цитологии и генетики СО РАН была реконструирована генная сеть регуляции внутриклеточного уровня холестерина и установлено четыре регуляторных контура, функционирующих при участии транскрипционных факторов подсемейства SREBP (sterol regulatory element-binding proteins) (Колчанов и др., 2013; Меркулова и др., 2013). Активность транскрипционных факторов подсемейства SREBP регулируется в обратной зависимости от уровня холестерина в клетке. Этот механизм реализуется при участии ряда белков, которые в совокупности с транскрипционными факторами подсемейства SREBP образуют так называемый холестериновый сенсор. Схема работы «холестеринового сенсора» представлена на рис. 1.

Активность транскрипционных факторов подсемейства SREBP может также регулироваться в ответ на сигналы, которые воздействуют на клетку извне, например инсулин и ростовые факторы (Sundqvist et al., 2005; Arito et al., 2008; Peterson et al., 2011). За счет такой регуляции осуществляется тонкая настройка функционирования факторов подсемейства SREBP в зависимости от состояния клетки и организма в целом. В свою очередь, факторы подсемейства SREBP контролируют экспрессию белков, участвующих в регуляции большого количества биологических процессов, объединяя локальные генные сети, контролирующие отдельные аспекты функционирования клеток (Jeon, Osborne, 2012).

Повышенный уровень холестерина является фактором риска развития сердечно-сосудистых заболеваний (атеросклероза, ишемической болезни сердца) (Vargas-Alarcon et al., 2019; Macvanin et al., 2024), а также может выступать в качестве сопутствующего фактора при ожирении (Kim et al., 2010), диабете (Zhang F. et al., 2018), жировом перерождении печени (Li et al., 2023), гепатокарциноме (Paul et al., 2022), опухолевых процессах (Jiang et al., 2020) и воспалении (Shimano, Sato, 2017). Получение новых знаний о генной сети регуляции биосинтеза холестерина как едином объекте чрезвычайно важно в контексте понимания связи данной системы с заболеваниями.

Цель настоящей работы – систематизация сведений о молекулярных механизмах, контролирующих активность факторов подсемейства SREBP и биосинтез холестерина, в формате генной сети и последующий анализ структурно-функциональной организации сети и эволюционных характеристик участвующих в ней генов.

Материалы и методы

Списки генов, использованные для построения генной сети. Список, включающий 24 гена человека, кодирующих ферменты биосинтеза холестерина (Приложение 1)[1], был составлен на основе данных базы WikiPathways (Agrawal et al., 2024).

Список, содержащий семь генов, кодирующих белки «холестеринового сенсора» (Приложение 2), формировали на основе описания механизма регуляции активности факторов SREBP1 и SREBP2 по данным, приведенным в публикациях (DeBose-Boyd, Ye, 2018; Jiang et al., 2020).

Список, включающий 31 ген человека, транскрипция которых регулируется факторами подсемейства SREBP (гены-мишени факторов SREBP1 и SREBP2), формировали на основе объединения данных базы TRRD (Kolchanov et al., 2002) и TRRUST (https://www.grnpedia.org/trrust/) (Han et al., 2018). В финальную версию списка генов-мишеней (Приложение 3) были включены гены, имевшие связи с белками SREBP1 и SREBP2, по данным системы ANDSystem (Ivanisenko et al., 2019).

Список генов, кодирующих белки, регулирующие активность белков и генов «холестеринового сенсора» («белки-регуляторы») (Приложение 4), формировали на основе данных системы ANDSystem (Ivanisenko et al., 2019). «Белки-регуляторы» были найдены в результате исполнения команды «Мастер путей» в программе ANDVisio (часть ANDSystem). Полученные таким образом связи между «белками-регуляторами» и белками либо генами «холестеринового сенсора» были верифицированы вручную.

Построение генной сети регуляции биосинтеза холестерина. Построение генной сети осуществляли с помощью компьютерной системы ANDSystem (Ivanisenko et al., 2019). На начальном этапе были построены генные сети, включающие небольшие группы генов (называемые нами в дальнейшем термином «малые генные сети»). Процедуры построения «малых сетей» описаны в Приложении 5. Информация о количестве объектов в сетях представлена в Приложении 6. Затем эти «малые генные сети» были объединены в общую сеть с помощью команды «Объединение графов» программы ANDVisio. Были объединены «малые» генные сети, включающие следующие связи: (1) от белков, регулирующих активность «холестеринового сенсора», к генам и белкам «холестеринового сенсора»; (2) от факторов подсемейства SREBP к генам-мишеням и от генов-мишеней к кодируемым ими белкам; (3) от белков, кодируемых генами-мишенями факторов подсемейства SREBP, к генам и белкам «холестеринового сенсора»; (4) от генов/белков «холестеринового сенсора» (кроме SREBP1 и SREBP2) к генам *SREBF1*, *SREBF2* и кодируемым белкам (SREBP1, SREBP2); (5) от ферментов биосинтеза холестерина к холестерину.

Поиск замкнутых регуляторных контуров. Поиск замкнутых регуляторных контуров длины 3, 4, 5 с участием факторов SREBP1 и SREBP2 выполняли с помощью команды «Мастер путей» в программе ANDVisio. При поиске ориентировались на шаблоны, представленные в Приложении 7. Беря за основу шаблон определенной длины (что соответствовало количеству объектов в регуляторном контуре), устанавливали количество и типы промежуточных объектов при поиске путей. Найденные пути дополняли связями между генами и кодируемыми белками (т. е. связями типа «экспрессия»), получая таким образом замкнутые регуляторные контуры.

Выявление тканей, где возможно функционирование регуляторных контуров. Использованы данные проекта GTEx (GTEx Consortium, 2020), экстрагированные из базы Expression Atlas (https://www.ebi.ac.uk/gxa/home). Были выбраны примеры тканей либо органов, где уровень экспрессии каждого гена, входящего в рассматриваемый контур, был не ниже 10 TPM (transcripts per million).

Анализ эволюционных характеристик генов. Анализ эволюционных характеристик генов осуществляли с использованием индекса PAI (phylostratigraphic age index). Величины PAI были рассчитаны для 19556 белок-кодирующих генов человека с помощью программы Orthoscape (Mustafin et al., 2017), как это описано в работе (Мустафин и др., 2021).

Результаты и обсуждение

Генная сеть регуляции биосинтеза холестерина

На первом этапе с помощью программы ANDVisio были построены так называемые «малые генные сети» (процедура построения описана в разделе «Материалы и методы» и Приложении 5). Далее «малые генные сети» объединяли с помощью программы ANDVisio. Таким образом, была получена генная сеть регуляции биосинтеза холестерина (рис. 2). Сеть включает: (1) гены *SREBF1*, *SREBF2* и кодируемые ими белки; (2) пять белков, регулирующих активность факторов SREBP1 и SREBP2 (INSIG1, INSIG2, SCAP, MBTPS1/S1P, MBTPS2/S2P), и кодирующие их гены («холестериновый сенсор»); (3) 62 белка-регулятора генов и белков «холестеринового сенсора»; (4) 31 ген-мишень факторов подсемейства SREBP (с учетом самого *SREBF2*) и кодируемые ими белки; (5) 243 взаимодействия между объектами (см. рис. 2).

Замкнутые регуляторные контуры с участием транскрипционных факторов подсемейства SREBP

Замкнутые регуляторные контуры с участием факторов подсемейства SREBP и имеющие длину 2, 3, 4. Регуляторные контуры этой категории представлены на рис. 3. Факторы подсемейства SREBP обозначены как SRBP1 и SRBP2. Один из трех отображенных на рис. 3 регуляторных контуров отражает положительную обратную связь, а два контура включают отрицательные обратные связи.

SREBP2 (белок) → *SREBF2* (ген) → SREBP2 (белок). Первый, самый короткий регуляторный контур, включавший два объекта (см. рис. 3, *а*), был обнаружен при анализе списка генов-мишеней транскрипционных факторов подсемейства SREBP (см. Приложение 3). По данным R. Sato с коллегами, промотор гена *SREBF2* человека содержит сайт связывания фактора SREBP2 (Sato et al., 1996), за счет чего осуществляется положительная авторегуляция экспрессии гена *SREBF2*.

[1] Приложения 1–9 см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx29.pdf

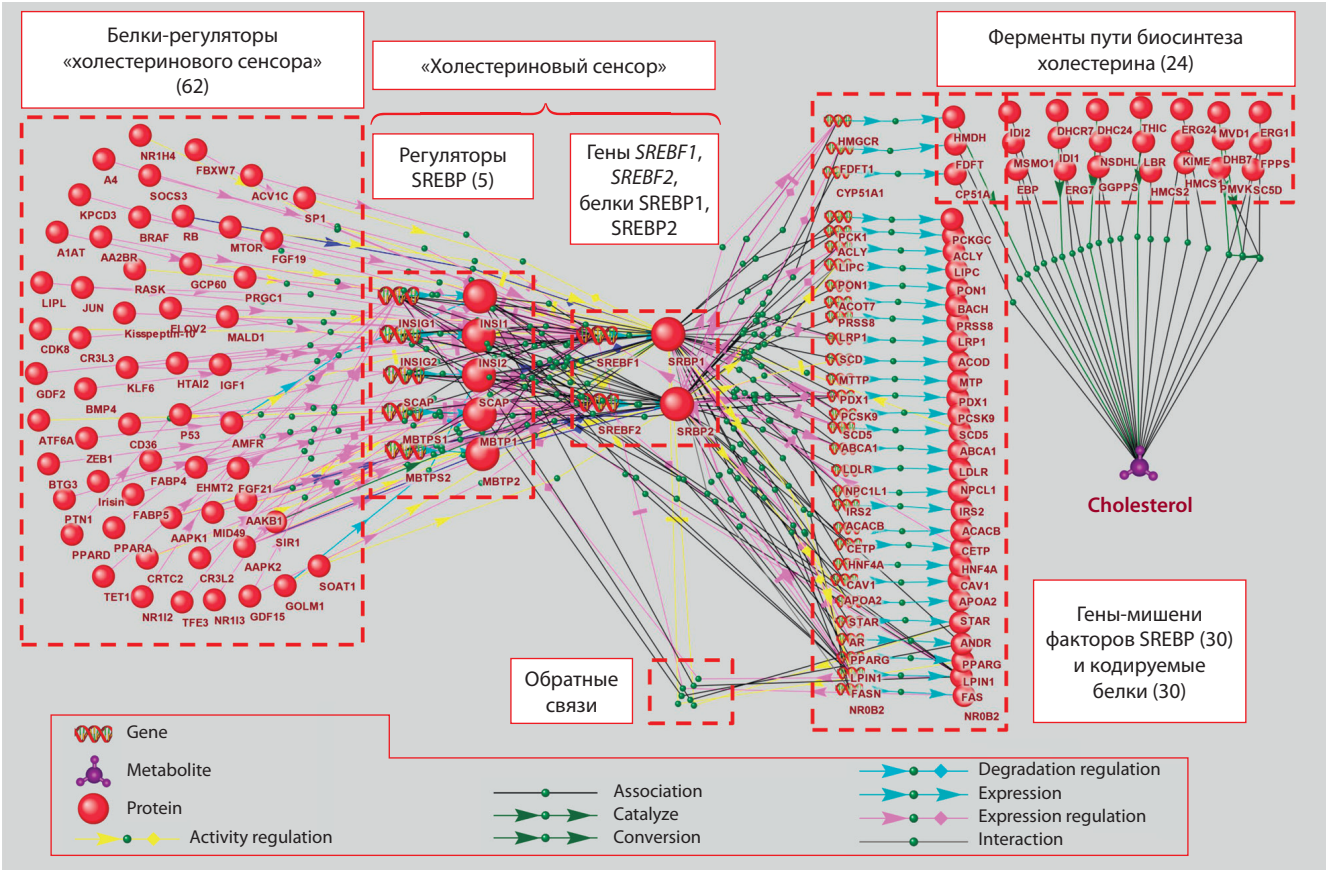


Рис. 2. Генная сеть регуляции биосинтеза холестерина, визуализированная в программе ANDVisio. Эта программа отображает белки SREBP1 и SREBP2 как объекты с обозначениями SRBP1 и SRBP2.

Списки генов каждой функциональной группы представлены в Приложениях 1–4. Приложение 3 содержит на один ген-мишень больше (т.е. 31 ген), этот 31-й ген (*SREBF2*) присутствует на рисунке в группе объектов, обозначенной как «холестериновый сенсор».

При поиске регуляторных контуров с участием факторов подсемейства SREBP мы ориентировались на шаблоны № 1–4, представленные в Приложении 7. В результате было найдено два регуляторных контура с участием фактора SREBP1 (см. рис. 3, *б*, *в*). Контуров с участием фактора SREBP2 не найдено.

SREBP1 (белок) → *LPIN1* (ген) → LPIN1 (белок) → SREBP1 (белок) (см. рис. 3, *б*). Этот регуляторный контур включает короткую петлю отрицательной обратной связи с участием гена *LPIN1* (lipin 1) и кодируемого им белка. Транскрипционный фактор SREBP1 (на рисунке обозначен как SRBP1) контролирует транскрипцию гена *LPIN1*: промотор гена *LPIN1* человека содержит мотив SRE (sterol regulatory element), и фактор SREBP1, связываясь с промотором, активирует транскрипцию гена *LPIN1* (Ishimoto et al., 2009). Белок LPIN1 подавляет активность SREBP1, предотвращая связывание SREBP1 с регуляторными районами его генов-мишеней, включая ген *LPIN1* (Mateus et al., 2021). Данный механизм осуществляется благодаря регулированию транспорта SREBP1 внутри ядра белком LPIN1. Этот белок способствует перемещению факторов подсемейства SREBP к ядерной ламине, при связывании с которой они инактивируются (Peterson et al., 2011). Активность LPIN1 контролируется киназой mTOR, участвующей в ответе клетки на ростовые факторы (Peterson et al., 2011). Таким образом, наличие регуляторного контура с участием LPIN1 свидетельствует о влиянии ростовых факторов на силу транскрипционного ответа генов при воздействии фактора SREBP1.

SREBP1 (белок) → *NR0B2/SHP1* (ген) → NR0B2/SHP1 (белок) → *SREBF1* (ген) → SREBP1 (белок) (см. рис. 3, *в*). Данная петля обратной связи реализуется с участием гена *NR0B2/SHP1* и кодируемого белка SHP1 (small heterodimer partner). Транскрипция гена *NR0B2/SHP1* человека активируется фактором SREBP1 (на рисунке обозначен как SRBP1) (Kim et al., 2004). Согласно данным базы UniProt (UniProt_ID = NR0B2_HUMAN), белок SHP1 может активироваться небольшими гидрофобными молекулами и является корепрессором транскрипции, он взаимодействует с рядом транскрипционных факторов, препятствуя их активации лигандами. Так, лиганд-зависимые транскрипционные факторы LRH-1, LXR и RXR активируют транскрипцию гена *SREBF1*. Однако белок SHP1 препятствует этой активации (Watanabe et al., 2004). Таким образом, существование регуляторного контура с участием генов и белков NR0B2/SHP1 означает, что транскрипционный ответ на понижение уровня холестерина зависит от внутриклеточного содержания других низкомолекулярных гидрофобных веществ, лигандов транскрипционных факторов LRH-1, LXR, RXR и корепрессора NR0B2/SHP1.

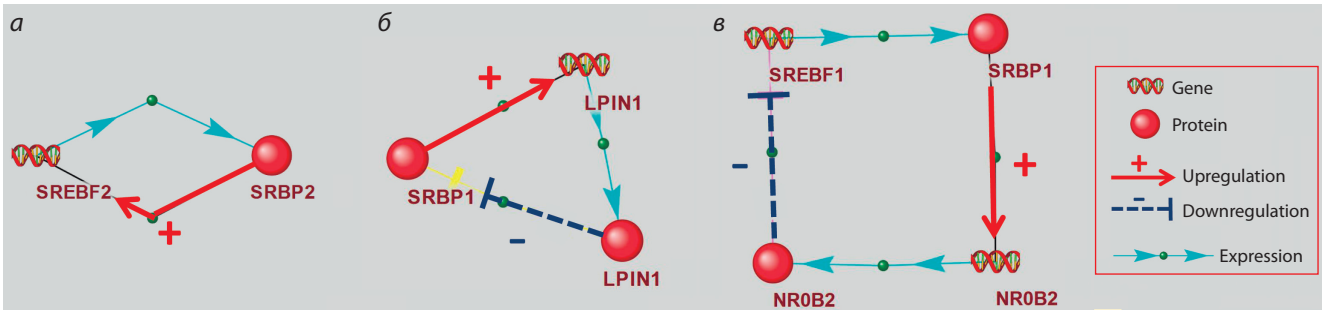


Рис. 3. Замкнутые регуляторные контуры с участием факторов подсемейства SREBP (обозначены на рисунках как SRBP1 и SRBP2).

а – положительная авторегуляция экспрессии гена *SREBF2*; *б* – контур с участием гена *LPIN1* и кодируемого им белка; *в* – контур с участием гена *NR0B2/SHP1* и кодируемого им белка.

Замкнутые регуляторные контуры с участием факторов подсемейства SREBP, а также белков «холестеринового сенсора», и имеющие длину 5. Нами выявлено три регуляторных контура с участием белков «холестеринового сенсора», которые, в свою очередь, влияют на уровень активности белков подсемейства SREBP (рис. 4). Эти контуры соответствовали шаблонам № 7 и 8, представленным в Приложении 7. Два контура включали белок SREBP1 (на рисунках обозначен как SRBP1) (см. рис. 4, *а*, *в*), и один контур включал белок SREBP2 (на рисунке обозначен как SRBP2) (см. рис. 4, *б*). Два из трех отображенных на рис. 4 регуляторных контуров отражают отрицательную обратную связь, а один контур соответствует положительной обратной связи.

SREBP1 (белок) → *PPARG* (ген) → PPARG (белок) → *INSIG1* (ген) → INSIG1 (белок) → SREBP1 (белок) (см. рис. 4, *а*).

SREBP2 (белок) → *PPARG* (ген) → PPARG (белок) → *INSIG1* (ген) → INSIG1 (белок) → SREBP2 (белок) (см. рис. 4, *б*).

Обнаружено два регуляторных контура с участием факторов подсемейства SREBP, а также генов *PPARG* и *INSIG1* и кодируемых ими белков. Белки SREBP1 и SREBP2 (на рис. 4, *а* и *б* обозначены как SRBP1 и SRBP2) способны взаимодействовать с сайтами связывания, расположенными в промоторе гена *PPARG* человека, и усиливать транскрипционную активность этого гена (Fajas et al., 1999). Белок PPARG является транскрипционным фактором, взаимодействующим с сайтом связывания (PPRE1) в промоторе гена *INSIG1* человека и усиливающим экспрессию данного гена (Kast-Woelbern et al., 2004). Увеличение экспрессии гена *INSIG1* приводит к повышению концентрации белка INSIG1, удерживающего preSREBP1 и preSREBP2 на мембране эндоплазматического ретикулума, тем самым подавляя активацию preSREBPs, которая происходит в аппарате Гольджи (Roth et al., 2008).

SREBP1 (белок) → *AR* (ген) → ANDR (белок) → *SCAP* (ген) → SCAP (белок) → SREBP1 (белок) (см. рис. 4, *в*).

В промоторной области гена *AR* человека, кодирующего белок-рецептор андрогенов (на рисунке – ANDR), содержится сайт связывания транскрипционного фактора SREBP1 (на рисунке обозначен как SRBP1), который усиливает экспрессию гена *AR* (Huang et al., 2010). Белок ANDR связывается с андроген-чувствительным районом, расположенным в восьмом интроне гена *SCAP* человека. Данное взаимодействие приводит к усилению экспрессии *SCAP* (Heemers et al., 2004), что способствует транслокации комплекса белков SCAP/SREBP1 в аппарат Гольджи, где происходит активация белка SREBP1 (Guo et al., 2019). Таким образом, этот регуляторный контур соответствует положительной обратной связи.

Обращение к данным по экспрессии генов из проекта GTEx (GTEx Consortium, 2020) показало, что найденные нами контуры (см. рис. 3 и 4) могут функционировать в широком наборе тканей. Примеры таких тканей приведены в Приложениях 8 и 9.

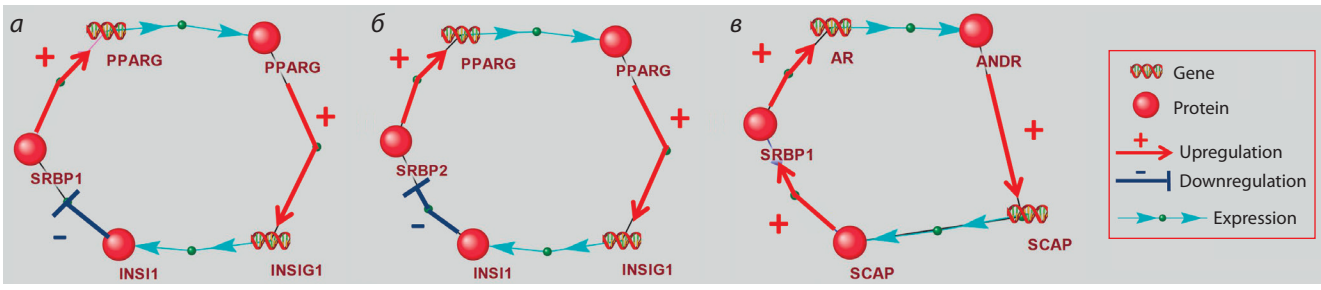
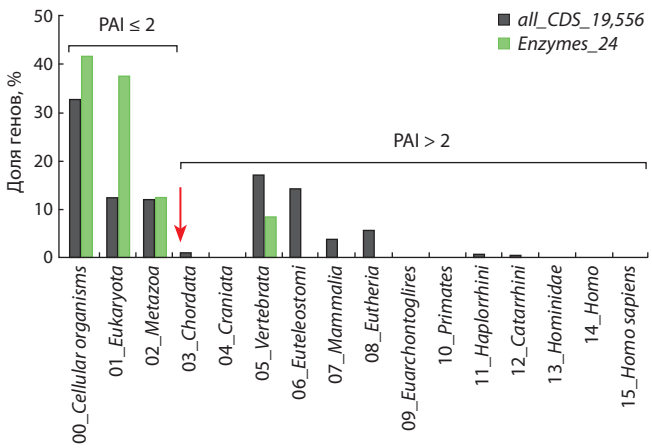


Рис. 4. Замкнутые регуляторные контуры с участием факторов подсемейства SREBP (обозначены как SRBP1 и SRBP2) и других генов и белков «холестеринового сенсора».

а – контур с участием белка SREBP1, генов *PPARG* и *INSIG1* и кодируемых ими белков; *б* – контур с участием белка SREBP2, генов *PPARG* и *INSIG1* и кодируемых ими белков; *в* – контур с участием белка SREBP1, генов *AR* и *SCAP* и кодируемых ими белков.

а



б

	all_CDS_19,556	*Enzymes_24*		χ^2 тест
		Наблюдаемое	Ожидаемое	*p*-value
PAI ≤ 2	6397	22	7.85	
PAI > 2	13 159	2	16.15	
Всего	19 556	24	24	$p < 0.001$

Рис. 5. Филостратиграфический возраст генов человека, кодирующих ферменты пути биосинтеза холестерина.

а – распределение значений индекса PAI (отложены по оси *X*) для всех белок-кодирующих генов человека (контрольная группа генов, обозначена как *all_CDS_19,556*, черные столбцы) и генов, кодирующих ферменты пути биосинтеза холестерина (эта група генов обозначена как *Enzymes_24*, зеленые столбцы); *б* – согласно критерию χ^2, наблюдаемые количества генов, кодирующих ферменты и имеющих PAI ≤ 2, достоверно отличаются от ожидаемых количеств ($p < 0.001$).

Филостратиграфический возраст генов, кодирующих ферменты биосинтеза холестерина и белки «холестеринового сенсора»

Для оценки филостратиграфического возраста генов был использован индекс PAI. Значения индекса PAI указывают на эволюционные стадии, соответствующие стадиям дивергенции определенных таксонов. Индекс PAI принимает значения от 1 до 15. Чем больше величина PAI исследуемого гена, тем моложе данный ген.

Гены, кодирующие ферменты биосинтеза холестерина. На рис. 5 представлены распределения по значениям индекса PAI для всех белок-кодирующих генов человека (черные столбцы, контрольная группа) и 24 генов, кодирующих ферменты пути биосинтеза холестерина (зеленые столбцы). Индексы PAI для генов, кодирующих ферменты пути биосинтеза холестерина, приведены в Приложении 1. Значения PAI для генов контрольной группы распределены неравномерно (см. рис. 5, *а*, черные столбцы). Примерно треть генов (~33 %) имели PAI, равный нулю (*Cellular organisms*, корень филостратиграфического дерева). А почти пятая часть (17 %) всех белок-кодирующих генов имела значение PAI, равное 5 (этап дивергенции *Vertebrata*).

При рассмотрении распределения величин PAI для выборки генов человека, кодирующих ферменты биосинтеза холестерина (см. Приложение 1), было обнаружено, что 22 гена из 24 (т. е. 92 %) имели величину PAI ≤ 2 (*Cellular organisms* (корень филостратиграфического дерева) и этапы дивергенции *Eukaryota* и *Metazoa*) (см. рис. 5, *а*, зеленые столбцы). Это количество статистически значимо отличалось от ожидаемого количества (7.85), рассчитанного исходя из распределения, полученного для выборки всех белок-кодирующих генов человека, содержавшей 19 556 генов ($p < 0.001$, см. рис. 5, *б*).

Таким образом, оказалось, что гены, кодирующие ферменты биосинтеза холестерина, характеризуются пониженными значениями индекса PAI по сравнению с набором всех белок-кодирующих генов человека, т. е. являются в среднем более «древними». Это хорошо согласуется с уже известными представлениями. Во-первых, холестерин найден в древних осадочных породах, и его производные используют в качестве биологических маркеров древней жизни на Земле (Simoneit, 2002). Во-вторых, установлено, что гены, кодирующие ферменты биосинтеза холестерина, были унаследованы многоклеточными организмами от их последнего общего эукариотического предка (Zhang T. et al., 2019). Кроме того, показано, что ферменты, принимающие участие в аминокислотном, углеводном и энергетическом обменах (включая и метаболизм липидов), обладают высокой консервативностью (Peregrín-Alvarez et al., 2009). Это связано с тем, что роль фермента заключается во взаимодействии с молекулой субстрата, т. е. трехмерные структуры фермента и субстрата должны пространственно подходить друг другу. Поэтому, как правило, эволюционным изменениям подвергается не структурная, а регуляторная часть гена, кодирующего фермент.

Гены, кодирующие белки «холестеринового сенсора». Как было указано выше и изображено на рис. 1, «холестериновый сенсор» представляет собой набор белков, совместное функционирование которых обеспечивает регуляцию транскрипции генов в зависимости от уровня холестерина в клетке. Набор генов, кодирующих белки этой группы, включает: (1) гены *SREBF1* и *SREBF2*, кодирующие транскрипционные факторы; (2) гены *SCAP*, *INSIG1* и *INSIG2*, кодирующие белки, которые изменяют свои конформационные свойства в ответ на изменение уровня холестерина, за счет чего регулируется скорость образования активных транскрипционных факторов подсемейства SREBP; (3) гены *MBTPS1* и *MBTPS2*, кодирующие ферменты-протеазы S1P и S2P, расщепляющие белки-предшественники факторов SREBP1 и SREBP2 (DeBose-Boyd, Ye, 2018; Jiang et al., 2020). Филостратиграфический возраст генов этой группы свидетельствует о достаточно раннем возникновении их предковых форм (см. цветовые обозначения объектов на рис. 1; Приложение 2).

Для четырех генов, *SCAP*, *INSIG1*, *MBTPS1/S1P* и *MBTPS2/S2P*, индекс PAI равен нулю (*Cellular organisms*, корень филостратиграфического дерева). Ген *INSIG2* имеет PAI, равный 2 (этап дивергенции *Metazoa*). Вместе с тем гены, кодирующие факторы *SREBP1* и *SREBP2*, являются более молодыми. Они имеют значение индекса PAI, равное 5 (этап дивергенции *Vertebrata*). Таким образом, хотя биосинтез холестерина был присущ даже самым древним организмам (Simoneit, 2002; Zhang T. et al., 2019), молекулярный механизм, позволяющий контролировать уровень холестерина в клетке, мог сформироваться только

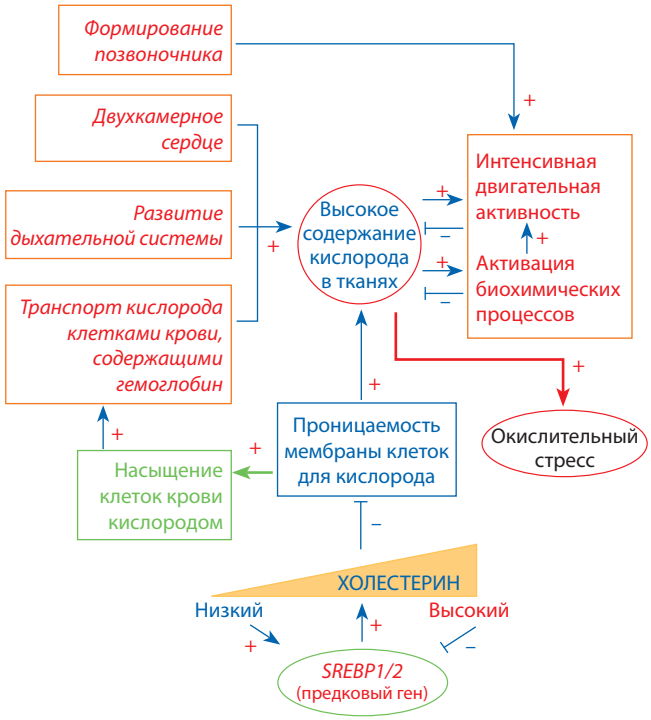


Рис. 6. Характерные особенности строения опорно-двигательного аппарата, а также кровеносной и дыхательной систем организма животных, сформировавшиеся на эволюционном этапе дивергенции *Vertebrata* (отображены курсивом), и значимая роль холестерина как фактора, снижающего проницаемость мембраны клеток для кислорода.

на значительно более поздних этапах эволюции. Это могло произойти не ранее, чем возникли первые позвоночные животные.

Этап дивергенции *Vertebrata* характеризуется существенным совершенствованием ряда физиологических систем организма (рис. 6). Формирование позвоночника на данном эволюционном этапе способствовало совершенствованию опорно-двигательного аппарата и, следовательно, большей двигательной активности. При этом возрастает потребность в снабжении мышц и других тканей кислородом. Снабжение клеток организма кислородом интенсифицируется за счет сокращений сердца, которое впервые формируется именно у позвоночных (Stephenson et al., 2017). На этом этапе эволюции совершенствуется дыхательная система и появляются специализированные клетки крови, переносящие кислород (эритроциты) (Snyder, Sheafor, 1999; Svoboda, Bartunek, 2015). Более эффективное снабжение тканей организма кислородом, с одной стороны, способствует интенсификации метаболических процессов, с другой стороны, может служить предпосылкой развития окислительного стресса.

Содержание холестерина в клеточной мембране существенным образом отражается на способности мембраны пропускать кислород внутрь клетки – его повышенная концентрация в мембране делает мембрану более плотной и менее проницаемой для кислорода (Zuniga-Hertz, Patel, 2019). Это, с одной стороны, защищает клетки от окислительного стресса, но, с другой стороны, препятствует проникновению кислорода в эритроциты и негативно отражается на биохимических процессах, протекающих с потреблением кислорода. Таким образом, возникает необходимость контроля уровня холестерина в клетке в определенных физиологических границах. С некоторого эволюционного этапа эту функцию начинают выполнять транскрипционные факторы подсемейства SREBP.

Заключение

В работе представлена генная сеть регуляции биосинтеза холестерина у человека. В этой сети систематизированы данные: (1) о полном наборе ферментов, осуществляющих биосинтез холестерина; (2) белках «холестеринового сенсора», с участием которых реализуется механизм регуляции экспрессии этих ферментов в зависимости от внутриклеточного уровня холестерина (включая транскрипционные факторы подсемейства SREBP); (3) белках, регулирующих активность белков «холестеринового сенсора»; (4) генах, кодирующих белки этих групп; (5) генах-мишенях транскрипционных факторов подсемейства SREBP. Найдены замкнутые регуляторные контуры, контролирующие активность факторов подсемейства SREBP, что свидетельствует о сложной природе молекулярно-генетических механизмов, управляющих процессом биосинтеза холестерина. В дальнейшем мы планируем расширить сеть за счет включения регуляторных влияний более высокого уровня («регуляторы регуляторов»), что обеспечит возможность выявления новых регуляторных контуров.

Исследование филостратиграфического возраста генов показало, что большинство генов человека, кодирующих ферменты биосинтеза холестерина и белки «холестеринового сенсора», датируется индексом PAI к ранним эволюционным этапам (корень филостратиграфического дерева, а также этапы дивергенции *Eukaryota* и *Metazoa*). Однако филостратиграфический возраст генов транскрипционных факторов подсемейства SREBP соответствует этапу дивергенции *Vertebrata*. Этот факт позволяет сделать вывод о том, что механизм регуляции транскрипции генов в ответ на изменение уровня холестерина мог сформироваться на более поздних эволюционных этапах, т. е. не ранее, чем этап дивергенции *Vertebrata*.

Список литературы / References

Колчанов Н.А., Игнатьева Е.В., Подколодная О.А., Лихошвай В.А., Матушкин Ю.Г. Генные сети. *Вавиловский журнал генетики и селекции*. 2013;17(4/2):833-850
[Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/2):833-850 (in Russian)]

Меркулова Т.И., Ананько Е.А., Игнатьева Е.В., Колчанов Н.А. Регуляторные коды транскрипции геномов эукариот. *Генетика*. 2013;49(1):37-54. doi 10.7868/S0016675813010074
[Merkulova T.I., Ananko E.A., Ignatieva E.V., Kolchanov N.A. Transcription regulatory codes of eukaryotic genomes. *Russ. J. Genet*. 2013;49(1):29-45. doi 10.1134/S1022795413010079]

Мустафин З.С., Лашин С.А., Матушкин Ю.Г. Филостратиграфический анализ генных сетей заболеваний человека. *Вавиловский журнал генетики и селекции*. 2021;25(1):46-56. doi 10.18699/VJ21.006
[Mustafin Z.S., Lashin S.A., Matushkin Yu.G. Phylostratigraphic analysis of gene networks of human diseases. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(1):46-56. doi 10.18699/VJ21.006]

Agrawal A., Balci H., Hanspers K., Coort S.L., Martens M., Slenter D.N., Ehrhart F., Digles D., Waagmeester A., Wassink I., Abbassi-Daloii T., Lopes E.N., Iyer A., Acosta J.M., Willighagen L.G., Nishida K., Riutta A., Basaric H., Evelo C.T., Willighagen E.L., Kutmon M., Pico A.R. WikiPathways 2024: next generation pathway database. *Nucleic Acids Res*. 2024;52(D1):D679-D689. doi 10.1093/nar/gkad960

Arito M., Horiba T., Hachimura S., Inoue J., Sato R. Growth factor-induced phosphorylation of sterol regulatory element-binding proteins inhibits sumoylation, thereby stimulating the expression of their target genes, low density lipoprotein uptake, and lipid synthesis. *J. Biol. Chem*. 2008;283(22):15224-15231. doi 10.1074/jbc.M800910200

Choy H.L., Gaylord E.A., Doering T.L. Ergosterol distribution controls surface structure formation and fungal pathogenicity. *mBio.* 2023;14(4):e0135323. doi 10.1128/mbio.01353-23

DeBose-Boyd R.A., Ye J. SREBPs in lipid metabolism, insulin signaling, and beyond. *Trends Biochem. Sci*. 2018;43(5):358-368. doi 10.1016/j.tibs.2018.01.005

Desmond E., Gribaldo S. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol*. 2009;10(1):364-381. doi 10.1093/gbe/evp036

Fajas L., Schoonjans K., Gelman L., Kim J.B., Najib J., Martin G., Fruchart J.C., Briggs M., Spiegelman B.M., Auwerx J. Regulation of peroxisome proliferator-activated receptor gamma expression by adipocyte differentiation and determination factor 1/sterol regulatory element binding protein 1: implications for adipocyte differentiation and metabolism. *Mol. Cell. Biol.* 1999;19(8):5495-5503. doi 10.1128/MCB.19.8.5495

Ferrer A., Altabella T., Arró M., Boronat A. Emerging roles for conjugated sterols in plants. *Prog. Lipid Res*. 2017;67:27-37. doi 10.1016/j.plipres.2017.06.002

GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-1330. doi 10.1126/science.aaz1776

Guo D., Wang Y., Wang J., Song L., Wang Z., Mao B., Tan N. RA-XII suppresses the development and growth of liver cancer by inhibition of lipogenesis via SCAP-dependent SREBP supression. *Molecules*. 2019;24(9):1829. doi 10.3390/molecules24091829

Han H., Cho J.W., Lee S., Yun A., Kim H., Bae D., Yang S., Kim C.Y., Lee M., Kim E., Lee S., Kang B., Jeong D., Kim Y., Jeon H.N., Jung H., Nam S., Chung M., Kim J.H., Lee I. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res*. 2018;46(D1):D380-D386. doi 10.1093/nar/gkx1013

Heemers H., Verrijdt G., Organe S., Claessens F., Heyns W., Verhoeven G., Swinnen J.V. Identification of an androgen response element in intron 8 of the sterol regulatory element-binding protein cleavage-activating protein gene allowing direct regulation by the androgen receptor. *J. Biol. Chem*. 2004;279(29):30880-30887. doi 10.1074/jbc.M401615200

Huang W.C., Zhau H.E., Chung L.W.K. Androgen receptor survival signaling is blocked by anti-β2-microglobulin monoclonal antibody via a MAPK/lipogenic pathway in human prostate cancer cells. *J. Biol. Chem*. 2010;285(11):7947-7956. doi 10.1074/jbc.M109.092759

Ishimoto K., Nakamura H., Tachibana K., Yamasaki D., Ota A., Hirano K.I., Tanaka T., Hamakubo T., Sakai J., Kodama T., Doi T. Sterol-mediated regulation of human lipin 1 gene expression in hepatoblastoma cells. *J. Biol. Chem*. 2009;284(33):22195-22205. doi 10.1074/jbc.M109.028753

Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl. 1):34. doi 10.1186/s12859-018-2567-6

Jeon T.I., Osborne T.F. SREBPs: metabolic integrators in physiology and metabolism. *Trends Endocrinol. Metab*. 2012:23(2):65-72. doi 10.1016/j.tem.2011.10.004

Jiang T., Zhang G., Lou Z. Role of the sterol regulatory element binding protein pathway in tumorigenesis. *Front. Oncol*. 2020;10:1788. doi 10.3389/fonc.2020.01788

Kast-Woelbern H.R., Dana S.L., Cesario R.M., Sun L., de Grandpre L.Y., Brooks M.E., Osburn D.L., Reifel-Miller A., Klausing K., Leibowitz M.D. Rosiglitazone induction of Insig-1 in white adipose tissue reveals a novel interplay of peroxisome proliferator-activated receptor gamma and sterol regulatory element-binding protein in the regulation of adipogenesis. *J. Biol. Chem*. 2004;279(23):23908-23915. doi 10.1074/jbc.M403145200

Kim H.J., Kim J.Y., Kim J.Y., Park S.K., Seo J.H., Kim J.B., Lee I.K., Kim K.S., Choi H.S. Differential regulation of human and mouse orphan nuclear receptor small heterodimer partner promoter by sterol regulatory element binding protein-1. *J. Biol. Chem*. 2004;279(27):28122-228131. doi 10.1074/jbc.M313302200

Kim H., Hiraishi A., Tsuchiya K., Sakamoto K. (-) Epigallocatechin gallate suppresses the differentiation of 3T3-L1 preadipocytes through transcription factors FoxO1 and SREBP1c. *Cytotechnology*. 2010;62(3):245-255. doi 10.1007/s10616-010-9285-x

Kolchanov N.A., Ignatieva E.V., Ananko E.A., Podkolodnaya O.A., Stepanenko I.L., Merkulova T.I., Pozdnyakov M.A., Podkolodny N.L., Naumochkin A.N., Romashchenko A.G. Transcription Regulatory Regions Database (TRRD): its status in 2002. *Nucleic Acids Res*. 2002;30(1):312-317. doi 10.1093/nar/30.1.312

Koolman J., Roehm K.H. (Eds). Color Atlas of Biochemistry. Stuttgart; New York: Thieme, 2005

Li N., Li X., Ding Y., Liu X., Diggle K., Kisseleva T., Brenner D.A. SREBP regulation of lipid metabolism in liver disease, and therapeutic strategies. *Biomedicines*. 2023;11(12):3280. doi 10.3390/biomedicines11123280

Luo J., Yang H., Song B.L. Mechanisms and regulation of cholesterol homeostasis. *Nat. Rev. Mol. Cell Biol.* 2020;21(4):225-245. doi 10.1038/s41580-019-0190-7

Macvanin M.T., Gluvic Z.M., Klisic A.N., Manojlovic M.S., Suri J.S., Rizzo M., Isenovic E.R. The link between miRNAs and PCKS9 in atherosclerosis. *Curr. Med. Chem*. 2024;31(42):6926-6956. doi 10.2174/0109298673262124231102042914

Mateus T., Martins F., Nunes A., Herdeiro M.T., Rebelo S. Metabolic alterations in myotonic dystrophy type 1 and their correlation with lipin. *Int. J. Environ. Res. Public. Health*. 2021;18(4):1794. doi 10.3390/ijerph18041794

Mustafin Z.S., Lashin S.A., Matushkin Y.G., Gunbin K.V., Afonnikov D.A. Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles. *BMC Bioinformatics*. 2017;18(Suppl. 1):1427. doi 10.1186/s12859-016-1427-5

Nes W.D. Biosynthesis of cholesterol and other sterols. *Chem. Rev*. 2011;111(10):6423-6451. doi 10.1021/cr200021m

Paul B., Lewinska M., Andersen J.B. Lipid alterations in chronic liver disease and liver cancer. *JHEP Rep*. 2022;4(6):100479. doi 10.1016/j.jhepr.2022.100479

Peregrín-Alvarez J.M., Sanford C., Parkinson J. The conservation and evolutionary modularity of metabolism. *Genome Biol*. 2009;10:R63. doi 10.1186/gb-2009-10-6-r63

Peterson T.R., Sengupta S.S., Harris T.E., Carmack A.E., Kang S.A., Balderas E., Guertin D.A., Madden K.L., Carpenter A.E., Finck B.N., Sabatini D.M. mTOR complex 1 regulates lipin 1 localization to control the SREBP pathway. *Cell*. 2011;146(3):408-420. doi 10.1016/j.cell.2011.06.034

Roth A., Looser R., Kaufmann M., Blättler S.M., Rencurel F., Huang W., Moore D.D., Meyer U.A. Regulatory cross-talk between drug metabolism and lipid homeostasis: constitutive androstane receptor and pregnane X receptor increase Insig-1 expression. *Mol. Pharmacol*. 2008;73(4):1282-1289. doi 10.1124/mol.107.041012

Sato R., Inoue J., Kawabe Y., Kodama T., Takano T., Maeda M. Sterol-dependent transcriptional regulation of sterol regulatory element-binding protein-2. *J. Biol. Chem*. 1996;271(43):26461-26464. doi 10.1074/jbc.271.43.26461

Schade D.S., Shey L., Eaton R.P. Cholesterol review: a metabolically important molecule. *Endocr. Pract*. 2020;26(12):1514-1523. doi 10.4158/EP-2020-0347

Shimano H., Sato R. SREBP-regulated lipid metabolism: convergent physiology – divergent pathophysiology. *Nat. Rev. Endocrinol*. 2017;13(12):710-730. doi 10.1038/nrendo.2017.91

Simoneit B.R. Molecular indicators (biomarkers) of past life. *Anat. Rec*. 2002;268(3):186-195. doi 10.1002/ar.10153

Snyder G.K., Sheafor B. Red blood cells: Centerpiece in the evolution of the vertebrate circulatory system. *Integr. Comp. Biol*. 1999; 39(2):189-198. doi 10.1093/icb/39.2.189

Stephenson A., Adams J.W., Vaccarezza M. The vertebrate heart: an evolutionary perspective. *J. Anat*. 2017;231(6):787-797. doi 10.1111/joa.12687

Sundqvist A., Bengoechea-Alonso M.T., Ye X., Lukiyanchuk V., Jin J., Harper J.W., Ericsson J. Control of lipid metabolism by phosphorylation-dependent degradation of the SREBP family of transcription factors by SCF(Fbw7). *Cell Metab*. 2005;1(6):379-391. doi 10.1016/j.cmet.2005.04.010

Svoboda O., Bartunek P. Origins of the vertebrate erythro/megakaryocytic system. *Biomed. Res. Int*. 2015;2015:632171. doi 10.1155/2015/632171

Vargas-Alarcon G., Gonzalez-Pacheco H., Perez-Mendez O., Posadas-Sanchez R., Cardoso-Saldaña G., Ramirez-Bello J., Escobedo G., Nieto-Lima B., Fragoso J.M. *SREBF1c* and *SREBF2* gene polymorphisms are associated with acute coronary syndrome and blood lipid levels in Mexican population. *PLoS One*. 2019;14(9): e0222017. doi 10.1371/journal.pone.0222017

Waller D.D., Park J., Tsantrizos Y.S. Inhibition of farnesyl pyrophosphate (FPP) and/or geranylgeranyl pyrophosphate (GGPP) biosynthesis and its implication in the treatment of cancers. *Crit. Rev. Biochem. Mol. Biol*. 2019;54(1):41-60. doi 10.1080/10409238.2019.1568964

Watanabe M., Houten S.M., Wang L., Moschetta A., Mangelsdorf D.J., Heyman R.A., Moore D.D., Auwerx J. Bile acids lower triglyceride levels via a pathway involving FXR, SHP, and SREBP-1c. *J. Clin. Invest*. 2004;113(10):1408-1418. doi 10.1172/JCI21025

Zhang F., Sun W., Chen J., Jiang L., Yang P., Huang Y., Gong A., Liu S., Ma S. SREBP-2, a new target of metformin? *Drug Des. Devel. Ther*. 2018;12:4163-4170. doi 10.2147/DDDT.S190094

Zhang T., Yuan D., Xie J., Lei Y., Li J., Fang G., Tian L., Liu J., Cui Y., Zhang M., Xiao Y., Xu Y., Zhang J., Zhu M., Zhan S., Li S. Evolution of the cholesterol biosynthesis pathway in animals. *Mol. Biol. Evol*. 2019;36(11):2548-2556. doi 10.1093/molbev/msz167

Zuniga-Hertz J.P., Patel H.H. The evolution of cholesterol-rich membrane in oxygen adaption: The respiratory system as a model. *Front. Physiol*. 2019;10:1340. doi 10.3389/fphys.2019.01340

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 06.11.2024. После доработки 15.11.2024. Принята к публикации 18.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):874-881

doi 10.18699/vjgb-24-95

Orthoweb: программный комплекс для эволюционного анализа генных сетей

Р.А. Иванов [1], А.М. Мухин [1, 2], Ф.В. Казанцев [1, 2], З.С. Мустафин [2], Д.А. Афонников [1, 2], Ю.Г. Матушкин [1], С.А. Лашин [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
ivanovromanart@bionet.nsc.ru

Аннотация. В данной статье описывается Orthoweb (https://orthoweb.sysbio.cytogen.ru/) – программный комплекс, разработанный для вычисления эволюционных индексов, включая филостратиграфические индексы и индексы дивергенции (K_a/K_s) как отдельных генов, так и генных сетей. Индекс филостратиграфического возраста (PAI) позволяет оценить эволюционную стадию появления гена (при этом косвенно оценив приблизительное время его возникновения – так называемый эволюционный возраст) на основе анализа ортологичных генов у близкородственных и дальнородственных таксонов. Кроме того, Orthoweb поддерживает расчет индексов возраста транскриптома (TAI) и дивергенции транскриптома (TDI). Эти индексы важны для понимания динамики экспрессии генов и ее последствий для развития и адаптации организмов. Orthoweb содержит также дополнительные аналитические функции, такие как возможность анализа терминов Gene Ontology (GO), что позволяет проводить функциональное обогащение и связывать эволюционное происхождение генов с биологическими процессами. Помимо этого, доступна возможность анализа обогащения по однонуклеотидным полиморфизмам (SNP), который помогает исследовать эволюционное значение генетических вариантов в конкретных геномных регионах. Одной из ключевых особенностей Orthoweb является интеграция перечисленных индексов с анализом генетических сетей. Программный пакет предлагает расширенные средства визуализации, такие как картирование генетических сетей и графическое представление распределения филостратиграфических индексов элементов сетей, что облегчает интуитивную интерпретацию сложных эволюционных связей. Для упрощения рабочих процессов в Orthoweb включена база данных с предварительно рассчитанными индексами для множества таксонов, доступная через API. Эта функция позволяет эффективно получать готовые данные по филостратиграфическим индексам и индексам дивергенции, значительно сокращая время вычислений.

Ключевые слова: генные сети; эволюция; филостратиграфия.

Для цитирования: Иванов Р.А., Мухин А.М., Казанцев Ф.В., Мустафин З.С., Афонников Д.А., Матушкин Ю.Г., Лашин С.А. Orthoweb: программный комплекс для эволюционного анализа генных сетей. *Вавиловский журнал генетики и селекции*. 2024;28(8):874-881. doi 10.18699/vjgb-24-95

Финансирование. Работа выполнена при поддержке бюджетного проекта № FWNR-2022-0020.

Orthoweb: a software package for evolutionary analysis of gene networks

R.A. Ivanov [1], A.M. Mukhin [1, 2], F.V. Kazantsev [1, 2], Z.S. Mustafin [2], D.A. Afonnikov [1, 2], Y.G. Matushkin [1], S.A. Lashin [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
ivanovromanart@bionet.nsc.ru

Abstract. This article introduces Orthoweb (https://orthoweb.sysbio.cytogen.ru/), a software package developed for the calculation of evolutionary indices, including phylostratigraphic indices and divergence indices (K_a/K_s) for individual genes as well as for gene networks. The phylostratigraphic age index (PAI) allows the evolutionary stage of a gene's emergence (and thus indirectly the approximate time of its origin, known as "evolutionary age") to be assessed based on the analysis of orthologous genes across closely and distantly related taxa. Additionally, Orthoweb supports the calculation of the transcriptome age index (TAI) and the transcriptome divergence index (TDI). These indices are important for understanding the dynamics of gene expression and its impact on the development and adaptation of organisms. Orthoweb also includes optional analytical features, such as the ability to explore Gene Ontology (GO) terms associated with genes, facilitating functional enrichment analyses that link evolutionary origins of genes to biological processes. Furthermore, it offers tools for SNP enrichment analysis, enabling the users to assess the evolutionary significance of genetic variants within specific genomic regions. A key feature of Orthoweb is its ability to

© Иванов Р.А., Мухин А.М., Казанцев Ф.В., Мустафин З.С., Афонников Д.А., Матушкин Ю.Г., Лашин С.А., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

integrate these indices with gene network analysis. The software offers advanced visualization tools, such as gene network mapping and graphical representations of phylostratigraphic index distributions of network elements, ensuring intuitive interpretation of complex evolutionary relationships. To further streamline research workflows, Orthoweb includes a database of pre-calculated indices for numerous taxa, accessible via an application programming interface (API). This feature allows the users to retrieve pre-computed phylostratigraphic and divergence data efficiently, significantly reducing computational time and effort.

Key words: gene networks; evolution; phylostratigraphy.

For citation: Ivanov R.A., Mukhin A.M., Kazantsev F.V., Mustafin Z.S., Afonnikov D.A., Matushkin Y.G., Lashin S.A. Orthoweb: a software package for evolutionary analysis of gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii* = *Vavilov Journal of Genetics and Breeding*. 2024;28(8):874-881. doi 10.18699/vjgb-24-95

Введение

Эволюционный анализ генных сетей позволяет исследовать происхождение и развитие биологических систем в контексте эволюции. Один из ключевых аспектов этого анализа – изучение индексов возраста генов, которые позволяют определить временные рамки появления и диверсификации генов в различных филогенетических линиях. Филостратиграфия, методология, основанная на оценке эволюционного возраста генов, предоставляет возможность для поиска древних и недавно возникших генов, а также для понимания их функциональной значимости в биологических процессах (Domazet-Lošo, Tautz, 2008; Tautz, Domazet-Lošo, 2011; Šestak et al., 2013; Xie et al., 2017). Целью филостратиграфического анализа является определение возраста возникновения гена-основателя на основе оценки распределения гомологичных ему генов в геномах организмов, принадлежащих к различным таксономическим группам. Для оценки времени происхождения генов в филостратиграфии используется индекс филостратиграфического возраста (Phylostratigraphic Age Index, PAI), соответствующий самой древней филострате, в которую входят гомологичные последовательности искомых генов.

Поиск генов с ограниченной в пределах таксонов гомологией особенно интересен с точки зрения эволюционной биологии, поскольку в ряде исследований было показано, что новые гены могут играть значительную роль в появлении новых эволюционных признаков и быть связаны с формированием новых морфологических признаков наземных растений (Bowles et al., 2020) и многоклеточных животных (Paps, Holland, 2018). Кроме того, продемонстрировано, что эволюционно новые гены связаны с каскадами развития органов, в частности с развитием мозговых тканей (An et al., 2023), и что гены, специфичные для таксона, перепредставлены в системах борьбы со стрессом и иммунной системе (Dornburg, Yoder, 2022). Некоторые исследователи предполагают, что гены, специфичные для таксона, связаны с экологической специализацией в различных таксонах (Baalsrud et al., 2018).

Однако классический подход филостратиграфии сталкивается с рядом ограничений, связанных с увеличением объема получаемых геномных данных и недостаточной точностью алгоритма BLASTP в определении гомологов. Эти факторы, а также высокая вычислительная сложность приводят к тому, что филостратиграфический анализ полногеномных данных с использованием BLASTP может занимать до нескольких недель (Buchfink et al., 2021). В связи с этим возникла потребность в разработке новых программных решений для филостратиграфического анализа.

Современные программные инструменты, такие как fagin (Arendsee et al., 2019), GenEra (Barrera-Redondo et al., 2023) и oggmap (Ullrich, Glytnasi, 2023), предлагают альтернативные подходы к филостратиграфическому анализу, позволяя преодолеть некоторые из ограничений классических методов. Программа fagin, написанная на языке R, использует подход к поиску гомологов, основанный на определении синтенных геномных интервалов в целевом геноме и дальнейшем поиске гомологии как в аминокислотных, так и в нуклеотидных последовательностях. Разработчики программного комплекса GenEra внесли ряд изменений в классический метод определения гомологов в филостратиграфии, заменив традиционный метод поиска BLASTP на алгоритм DIAMOND v2, что способствует лучшему определению отдаленных гомологий из-за снятия ограничений по числу лучших совпадений последовательностей при выравнивании. Также была добавлена возможность проверки ошибки поиска гомологии и возможность оценить таксономическую репрезентативность – показатель, который учитывает наличие гомологов гена хотя бы у одного репрезентативного вида для каждого промежуточного таксономического уровня между наиболее отдаленно родственным родом и искомым видом. Программа oggmap (Ullrich, Glytnasi, 2023), реализованная в виде пакета на языке программирования Python, предназначена для получения карт ортологий (ортокарт) или, другими словами, значений индекса филостратиграфического возраста заданных групп ортологов из результатов, полученных с помощью таких инструментов, как OrthoFinder (Emms, Kelly, 2019) и eggNOG (Huerta-Cepas et al., 2019). В отличие от классического метода филостратиграфии, данный подход не подразумевает шага поиска ортологов при помощи инструментов выравнивания. Вместо этого предполагается использование готовых результатов поиска ортологов в виде ортокарт и дальнейшее их применение в расчете возраста генов. Такие ортокарты содержат информацию о возрасте генов в каждой группе ортологов.

Для полноценного эволюционного анализа эти инструменты и подходы требуют знаний языков программирования. Кроме того, большая часть программных средств использует алгоритмы выравнивания BLAST, продолжительность работы которых в ряде случаев существенно замедляет анализ. Наконец, существующие на сегодняшний день программные реализации метода расчета филостратиграфических индексов не способны на полноценный и быстрый эволюционный анализ компонентов генных сетей. В этой статье мы представляем Orthoweb – программный комплекс для эволюционного

анализа генных сетей и отдельных генов, реализованный в веб-приложении по адресу https://orthoweb.sysbio.cytogen.ru.

Материалы и методы

Orthoweb разработан на языке программирования Java с использованием фреймворка Spring для реализации серверной части приложения и фреймворков Vue.JS и webix для реализации клиентской части. Для визуализации сетей используется группа библиотек cytoscape.js. В качестве СУБД для хранения данных из базы KEGG (таксоны, список ортологов, кодирующие последовательности и т. д.) и промежуточных результатов анализа применяется MongoDB, что значительно увеличивает скорость последующей работы с этими данными.

Рассчитанные индексы хранятся в базе данных на основе СУБД PostgreSQL. Доступ к этим данным реализован по технологии REST API с применением библиотеки FLASK (flask.palletsprojects.com). Благодаря этому программному интерфейсу возможно запросить данные при помощи разных инженерных сред моделирования (Matlab, Octave, Statistica) или стандартными библиотеками скриптовых языков программирования (R, Python).

Результаты

Функциональность Orthoweb

Расчет эволюционных индексов возраста единичных генов. Основная функция Orthoweb – определение индексов филостратиграфического возраста (PAI) генов.

В Orthoweb реализованы два метода определения PAI: 1) на основе анализа метрик идентичности гомологических последовательной; 2) с использованием классификации белков на ортологические группы БД KEGG (KEGG Orthology, KO). На основе информации KO из БД KEGG (Kanehisa et al., 2016) Orthoweb позволяет для каждой последовательности белка найти соответствующие ей ортологи и определить виды, в геномах которых они были найдены. Таксономические линии выявленных видов последовательно сравниваются с линией исследуемого вида, чтобы определить их эволюционную родословную и установить последнего общего предка для рассматриваемого гена, чей номер от корня таксономического дерева рассчитывается как PAI (рис. 1). Таксономические линии ортологов уже отобраны в базе данных KEGG и практически не требуют дополнительных настроек от пользователя. Рассчитанные индексы PAI сохраняются в регулярно обновляемую базу данных, о которой более подробно пойдет речь ниже. Поскольку данные по ортогруппам KEGG обновляются часто, то в Orthoweb реализована также возможность рассчитывать индексы PAI по группам ортологов напрямую из базы данных KEGG для доступа к самой актуальной информации. Однако не для всех генов такая информация доступна. В частности, для человека только примерно 2/3 генов, представленных в KEGG, ассоциированы с KO-группой.

Вторым способом расчета PAI является использование таблицы сходства последовательностей (Best Similarity Table), доступной для большинства генов, представленных в KEGG (Kanehisa et al., 2016). При помощи этого способа пользователь может отобрать гомологичные гены на основе таких параметров, как сходство аминокислотных последовательностей кодируемых генами белков и результат работы алгоритма Смита–Ватермана по локальному выравниванию последовательностей.

Расчет индексов дивергенции. Orthoweb поддерживает также опцию расчета соотношения несинонимичных и синонимичных замен (соотношение d_N/d_S) между

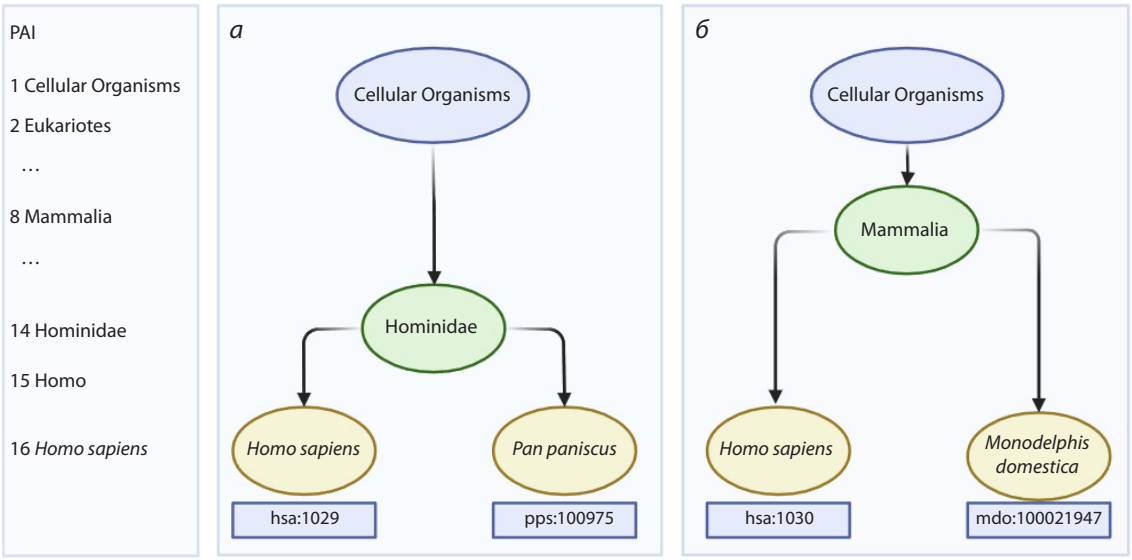


Рис. 1. Пример определения PAI для двух генов *Homo sapiens*.

а – эволюционно молодой ген hsa:1029 (CDKN2A); наиболее отдаленный от исследуемого организм, у которого был найден ортолог этого гена, – *Pan paniscus* (шимпанзе бонобо); *б* – эволюционно более древний ген hsa:1030 (CDKN2B); наиболее отдаленный от исследуемого организм, у которого был найден ортолог этого гена, – *Monodelphis domestica* (домовый опоссум). Можно заключить, что ген на примере (*а*) эволюционно моложе гена на примере (*б*). Шкала слева показывает индекс PAI, который соответствует глубине узла таксономического дерева, адаптировано из (Mustafin et al., 2021).

последовательностью изучаемого гена и каждым из его гомологов у близкородственных видов, отраженного в индексе дивергенции (divergence index, DI). Этот индекс позволяет определить тип отбора, которому подвержен ген. Индекс вычисляется на основании отношения d_N/d_S (в литературе также известен как K_a/K_s), где d_N – доля несинонимичных замен в последовательностях исследуемого гена и его гомологов, т. е. таких замен, которые приводят к смене кодируемой данным триплетом аминокислоты; d_S – доля синонимичных замен, не приводящих к замене кодируемой аминокислоты. Считается, что значение индекса меньше 1 указывает на то, что ген подвержен стабилизирующему отбору, значения близкие к 1 – нейтральной эволюции, а значения больше 1 – движущему отбору (Yang, Nielsen, 2000).

В случае сравнения с одной гомологичной последовательностью DI равен d_N/d_S. При наличии нескольких гомологов DI равен среднему значению d_N/d_S для всех сравнений. При подсчете индекса DI пользователи Orthoweb могут выбрать таксономическую глубину анализа, чтобы учитывать эволюционную изменчивость гена среди более или менее эволюционно отдаленных организмов. Для подсчета отношения d_N/d_S используется программа PAML (Yang, 2007).

Расчет насыщенности гена однонуклеотидными полиморфизмами и анализ терминов генной онтологии. Orthoweb также интегрирует информацию об ассоциированных с генами терминах генной онтологии и насыщенности исследуемых генов однонуклеотидными полиморфизмами (SNP). Для получения информации по терминам генной онтологии используется ресурс http://geneontology.org/ (Ashburner et al., 2000; Carbon et al., 2021). Загрузка происходит с помощью предоставляемого API (application programming interface – интерфейс программного доступа к ресурсам). Например, запрос для гена TBP формируется как http://api.geneontology.org/api/bioentity/gene/NCBIGene:6908/function, в нем указана база данных и идентификатор гена в ней. Orthoweb предоставляет эту информацию самостоятельно, для большинства модельных организмов используются связанные с ними базы данных (например, TAIR для *Arabidopsis thaliana*, FlyBase для *Drosophila melanogaster* и т. д.), а для остальных организмов – база UniProt. Если в Gene Ontology есть данные по исследуемому гену и в KEGG имеется необходимый идентификатор, что выполняется практически для всех хорошо изученных генов, то будут загружены идентификаторы ассоциированных с геном терминов GO и их наименования.

Для получения данных о насыщенности искомых генов однонуклеотидными полиморфизмами реализована автоматическая система запросов к базе данных NCBI SNP (Sayers et al., 2022). Запрос конструируется на основании идентификатора гена и, например, для гена TBP с идентификатором hsa:6908 имеет следующий вид: https://www.ncbi.nlm.nih.gov/snp/?term=6908. В результате запроса пользователю предоставляется количество найденных SNP. Следует отметить, что в текущей версии Orthoweb поиск SNP осуществляется только для генов человека.

Расчет эволюционных индексов группы генов. Еще одной функцией Orthoweb является ввод данных об экспрессии генов для расчета филотранскриптомных индексов. Анализ филотранскриптомных индексов – это подход, объединяющий информацию об эволюционном возрасте генов и данные об уровне их экспрессии. Этот анализ позволяет исследовать взаимосвязь между индексом PAI генов и изменениями в их активности в контексте различных физиологических состояний, адаптивных ответов или этапов развития организмов. Используя филотранскриптомный анализ, можно выявить, как эволюционные характеристики генома связаны с транскрипционной регуляцией и функциональной динамикой генов в разнообразных биологических контекстах. К филотранскриптомным индексам относятся два эволюционных индекса: Transcriptome Age Index (Domazet-Lošo, Tautz, 2010) и Transcriptome Divergence Index (Quint et al., 2012).

Transcriptome Age Index (TAI) – это индекс возраста транскриптома, который отражает взвешенный средний возраст транскриптома в выбранном биологическом процессе. Данные об экспрессии выступают в качестве дополнительного множителя и используются для нормализации результата, так что чем выше итоговое значение TAI/TDI, тем больше вклад эволюционно молодых/изменчивых генов. Формулы для расчета индексов выглядят следующим образом:

$$\mathrm{TAI} = \frac{\sum_{i=1}^{n} ps_i e_i}{\sum_{i=1}^{n} e_i},$$

где ps_i – целое число, отражающее PAI для гена с индексом i; e_i – значение уровня экспрессии, полученное с транскриптомных данных по анализу экспрессии гена i; n – общее число генов.

Transcriptome Divergence Index (TDI) – индекс дивергенции транскриптома, количественно отражает консервативность транскриптома в искомом процессе и помогает выявить биологические процессы или стадии развития организма, в которых наблюдается повышенная экспрессия более консервативных или более молодых генов:

$$\mathrm{TDI} = \frac{\sum_{i=1}^{n} \mathrm{DI}_i e_i}{\sum_{i=1}^{n} e_i}.$$

Здесь DI_i – индекс дивергенции для гена i; e_i – уровень экспрессии гена i; n – количество генов.

Примеры использования Orthoweb

Для иллюстрации работы Orthoweb мы опишем порядок действий и примеры его применения в филостратиграфическом анализе.

Анализ характеристик единичных генов. В анализе эволюционных индексов отдельных генов Orthoweb принимает несколько форматов входных файлов: список генов из формы на сайте, список генов из файла или файл со взаимодействиями между элементами генной сети в форматах .txt или .tsv. Пользователи могут выбрать желаемый формат входных данных в соответствующей форме – *Choose the type of input data* (рис. 2). Для корректного анализа в Orthoweb требуется подавать идентификаторы генов KEGG.

Следующим шагом является выбор режима анализа в форме *The type of orthology*. В этой форме можно выбрать одну из опций: расчет филостратиграфических индексов с

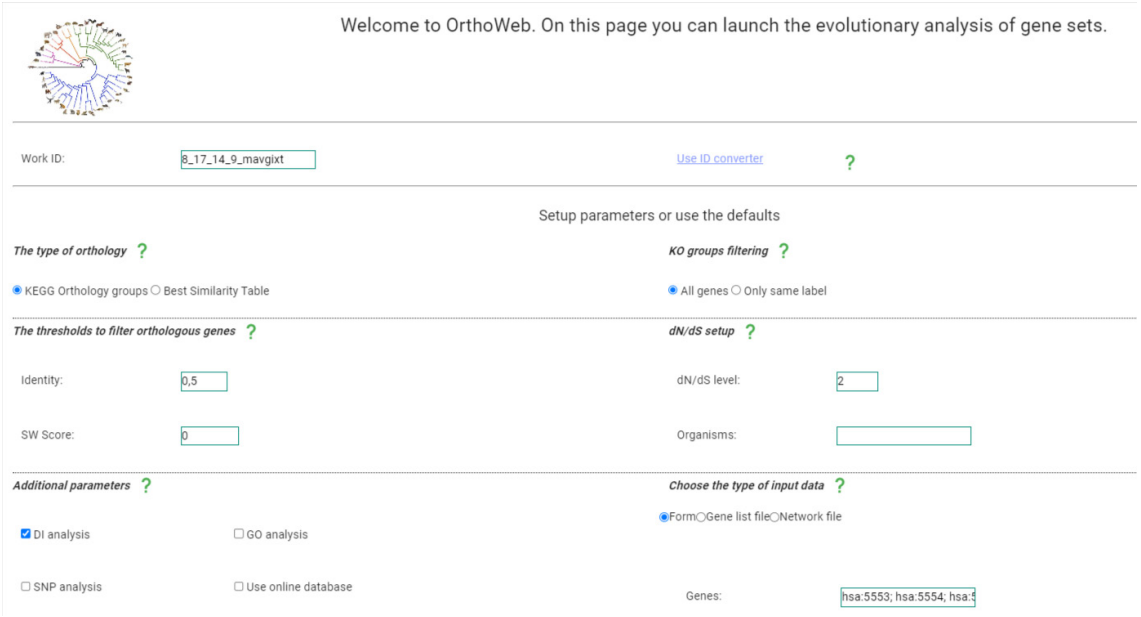


Рис. 2. Стартовая страница веб-сервиса Orthoweb.

использованием анализа семейств ортологов и KO-групп (опция *KEGG Orthology groups*) или с применением анализа гомологичных последовательностей (опция *Best Similarity Table*).

При выборе режима *KEGG Orthology groups* необходимо также принять решение относительно того, включать ли паралогичные гены в анализ в окне *KO groups filtering*.

При выборе режима расчета филостратиграфических индексов генов с применением анализа гомологов необходимо ввести в Orthoweb пороговые значения параметров идентичности аминокислотных последовательностей (по умолчанию установлено значение 0.5) и показателя алгоритма Смита–Ватермана для отбора гомологичных генов в опции *The thresholds to filter orthologous genes*.

В окне *Additional parameters* можно задать ряд дополнительных параметров для анализа: расчет индекса дивергенции DI в опции *DI analysis*, оценку обогащения однонуклеотидными полиморфизмами и идентификацию терминов Gene Ontology. Для расчета индексов дивергенции можно дополнительно настроить группы организмов, относительно которых вы хотите рассчитать значение индекса, в окне d_N/d_S *setup*. В этой опции есть два варианта настройки. Первый параметр, d_N/d_S *level*, определяет таксономический уровень, на котором проводится анализ d_N/d_S. В основном этот вид анализа используется для сравнения последовательностей близкородственных организмов. Значение параметра, равное 1, ограничивает анализ организмами в пределах одного рода. Например, при анализе генов человека значение, равное 2, означает, что d_N/d_S будет рассчитываться относительно других организмов из семейства Hominidae. Второе поле, *Organisms*, позволяет вводить коды конкретных видов из базы данных KEGG. В частности, для сравнения последовательности изучаемого гена человека не со всеми гоминидами, а только с гориллой, в поле нужно ввести код "*ggo*".

Результатом работы Orthoweb в этих режимах анализа является файл архива, в котором будет табличный текстовый файл со столбцами данных: Gene – с идентификаторами генов KEGG ID; Label – с идентификаторами Entrez ID; PAI – со значениями индексов филостратиграфического возраста; и столбцы со значениями из дополнительных режимов анализа: DI, SNP и GO label.

Анализ характеристик групп генов. Для расчета индексов возраста (TAI) и дивергенции транскриптомов (TDI) необходимо выбрать опцию формата входных данных *Network file – Use expression*. В этом режиме пользователь должен предоставить табличный текстовый файл, содержащий колонку с названиями генов и несколько колонок с нормализованными значениями экспрессии, названными по условию, в котором проводился анализ экспрессии. Подаваемый файл может быть как файлом генной сети, так и файлом только со списком генов.

На выходе программа Orthoweb выдаст табличный текстовый файл с тремя столбцами: Data – с наименованиями условий, заданных во входном файле; TAI – со значениями индекса возраста транскриптома выбранного набора генов; TDI – со значениями индекса дивергенции транскриптома в заданных условиях.

Анализ генных сетей. Помимо анализа индексов отдельных генов и списков генов, в Orthoweb реализованы филостратиграфический анализ и визуализация генных сетей. Пользователи могут анализировать как сети, импортированные из баз данных KEGG Pathway (Kanehisa et al., 2017) и WikiPathways, так и сети, загруженные из текстовых файлов. Доступ к анализу сетей из баз данных предоставляется по ссылке: https://orthoweb.sysbio.cytogen.ru/pathway.html.

Orthoweb поддерживает импорт и анализ сетей из двух основных баз данных. Первая база данных – KEGG Pathway, содержит множество генных сетей и метаболических путей, классифицированных по различным критериям,

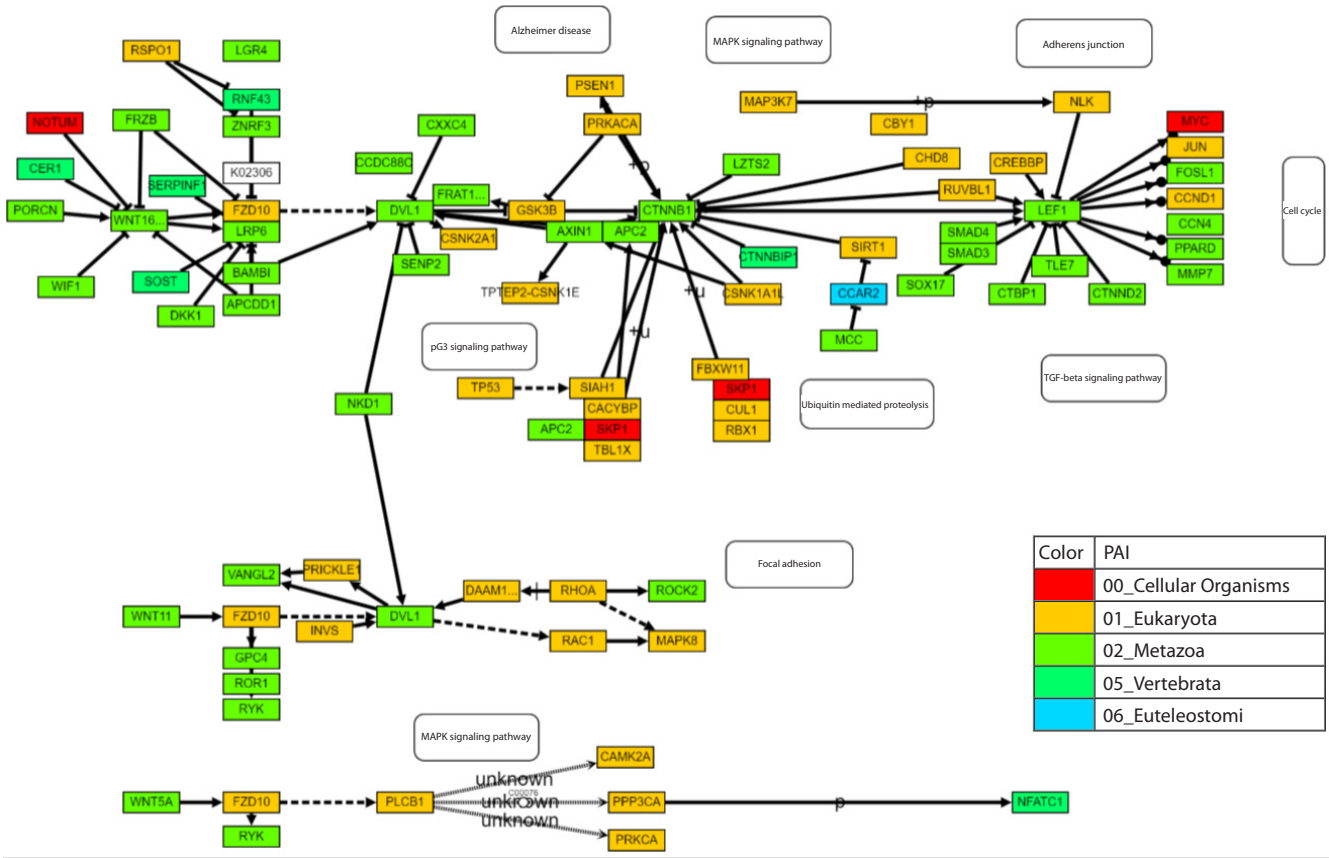


Рис. 3. Пример визуализации сети из базы данных KEGG pathway "Wnt signalling pathway" (https://www.kegg.jp/pathway/hsa 04310), проанализированной с помощью Orthoweb.

Цвет узла соответствует индексу PAI гена (белые элементы – пути и химические соединения). По умолчанию импортируется стандартная структура сети с сохранением координат элементов, но масштаб сети может быть изменен пользователем, и с каждым элементом можно взаимодействовать.

таким как метаболизм, функционирование разных систем организма и заболевания человека. Для запуска анализа пользователю необходимо указать код метаболического пути и организм, для которого будет импортирована сеть. В результате анализа сети из KEGG Pathway пользователь Orthoweb получит генную сеть, на узлах которой цветом будут отражены значения PAI, определенные на основании представленных в сети KO групп. Поскольку все элементы, представленные в сетях KEGG, описаны в самой базе KEGG, то импорт и анализ такой сети максимально удобны для Orthoweb, который большую часть информации для анализа загружает из KEGG.

В качестве примера работы этого режима Orthoweb мы проанализировали сеть Wnt/β-catenin сигнального каскада (рис. 3). WNT/β-catenin путь участвует в контроле клеточного цикла, адгезии, миграции и дифференцировке клеток. Активация пути начинается со связывания лигандов WNT с рецепторами Frizzled и LRP на поверхности клетки, что приводит к стабилизации и накоплению β-catenin в цитоплазме и его последующему перемещению в ядро, где он взаимодействует с факторами транскрипции и стимулирует экспрессию целевых генов (Davidson et al., 2009). Дисрегуляция данного пути связана с развитием множества типов рака (Zhan et al., 2017). Этот сигнальный каскад является одним из самых древних и включает в себя преимущественно гены, возникшие на этапе происхождения многоклеточных организмов и эукариот (PAI = 1, 2).

Вторая база данных, для которой реализован импорт сетей, – это WikiPathways. Сети, представленные в WikiPathways, содержат больше деталей, сущностей и вариантов взаимодействий, чем в KEGG, что усложняет их полноценный импорт и требует учета идентификаторов из множества различных баз данных.

Orthoweb предлагает пошаговый процесс импорта и анализа генных сетей, подготовленных пользователем. Пользователь может сначала импортировать сеть в формате TSV (текстовый табличный файл, где значения разделены табуляцией), а затем взаимодействовать с ней, например, перемещать элементы перед запуском анализа. Этот формат применяется в широко используемом инструменте STRING (von Mering et al., 2005), что обеспечивает совместимость и простоту интеграции данных из STRING в Orthoweb без дополнительной обработки. В случае сети, импортированной из STRING, колонка combined_score содержит достоверность найденного взаимодействия с весом от 0 до 1. По завершении анализа цвет генов изменяется в соответствии с показателем PAI (рис. 4). При включении опций дополнительных режимов анализа, описанных ранее в тексте, те также будут отражены в визуализации.

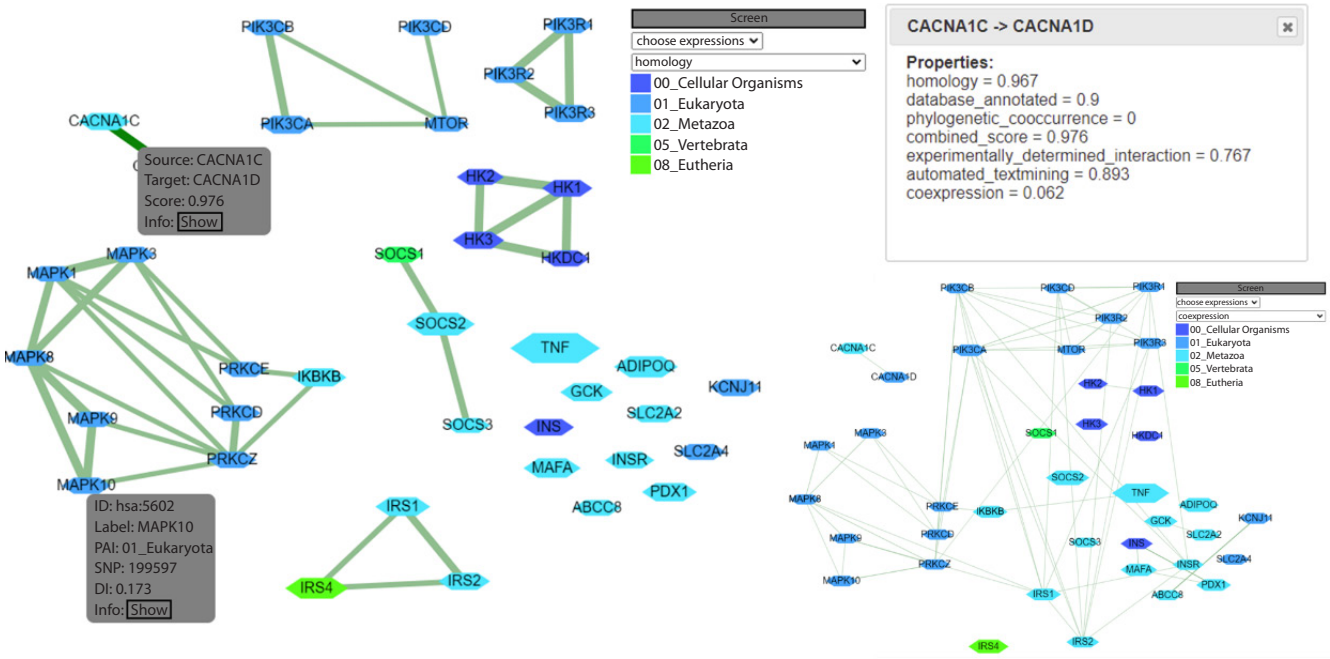


Рис. 4. Пример сети, импортированной из инструмента STRING, в котором цвет узла соответствует PAI индексу, а толщина дуги между ребрами – значению combined_score STRING. При выборке конкретной связи в сети предоставляется информация по уровням достоверностям этой связи в STRING.

База данных для хранения результатов

Для ускорения расчетов индексов и исключения многократных повторных вычислений в Orthoweb разработана база данных, содержащая таблицы: Организмы; Гены; Индексы PAI; Индексы DI; Термины генной онтологии (идентификаторы и наименования); SNP; Индексы PAI, определенные на основе KO-групп. Помимо использования этой базы при работе в интерактивном режиме, возможно ее использование через API (программный инструмент доступа) в рамках инженерных сред моделирования или популярных скриптовых языков программирования (Matlab, Octave, R, Python…). Таким образом, предоставляется доступ к имеющейся информации по всем посчитанным PAI и DI индексам для генов конкретных организмов, на основе которых можно выстроить сценарии отбора и визуализации данных. Благодаря API обращение к данным базы осуществляется через обычные URL-адреса специальной структуры. Результатом запроса является текстовый файл в структурированном формате JSON. С описанием ключей для API запросов и примером запроса к базе данных можно ознакомиться в электронном Приложении[1].

Заключение

В работе представлен программный комплекс Orthoweb, предназначенный для анализа филостратиграфических индексов и индексов дивергенции как отдельных генов, так и генных сетей. Orthoweb позволяет также интегрировать значения эволюционных индексов с данными по уровням экспрессии генов в различных условиях.

Одной из ключевых особенностей Orthoweb является его расширенная функциональность по визуализации данных. Инструменты для отображения эволюционных индексов на генных сетях значительно упрощают интерпретацию сложных эволюционных взаимосвязей, делая результаты анализа доступными для широкого круга исследователей.

[1] Приложение см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx30.pdf

Список литературы / References

An N.A., Zhang J., Mo F., Luan X., Tian L., Shen Q.S., Li X., Li C., Zhou F., Zhang B., Ji M., Qi J., Zhou W.-Z., Ding W., Chen J.-Y., Yu J., Zhang L., Shu S., Hu B., Li C.-Y. De novo genes with an lncRNA origin encode unique human brain developmental functionality. *Nat. Ecol. Evol.* 2023;7(2):264-278. doi 10.1038/s41559-022-01925-6

Arendsee Z., Li J., Singh U., Bhandary P., Seetharam A., Wurtele E.S. fagin: synteny-based phylostratigraphy and finer classification of young genes. *BMC Bioinformatics*. 2019;20(1):440. doi 10.1186/s12859-019-3023-y

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G.M., Sherlock G. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 2000;25(1):25-29. doi 10.1038/75556

Baalsrud H.T., Tørresen O.K., Solbakken M.H., Salzburger W., Hanel R., Jakobsen K.S., Jentoft S. *De novo* gene evolution of antifreeze glycoproteins in codfishes revealed by whole genome sequence data. *Mol. Biol. Evol.* 2018;35(3):593-606. doi 10.1093/molbev/msx311

Barrera-Redondo J., Lotharukpong J.S., Drost H.-G., Coelho S.M. Uncovering gene-family founder events during major evolutionary transitions in animals, plants and fungi using GenEra. *Genome Biol.* 2023;24(1):54. doi 10.1186/s13059-023-02895-z

Bowles A.M.C., Bechtold U., Paps J. The origin of land plants is rooted in two bursts of genomic novelty. *Curr. Biol.* 2020;30(3):530-536.e2. doi 10.1016/j.cub.2019.11.090

Buchfink B., Reuter K., Drost H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods*. 2021;18(4):366-368. doi 10.1038/s41592-021-01101-x

Carbon S., Douglass E., Good B.M., Unni D.R., Harris N.L., Mungall C.J., Basu S., Chisholm R.L., Dodson R.J., Hartline E., … Stein L., Howe D.G., Toro S., Westerfield M., Jaiswal P., Cooper L., Elser J. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 2021;49(D1):D325-D334. doi 10.1093/nar/gkaa1113

Davidson G., Shen J., Huang Y.-L., Su Y., Karaulanov E., Bartscherer K., Hassler C., Stannek P., Boutros M., Niehrs C. Cell cycle control of Wnt receptor activation. *Dev. Cell*. 2009;17(6):788-799. doi 10.1016/j.devcel.2009.11.006

Domazet-Lošo T., Tautz D. An ancient evolutionary origin of genes associated with human genetic diseases. *Mol. Biol. Evol.* 2008;25(12):2699-2707. doi 10.1093/molbev/msn214

Domazet-Lošo T., Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*. 2010;468(7325):815-819. doi 10.1038/nature09632

Dornburg A., Yoder J.A. On the relationship between extant innate immune receptors and the evolutionary origins of jawed vertebrate adaptive immunity. *Immunogenetics*. 2022;74(1):111-128. doi 10.1007/s00251-021-01232-7

Emms D.M., Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. doi 10.1186/s13059-019-1832-y

Huerta-Cepas J., Szklarczyk D., Heller D., Hernández-Plaza A., Forslund S.K., Cook H., Mende D.R., Letunic I., Rattei T., Jensen L.J., von Mering C., Bork P. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2019;47(D1):D309-D314. doi 10.1093/nar/gky1085

Kanehisa M., Sato Y., Kawashima M., Furumichi M., Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44(D1):D457-D462. doi 10.1093/nar/gkv1070

Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353-D361. doi 10.1093/nar/gkw1092

Mustafin Z.S., Lashin S.A., Matushkin Y.G. Phylostratigraphic analysis of gene networks of human diseases. *Vavilov J. Genet. Breed.* 2021;25(1):46-56. doi 10.18699/VJ21.006

Paps J., Holland P.W.H. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. *Nat. Commun.* 2018;9(1):1730. doi 10.1038/s41467-018-04136-5

Quint M., Drost H.G., Gabel A., Ullrich K.K., Bönn M., Grosse I. A transcriptomic hourglass in plant embryogenesis. *Nature*. 2012;490(7418):98-101. doi 10.1038/nature11394

Sayers E.W., Bolton E.E., Brister J.R., Canese K., Chan J., Comeau D.C., Connor R., Funk K., Kelly C., Kim S., Madej T., Marchler-Bauer A., Lanczycki C., Lathrop S., Lu Z., Thibaud-Nissen F., Murphy T., Phan L., Skripchenko Y., Tse T., Wang J., Williams R., Trawick B.W., Pruitt K.D., Sherry S.T. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2022;50(D1):D20-D26. doi 10.1093/nar/gkab1112

Šestak M.S., Božičević V., Bakarić R., Dunjko V., Domazet-Lošo T. Phylostratigraphic profiles reveal a deep evolutionary history of the vertebrate head sensory systems. *Front. Zool.* 2013;10(1):18. doi 10.1186/1742-9994-10-18

Tautz D., Domazet-Lošo T. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* 2011;12(10):692-702. doi 10.1038/nrg3053

Ullrich K.K., Glytnasi N.E. oggmap: a Python package to extract gene ages per orthogroup and link them with single-cell RNA data. *Bioinformatics*. 2023;39(11):btad657. doi 10.1093/bioinformatics/btad657

von Mering C., Jensen L.J., Snel B., Hooper S.D., Krupp M., Foglierini M., Jouffre N., Huynen M.A., Bork P. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 2005;33(D1):D433-D437. doi 10.1093/nar/gki005

Xie L., Draizen E.J., Bourne P.E. Harnessing big data for systems pharmacology. *Annu. Rev. Pharmacol. Toxicol.* 2017;57(1):245-262. doi 10.1146/annurev-pharmtox-010716-104659

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 2007;24(8):1586-1591. doi 10.1093/molbev/msm088

Yang Z., Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 2000;17(1):32-43. doi 10.1093/oxfordjournals.molbev.a026236

Zhan T., Rindtorff N., Boutros M. Wnt signaling in cancer. *Oncogene*. 2017;36(11):1461-1473. doi 10.1038/onc.2016.304

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 08.11.2024. После доработки 21.11.2024. Принята к публикации 22.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):882-896

doi 10.18699/vjgb-24-96

Изучение особенностей метаболизма тканей глиобластомы и перитуморального пространства при использовании таргетированного метаболомного скрининга методом ВЭЖХ-МС/МС и генных сетей

Н.В. Басов[1, 2 §], А.В. Адамовская[2, 3 §], А.Д. Рогачев[1, 2], Е.В. Гайслер[2], П.С. Деменков[2, 3], Т.В. Иванисенко[2, 3], А.С. Вензель[2, 3], С.В. Мишинов[4], В.В. Ступак[4], С.В. Чересиз[2], О.С. Олешко[2], Е.А. Бутикова[2, 5], А.Е. Осечкова[1, 6], Ю.С. Сотникова[1, 2, 6], Ю.В. Патрушев[2, 6], А.С. Поздняков[7], И.Н. Лаврик[3], В.А. Иванисенко[2, 3, 8], А.Г. Покровский[2]

[1] Новосибирский институт органической химии им. Н.Н. Ворожцова Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[4] Новосибирский научно-исследовательский институт травматологии и ортопедии им. Я.Л. Цивьяна Министерства здравоохранения Российской Федерации, Новосибирск, Россия
[5] Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[6] Федеральный исследовательский центр Институт катализа им. Г.К. Борескова Сибирского отделения Российской академии наук, Новосибирск, Россия
[7] Федеральный исследовательский центр «Иркутский институт химии им. А.Е. Фаворского Сибирского отделения Российской академии наук», Иркутск, Россия
[8] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

basov@nioch.nsc.ru

Аннотация. В ходе исследования проведен комплексный анализ метаболомных профилей глиобластомы и прилегающей ткани головного мозга, включавший 17 образцов глиобластомы и 15 образцов перитуморальной ткани. С использованием метода ВЭЖХ-МС/МС было проанализировано более 400 метаболитов, что позволило выявить значимые различия в их содержании между опухолевой и перитуморальной тканями. Статистический анализ, включавший тесты Манна–Уитни и Куккони, показал, что существенное количество метаболитов, в частности церамиды, значимо различается в тканях глиобластомы и перитуморального пространства. Анализ метаболических путей из базы данных KEGG, выполненный с помощью MetaboAnalyst 6.0, выявил статистически значимую перепредставленность метаболизма сфинголипидов, что указывает на важную роль этих липидных молекул в патогенезе глиобластомы. С использованием методов компьютерной системной биологии и искусственного интеллекта, реализованных в когнитивной системе ANDSystem, реконструированы молекулярно-генетические регуляторные пути, описывающие потенциальные механизмы нарушения функции ферментов метаболизма сфинголипидов. Реконструированные пути были объединены в регуляторную генную сеть. Данная сеть включала 15 генов, 329 белков и 389 взаимодействий, при этом 119 из 294 белков, регулирующих ключевые ферменты сфинголипидного метаболизма, оказались ассоциированы с глиобластомой. Анализ перепредставленности биологических процессов Gene Ontology показал статистическую значимость 184 процессов, в том числе апоптоза, сигнального пути NF-κB, пролиферации, миграции, ангиогенеза и пироптоза, многие из которых играют важную роль в онкогенезе. Результаты исследования подчеркивают ключевую роль метаболизма сфинголипидов в развитии глиобластомы и открывают новые перспективы для разработки терапевтических подходов, направленных на его модуляцию.

Ключевые слова: глиобластома; перитуморальная ткань; маркеры; метаболомика; ВЭЖХ-МС/МС; сфинголипиды; метаболические пути; генные сети; когнитивная система ANDSystem.

Для цитирования: Басов Н.В., Адамовская А.В., Рогачев А.Д., Гайслер Е.В., Деменков П.С., Иванисенко Т.В., Вензель А.С., Мишинов С.В., Ступак В.В., Чересиз С.В., Олешко О.С., Бутикова Е.А., Осечкова А.Е., Сотникова Ю.С., Патрушев Ю.В., Поздняков А.С., Лаврик И.Н., Иванисенко В.А., Покровский А.Г. Изучение особенностей метаболизма тканей глиобластомы и перитуморального пространства при использовании таргетированного метаболомного скрининга методом ВЭЖХ-МС/МС и генных сетей. *Вавиловский журнал генетики и селекции*. 2024;28(8):882-896. doi 10.18699/vjgb-24-96

Финансирование. Изготовление монолитных колонок для HPLC поддержано средствами проекта № FWUR-2024-0032. Отбор, пробоподготовка и анализ образцов методом ВЭЖХ-МС/МС поддержаны средствами государственного задания № FSUS-2020-0035. Биоинформатический анализ поддержан бюджетным проектом FWNR-2022-0020.

© Басов Н.В., Адамовская А.В., Рогачев А.Д., Гайслер Е.В., Деменков П.С., Иванисенко Т.В., Вензель А.С., Мишинов С.В., Ступак В.В., Чересиз С.В., Олешко О.С., Бутикова Е.А., Осечкова А.Е., Сотникова Ю.С., Патрушев Ю.В., Поздняков А.С., Лаврик И.Н., Иванисенко В.А., Покровский А.Г., 2024

§ Авторы внесли равный вклад в работу.

Контент доступен под лицензией Creative Commons Attribution 4.0

Investigation of metabolic features of glioblastoma tissue and the peritumoral environment using targeted metabolomics screening by LC-MS/MS and gene network analysis

N.V. Basov[1, 2 §], A.V. Adamovskaya [2, 3 §], A.D. Rogachev [1, 2], E.V. Gaisler[2], P.S. Demenkov [2, 3], T.V. Ivanisenko [2, 3], A.S. Venzel[2, 3], S.V. Mishinov [4], V.V. Stupak[4], S.V. Cheresiz[2], O.S. Oleshko[2], E.A. Butikova[2, 5], A.E. Osechkova [1, 6], Yu.S. Sotnikova[1, 2, 6], Y.V. Patrushev[2, 6], A.S. Pozdnyakov[7], I.N. Lavrik[3], V.A. Ivanisenko [2, 3, 8], A.G. Pokrovsky [2]

[1] N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[4] Novosibirsk Research Institute of Traumatology and Orthopedics named after Ya.L. Tsivyan of the Ministry of Health of the Russian Federation, Novosibirsk, Russia
[5] Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[6] Boreskov Institute of Catalysis of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[7] A.E. Favorsky Irkutsk Institute of Chemistry of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia
[8] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

basov@nioch.nsc.ru

Abstract. The metabolomic profiles of glioblastoma and surrounding brain tissue, comprising 17 glioblastoma samples and 15 peritumoral tissue samples, were thoroughly analyzed in this investigation. The LC-MS/MS method was used to analyze over 400 metabolites, revealing significant variations in metabolite content between tumor and peritumoral tissues. Statistical analyses, including the Mann–Whitney and Cucconi tests, identified several metabolites, particularly ceramides, that showed significant differences between glioblastoma and peritumoral tissues. Pathway analysis using the KEGG database, conducted with MetaboAnalyst 6.0, revealed a statistically significant overrepresentation of sphingolipid metabolism, suggesting a critical role of these lipid molecules in glioblastoma pathogenesis. Using computational systems biology and artificial intelligence methods implemented in a cognitive platform, ANDSystem, molecular genetic regulatory pathways were reconstructed to describe potential mechanisms underlying the dysfunction of sphingolipid metabolism enzymes. These reconstructed pathways were integrated into a regulatory gene network comprising 15 genes, 329 proteins, and 389 interactions. Notably, 119 out of the 294 proteins regulating the key enzymes of sphingolipid metabolism were associated with glioblastoma. Analysis of the overrepresentation of Gene Ontology biological processes revealed the statistical significance of 184 processes, including apoptosis, the NF-kB signaling pathway, proliferation, migration, angiogenesis, and pyroptosis, many of which play an important role in oncogenesis. The findings of this study emphasize the pivotal role of sphingolipid metabolism in glioblastoma development and open new prospects for therapeutic approaches modulating this metabolism.

Key words: glioblastoma; peritumoral tissue; markers; metabolomics; LC-MS/MS; sphingolipids; metabolic pathways; gene networks; cognitive system ANDSystem.

For citation: Basov N.V., Adamovskaya A.V., Rogachev A.D., Gaisler E.V., Demenkov P.S., Ivanisenko T.V., Venzel A.S., Mishinov S.V., Stupak V.V., Cheresiz S.V., Oleshko O.S., Butikova E.A., Osechkova A.E., Sotnikova Yu.S., Patrushev Y.V., Pozdnyakov A.S., Lavrik I.N., Ivanisenko V.A., Pokrovsky A.G. Investigation of metabolic features of glioblastoma tissue and the peritumoral environment using targeted metabolomics screening by LC-MS/MS and gene network analysis. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):882-896. doi 10.18699/vjgb-24-96

Введение

Глиобластома (ГБМ) – наиболее распространенная первичная опухоль головного мозга у взрослых, агрессивность которой во многом определяется ее инвазией – активной инфильтрацией отдельных злокачественных клеток или их групп в паренхиму мозга, окружающую опухоль (Vollmann-Zwerenz et al., 2020). Всемирная организация здравоохранения (ВОЗ) классифицирует глиомы в зависимости от их типа клеток и агрессивности: класс I включает доброкачественные опухоли, а класс IV включает наиболее агрессивные типы опухолей – глиобластомы (Louis et al., 2021). Плохая выживаемость пациентов с ГБМ даже после наиболее радикальной операции по удалению первичной опухоли, сопровождаемой мультимодальной химиорадиотерапией (Omuro, DeAngelis, 2013), связана с рецидивами опухоли, которые появляются непосредственно в послеоперационной полости или в ее 2-сантиметровой маргинальной зоне, или в виде отдаленных и множественных рецидивных опухолевых очагов. По-видимому, эти рецидивные опухоли формируются за счет того, что клетки ГБМ из перитуморальной зоны вновь мигрируют в полость первичной опухоли или в отдаленные участки мозга.

Несмотря на устоявшиеся подходы к верификации заболевания, проблема установления прогноза опухолевого роста и чувствительности опухоли к лечению в настоящие дни находится на стадии решения. Так, в 2016 г. ВОЗ представила новую систему классификации опухолей головного мозга, включающую генетические маркеры IDH1/IDH2, О-6-метилгуанин ДНК-метилтрансферазу (MGMT) и рецепторы эпидермального фактора роста (EGFR) (Louis et al., 2021). Использование результатов исследований маркерных молекул позволяет клиницистам дифференцировать опухоли не только по типу клеток и агрессив-

ности, как это было возможно при ранее существовавших методах, но и по генетическому фенотипу неопластических клеток, что обеспечивает лучшую корреляцию с прогнозом опухоли (Jaroch et al., 2021). Молекулярные биомаркеры стали неотъемлемой частью оценки глиальных опухолей, а их содержание определяет клинические решения при некоторых подтипах глиом, включая лечение. Обсуждается потенциал классификации глиом, в основе которой будут лежать молекулярные маркеры, что позволит более качественно внедрять персонализированные подходы в терапии (Siegal, 2015). Одним из перспективных направлений современной науки, ориентированной на поиск биомаркеров заболеваний, является применение омиксных технологий, среди прочего – метаболомного скрининга.

Как и большинство злокачественных опухолей, глиобластома обладает уникальным биоэнергетическим состоянием аэробного гликолиза, известным как эффект Варбурга (Siegal, 2015), при котором аэробный гликолиз служит основным источником АТФ для раковых клеток (Warburg, 1956). Представления о метаболизме раковых клеток активно развиваются, однако преимущества, которые дает раковым клеткам метаболическая трансформация, остаются неясными (Koppenol et al., 2011), как и неясен механизм воздействия гипоксии на метаболическое перепрограммирование опухолевых клеток. Недавние открытия связи между онкогенами и метаболическими процессами привели к возрождению интереса к работам Варбурга (Poteet et al., 2013). Появляется все больше данных, указывающих на то, что адаптация аэробного гликолиза в раковых клетках может способствовать накоплению биомассы и, следовательно, пролиферации злокачественных клеток (Heiden et al., 2009). Изучение метаболизма опухолевых клеток актуально для разработки моделей, отражающих состав микроокружения опухоли (Liberti, Locasale, 2016), и поиска новых эффективных терапевтических подходов. Метаболомный подход к изучению глиобластомы приобретает большой интерес как диагностический инструмент, а также инструмент для изучения метаболизма ГБМ, знание которого может помочь в разработке новых терапевтических подходов (Pandey et al., 2017; Zhou, Wahl, 2019). В некоторых аспектах метаболомный анализ превосходит анализ экспрессии генов, поскольку функция генов может быть затронута на эпигенетическом и посттрансляционном уровнях, тогда как метаболиты служат непосредственными индикаторами ферментативной активности и биохимических процессов в клетке (Pandey et al., 2017).

Наиболее достоверным подходом к изучению особенностей метаболизма глиобластомы представляется анализ метаболических различий между разными областями глиобластомы (Wolf et al., 2010; Chinnaiyan et al., 2012), в частности между центральной областью опухоли и перитуморальной зоной, образцы которых можно получить интраоперационно в ходе резекции опухоли (Youngblood et al., 2021). Однако таких исследований в литературе представлено мало.

Для выявления молекулярно-генетических механизмов, объясняющих наблюдаемые метаболомные данные, применяют различные методы, основанные на анализе метаболических путей и генных сетей. Генные сети могут дать ценную информацию о генетической регуляции идентифицированных метаболических путей, составляющих основу для интеграции метаболомных и геномных данных (Колчанов и др., 2013). Нами ранее была разработана программно-информационная система ANDSystem для автоматизированного извлечения биологических и медицинских знаний из научных публикаций на основе методов искусственного интеллекта (Demenkov et al., 2011; Ivanisenko V.A. et al., 2015, 2019; Ivanisenko T.V. et al., 2020, 2022). Эта программа позволяет пользователям проводить реконструкцию, расширение, графическую визуализацию генных сетей, осуществлять фильтрацию данных, а также с помощью шаблонов проводить поиск регуляторных путей в глобальной генной сети. ANDSystem применялась для анализа молекулярно-генетических механизмов широкого круга заболеваний, включая коморбидные состояния, ответ организма на стресс, поиск фармакологических мишеней, и других задач (Bragina et al., 2014, 2016, 2023; Popik et al., 2016; Saik et al., 2016, 2018a, b, 2019; Zolotareva et al., 2019; Антропова и др., 2022; Деменков и др., 2023).

С использованием ANDSystem проводился анализ метаболомных и протеомных данных (Pastushkova et al., 2013, 2019; Binder et al., 2014; Larina et al., 2015; Rogachev et al., 2021; Ivanisenko V.A. et al., 2022; Иванисенко и др., 2023). В частности, на основе метаболомного анализа плазмы крови пациентов с COVID-19 была установлена роль неструктурных вирусных белков в нарушениях метаболизма при данной патологии, что повлияло на изменение метаболомного профиля (Ivanisenko V.A. et al., 2022). Анализ метаболомного профиля пациентов с постоперационным делирием, выполненный с помощью ANDSystem, позволил идентифицировать потенциальные маркеры, представленные рядом сфинголипидов, и установить молекулярно-генетические механизмы нарушения их метаболизма (Иванисенко и др., 2023).

В настоящей работе был проведен таргетированный скрининг широкого профиля метаболитов в глиобластоме и перитуморальной ткани опухоли. В ходе статистического анализа данных скрининга были выявлены метаболиты, участвующие в метаболизме сфинголипидов, содержание которых в опухоли значительно отличалось от содержания в перитуморальной ткани. На основе реконструкции генных сетей определены гены, оказывающие наибольшее регуляторное влияние на функции и экспрессию ключевых ферментов метаболизма сфинголипидов, среди которых оказались как известные опухолевые маркеры (p53, TNF-α, VEGF и др.), так и перспективные кандидаты-маркеры (KLF4, E2F4 и др.). Нарушение функции этих генов при глиобластоме может объяснять наблюдаемые изменения метаболомного профиля.

Материалы и методы

Реактивы и материалы. Метанол и ацетонитрил, использованные для пробоподготовки и анализа, относились к градиентной ВЭЖХ и были приобретены в компании «Химмед» (Москва, Россия). Очищенную воду готовили на установке Sartorius arium 611DI. Элюент А готовили в соответствии с протоколом, описанным в (Li et al., 2017).

Пациенты – участники исследования. Для получения ткани опухолей в исследование были включены пациенты, подвергшиеся оперативному лечению в нейрохирургическом отделении Новосибирского научно-исследовательского института травматологии и ортопедии (НИИТО) им. Я.Л. Цивьяна (Новосибирск, Россия) по поводу впервые выявленной ГБМ в 2019–2022 гг., в том числе пациенты с глиомами IV класса, госпитализированные для хирургической резекции опухоли. Диагноз был подтвержден МРТ и гистопатологическим исследованием эксцизионного образца биопсии. Окончательный диагноз ставился на основании гистологического исследования и заключения двух патоморфологов в соответствии с классификацией ВОЗ. Образцы опухолей были собраны интраоперационно и анонимизированы для исследователей. Идентификатор клинических исследований NCT03865355. Всего в исследование было включено 17 пациентов. Ниже представлена их половозрастная характеристика:

Пол (М/Ж)	Возраст, лет				
	мин.	макс.	средний	медиана	станд. отклон.
8/9	28.2	69.6	54.9	63.22	15.9

Сбор образцов опухолевой ткани от пациентов. Образцы опухолевой ткани получали во время циторедуктивного оперативного вмешательства. После отбора образцы немедленно помещали в среду для культивирования клеток RPMI1 640 без добавок и хранили при +4 °C до обработки. Части опухоли из разных областей (центр опухоли и перитуморальная ткань) разделяли с помощью хирургического инструмента на фрагменты размером от 2×2×2 до 5×5×5 мм, заворачивали в пакетики из стерильной фольги, замораживали в жидком азоте и хранили в дальнейшем в низкотемпературном холодильнике (–80 °C). В общей сложности в исследовании были использованы образцы глиобластомы от 17 пациентов и образцы ткани перитуморального пространства от 15 пациентов.

Соответствие этическим стандартам. Исследование было рассмотрено Комитетом по этике Института медицины и психологии Зельмана Новосибирского государственного университета (протокол заседания от 04.01.2018), и все экспериментальные протоколы были утверждены. Все процедуры с участием людей признаны соответствующими этическим стандартам институционального исследовательского комитета и Хельсинкской декларации 1964 г. и последующим поправкам к ней или аналогичным этическим стандартам. Форма информированного согласия была заполнена и подписана каждым пациентом. Исследование одобрено Локальным этическим комитетом Новосибирского НИИТО им. Я.Л. Цивьяна (протокол заседания от 11.09.2017 № 050/17), информированное добровольное согласие получено у всех участников.

Пробоподготовка образцов глиобластомы и ткани перитуморального пространства. Экстракцию метаболитов из образцов глиобластомы проводили одновременно в соответствии с модифицированным протоколом, описанным в (Yuan et al., 2012; Li et al., 2017). В пробирке объемом 1.5 мл к образцу ткани массой от 9 до 33 мг добавляли 250 мкл 80 % (об.) охлажденного метанола на 10 мг ткани. Затем образцы гомогенизировали в течение 2 мин на тканевом гомогенизаторе Bertin Minilys (Rockville, Мэриленд, США) (для улучшения дезинтеграции образцов добавляли гранитную крошку), после чего инкубировали 24 ч при –70 °C. Затем образцы перемешивали на вортексе и центрифугировали при +4 °C и 16000 g в течение 15 мин. Супернатант количественно переносили в новую полипропиленовую пробирку. К осадку добавляли тот же объем охлажденного 80 % (об/об) метанола, встряхивали на вортексе 1 мин, выдерживали при –70 °C в течение 30 мин и центрифугировали в аналогичных условиях. После центрифугирования супернатанты из обоих экстрактов объединяли, отбирали аликвоту объемом 500 мкл и упаривали досуха в вакуумной центрифуге SpeedVac concentrator (Thermo Fisher Scientific/Savant, Уолтем, США). Высушенные образцы восстанавливали в 20 мкл воды MilliQ и анализировали.

Высокоэффективная жидкостная хроматография с масс-спектрометрической детекцией. Образцы анализировали методом высокоэффективной жидкостной хроматографии с тандемной масс-спектрометрической детекцией (ВЭЖХ-МС/МС) по методике, описанной в работе (Basov et al., 2024). Хроматографическое разделение проводили с использованием хроматографа LC-20AD Prominence (Shimadzu, Япония), оснащенного автодозатором SIL-20AC (Shimadzu, Япония), термостатированным при 10 °C. Объем пробы составлял 2 мкл. Элюент А – 5 % ацетонитрил в водном растворе 20 мМ $(NH_4)_2CO_3$, доведенный до pH = 9.8 водным раствором аммиака. Элюент Б – 100 % ацетонитрил. Согласно методике, каждый образец анализировали дважды: в режимах гидрофильной (ГФХ) и обращенно-фазовой (ОФХ) хроматографии. Условия хроматографического разделения были следующие. Градиент ГФХ: 0 мин – 98 % Б, 2 мин – 98 % Б, 6 мин – 0 % Б, 10 мин – 0 % Б. Далее колонка уравновешивалась на протяжении 4 мин. Градиент ОФХ: 0 мин – 0 % Б, 1 мин – 0 % Б, 6 мин – 98 % Б, 16 мин – 98 % Б. Далее колонка уравновешивалась на протяжении 3 мин. Скорость потока в каждом из анализов составляла 300 мкл/мин. Анализ в обоих режимах хроматографии проводился при использовании монолитной колонки на основе 1-винил-1,2,4-триазола, 2 × 60 мм. Монолитный материал колонки был синтезирован в соответствии с методикой, описанной в работе (Patrushev et al., 2018): в стеклянной трубке с внутренним диаметром 2 мм выполнялась сополимеризация смеси стирола, дивинилбензола и 1-винил-1,2,4-триазола, взятых в объемном соотношении 10 : 50 : 40 соответственно.

Масс-спектрометрическую детекцию 489 метаболитов проводили в режиме MRM (multiple reaction monitoring – мониторинг множественных реакций) в виде положительных и отрицательных ионов с использованием масс-спектрометра API 6500 QTRAP (AB SCIEX, США), оснащенного источником ионизации электрораспылением. Основные масс-спектрометрические параметры: напряжение ионного распыления (IS) = 5500 В для режима положительной ионизации и –4500 В для режима отрицательной ионизации; температура газа-осушителя (TEM) –

Таблица 1. Схемы шаблонов молекулярно-генетических путей регуляции ферментов метаболических путей белками человека

№	Название шаблона	Схема шаблона регуляторного пути
P1	Белок-белковые взаимодействия	Hp – белок-белковые взаимодействия → Kp
P2	Регуляция функции белка	Hp – регуляция активности/деградации/протеолиза/транспорта → Kp
P3	Регуляция экспрессии	Hp – регуляция экспрессии → Kg – экспрессия → Kp

Примечание. Hp – белки человека; Kg – гены, кодирующие ферменты метаболического пути KEGG; Kp – ферменты метаболического пути KEGG.

475 °C, газ в ячейке соударений (CAD) – средний, давление газа-распылителя (GS1), газа-осушителя (GS2) и газовой завесы (CUR) – 33, 33 и 30 psi соответственно. Потенциал декластеризации (DP) составлял ±91 В, потенциал входа (EP) – ±10 В, потенциал на выходе из ячейки соударений (CXP) – ±9 В. Время сканирования перехода MRM – 5 мс. Переходы ионов-предшественников и фрагментных ионов, названия метаболитов, время фрагментации и соответствующие энергии столкновений были адаптированы из работ (Yuan et al., 2012; Li et al., 2017) (Приложение 1)[1]. Управление устройством и сбор информации осуществляли с помощью программного обеспечения Analyst 1.6.3 (AB SCIEX). Хроматограммы обрабатывали в программном обеспечении Skyline 24.0 (Adams et al., 2020).

Предварительная обработка и статистический анализ данных. Статистический анализ результатов метаболомного скрининга глиобластомы и перитуморального пространства проводился с помощью тестов Манна–Уитни и Куккони (Cucconi test), реализованных в пакетах Python (SciPy и Nonparstat). Тест Манна–Уитни использовали для выявления значимых различий между группами, тогда как тест Куккони – для дополнительного подтверждения выявленных различий в условиях неоднородности выборки. Коррекция выбросов осуществлялась следующим образом. Выбросом считалось значение, выходящее за 1.5 межквартильного интервала. Если значение было определено как выброс, оно заменялось на скорректированное по следующему правилу: замена производилась на значение, равное 1.5 × значение межквартильного интервала ± 10^{-5} (значение вычиталось при выбросах «вверх» и прибавлялось при выбросах «вниз»).

Для выявления перепредставленных метаболических путей на основе установленных метаболитов с высокой значимостью использовалось программное онлайн-обеспечение MetaboAnalyst 6.0 (http://www.metaboanalyst.ca/) (Pang et al., 2021) с инструментом «Анализ обогащения» (Enrichment Analysis).

Реконструкция генных сетей проводилась с помощью графического пользовательского интерфейса ANDVisio программно-информационной системы ANDSystem (http://www-bionet.sscc.ru/andvisio/). В модуле «Мастер путей» (Pathway Wizard) программы ANDVisio были созданы шаблоны путей регуляции ферментов выявленных метаболических путей белками человека (табл. 1). Список идентификаторов белков человека извлечен из базы данных SwissProt (https://www.uniprot.org/).

Список белков, ассоциированных с глиобластомой, получен из базы знаний ANDSystem. Анализ сверхпредставленности биологических процессов Gene Ontology проводился с использованием веб-сервиса DAVID (https://david.ncifcrf.gov/) с настройками, заданными по умолчанию.

Результаты

В рамках исследования были собраны и проанализированы образцы глиобластомы и ткани головного мозга, прилегающей к опухоли (17 образцов глиобластомы и 15 образцов перитуморальной ткани). Метаболомный анализ проводился с помощью разработанного нами ранее подхода с использованием метода ВЭЖХ-МС/МС (Basov et al., 2024). Полученные хроматограммы обрабатывали, интегрируя пик каждого метаболита, после чего сравнивали сигнал каждого метаболита между образцами глиобластомы и перитуморальной ткани. По результатам анализа были получены значения площадей пиков 446 метаболитов.

В качестве основного подхода к статистическому анализу результатов метаболомного скрининга был использован тест Манна–Уитни с критическим значением $p < 0.05$. Как дополнительный тест сравнения групп был выбран непараметрический тест Куккони. Мы сопоставили списки метаболитов, удовлетворяющих каждому из тестов при $p < 0.05$, и получили группу метаболитов, удовлетворяющих обоим тестам (табл. 2).

Для полученного набора метаболитов с использованием MetaboAnalyst 6.0 был проведен анализ перепредставленности метаболических путей KEGG (табл. 3). Согласно данному анализу, статистически значимым перепредставленным метаболическим путем является метаболизм сфинголипидов (Sphingolipid metabolism). Еще одним обогащенным маркерами путем оказался метаболический путь KEGG «Биосинтез пантотеновой кислоты и КоА» (p-value = 0.012), однако после поправки на множественное сравнение значение p-value превысило порог значимости 0.05 и составило 0.46.

Среди метаболитов – потенциальных маркеров 5 из 22 (~23 %) относились к церамидам – классу липидных молекул, являющихся основным компонентом клеточных мембран, а 3 метаболита (4-фосфопантотеновая кислота, малонил-КоА и кофермент А) – к основным предшественникам в биосинтезе липидов *de novo* (см. табл. 2).

[1] Приложения 1–5 см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx31.xlsx

Таблица 2. Метаболиты, содержание которых достоверно ($p < 0.05$) различается в образцах глиобластомы и перитуморального пространства

№ п/п	Название метаболита		*p*-value	
	на английском языке	на русском языке (при наличии)	Тест Куккони	Тест Манна–Уитни
1	Malonyl-CoA	Малонил-КоА	9.99×10^{-4}	1.17×10^{-4}
2	SM (d18:1/22:0 OH)	Сфингомиелин (d18:1/22:0 OH)	3.00×10^{-3}	2.52×10^{-3}
3	Octanoylcarnitine	Октаноилкарнитин	8.99×10^{-3}	4.10×10^{-3}
4	Pyroglutamic acid	Пироглутамовая кислота	2.00×10^{-3}	4.62×10^{-3}
5	Ceramide (d18:1/16:0 OH)*	Церамид (d18:1/16:0 OH)	3.00×10^{-3}	4.62×10^{-3}
6	3-Phosphoglyceric acid	3-Фосфоглицериновая кислота	1.20×10^{-2}	5.20×10^{-3}
7	THC 18:1/20:0	–	1.60×10^{-2}	7.33×10^{-3}
8	Hexose Disaccharide Pool	Пул дисахаридов гексоз	1.30×10^{-2}	8.20×10^{-3}
9	4-Phosphopantothenate	4-Фосфопантотеновая кислота	2.80×10^{-2}	8.20×10^{-3}
10	2-Octenoylcarnitine	2-Октеноилкарнитин	2.80×10^{-2}	4.10×10^{-3}
11	Ceramide (d18:1/16:0)*	Церамид (d18:1/16:0)	4.00×10^{-3}	9.17×10^{-3}
12	Ceramide (d18:1/22:0)*	Церамид (d18:1/22:0)	8.99×10^{-3}	1.14×10^{-2}
13	Coenzyme A	Кофермент А (КоА)	4.00×10^{-2}	1.27×10^{-2}
14	Pyridoxal	Пиридоксаль	3.10×10^{-2}	1.41×10^{-2}
15	N-carbamoyl-L-aspartate	N-карбамоил-L-аспартат	4.40×10^{-2}	1.74×10^{-2}
16	Citrulline	Цитруллин	2.50×10^{-2}	1.92×10^{-2}
17	Decanoylcarnitine	Деканоилкарнитин	4.70×10^{-2}	2.12×10^{-2}
18	GC (18:2/16:0)	Глицерохолин (18:2/16:0)	2.40×10^{-2}	2.35×10^{-2}
19	Purine	Пурин	2.00×10^{-2}	3.45×10^{-2}
20	Ceramide (d18:1/16:2)*	Церамид (d18:1/16:2)	3.20×10^{-2}	4.14×10^{-2}
21	Ceramide (d18:1/16:1 OH)*	Церамид (d18:1/16:1 OH)	2.50×10^{-2}	4.53×10^{-2}
22	Glycerophosphocholine	Глицерофосфохолин	3.00×10^{-2}	4.95×10^{-2}

* Метаболиты, относящиеся к классу церамидов.

Таблица 3. Перепредставленные метаболические пути KEGG для набора метаболомных маркеров

Метаболический путь (KEGG)	*p*-value	FDR*
Метаболизм сфинголипидов	7.95×10^{-5}	6.36×10^{-3}
Биосинтез пантотеновой кислоты и КоА	11.5×10^{-3}	4.6×10^{-1}

* FDR (False Discovery Rate) – поправка на множественное сравнение.

Содержание церамидов было в два раза и более выше в ткани опухоли, чем в перитуморальной ткани. Кроме того, в группе образцов глиобластомы дисперсия была значительно выше, что свидетельствует о большей гетерогенности группы (рис. 1).

Содержание метаболитов пути синтеза пантотеновой кислоты и КоА в тканях опухоли было достоверно снижено (рис. 2), что может свидетельствовать об их активном расходовании на биосинтез липидов.

На следующем шаге нашего исследования был проведен анализ потенциальных механизмов нарушения функции ферментов метаболизма сфинголипидов. Для этого с помощью ANDSystem были реконструированы молекулярно-генетические регуляторные пути по шаблонам, приведенным в табл. 1. Они описывают потенциальную регуляцию ферментов, входящих в метаболизм сфинголипидов, белками человека (Приложения 2–4). Первым звеном реконструированных регуляторных путей являлись все белки человека, последним – ферменты метаболизма сфинголипидов из базы данных KEGG, участвующие в метаболизме церамидов (ceramide), сфингомиелина (SM, sphingomyelin), GC (glucosylceramide) и THC (trihexosylceramide). Эти ферменты метаболизма сфинголипидов далее будем называть ключевыми. Мы рассмотрели такие взаимодействия между участниками регуляторных путей, как белок-белковые взаимодействия,

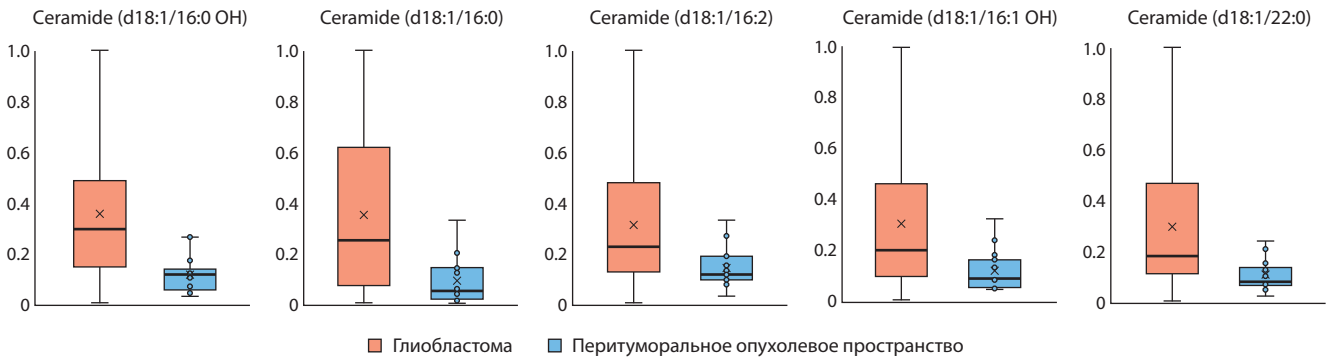


Рис. 1. Содержание маркерных церамидов в ткани опухоли и перитуморального пространства.

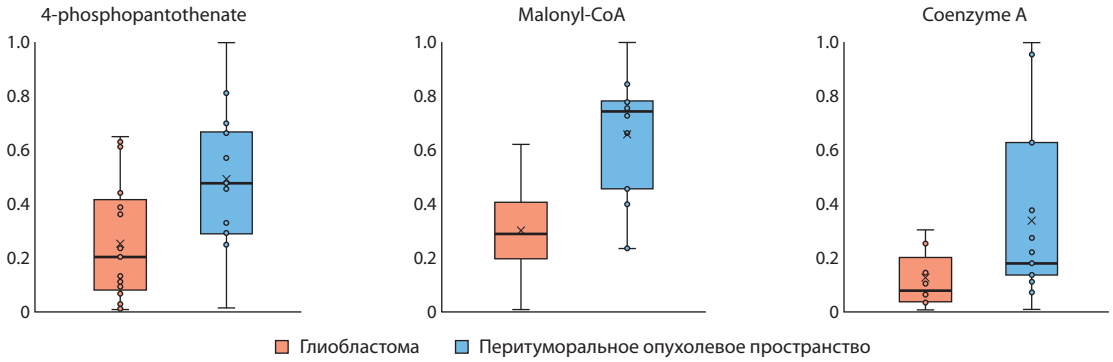


Рис. 2. Содержание КоА и связанных с ним метаболитов в ткани опухоли и перитуморального пространства.

регуляция экспрессии гена, регуляция активности/деградации/транспорта белка.

Реконструированные регуляторные пути были объединены в общую генную сеть (рис. 3). Регуляторная генная сеть включала 15 генов, 329 белков, среди которых 35 ферментов метаболизма сфинголипидов, и 389 взаимодействий между ними.

Согласно базе знаний ANDSystem, для 119 из 294 белков генной сети, регулирующих ключевые ферменты метаболизма сфинголипидов, в литературе приводятся данные о нарушении их функций при глиобластоме. Подсеть регуляторной генной сети, описывающая взаимодействия этих белков с ключевыми ферментами метаболизма сфинголипидов, показана на рис. 4.

Всего в ANDSystem хранится информация об ассоциации 2393 белков человека с глиобластомой, из которых 119 вошли в регуляторную генную сеть. Таким образом, согласно гипергеометрическому тесту, реконструированная генная сеть статистически значимо ассоциирована с глиобластомой (p-value < 10^{-35}).

Анализ перепредставленности биологических процессов Gene Ontology, выполненный для генов результирующей генной сети, показал статистическую значимость 184 процессов, включая апоптоз, сигнальный путь NF-κB, пролиферацию, миграцию и ангиогенез, нарушаемые при многих онкологических заболеваниях, а также такие процессы, как пироптоз, роль которых при глиобластоме активно изучается (Приложения 5).

Обсуждение

Восприимчивость клеток глиобластомы к изменениям содержания метаболитов кофермента А

В нашей работе было установлено пониженное содержание КоА и малонил-КоА в тканях глиобластомы в сравнении с тканями перитуморального пространства (см. рис. 2). Для синтеза жирных кислот *de novo* глиобластома должна генерировать цитозольный ацетил-КоА, который может образовываться из цитрата под действием АТФ-цитратлиазы или из ацетата под действием ацетил-КоА-синтетазы (Santos, Schulze, 2012). В работе (Mashimo et al., 2014) показано, что опухоли головного мозга разного клеточного происхождения обладают способностью окислять введенный ацетат. Авторы утверждают, что для окисления ацетата клетка путем повышения экспрессии активирует изоформу ацетил-КоА-синтетазы ACSS2. Повышенная экспрессия ACSS2 при ГБМ по сравнению с глиомами более низкой степени злокачественности подтверждает утверждение о том, что активация этого фермента связана с увеличением окисления ацетата опухолью. При дефиците ACSS-2 у моделей мышей с гепатоцеллюлярной карциномой снижалась опухолевая нагрузка и ингибировался рост опухоли (Comerford et al., 2014).

Содержание малонил-КоА определяет направление метаболизма жирных кислот: синтез триглицеридов или окисление (Clarke S.D., Nakamura, 2004). Ранее в литературе отмечалось, что ингибирование β-окисления жирных

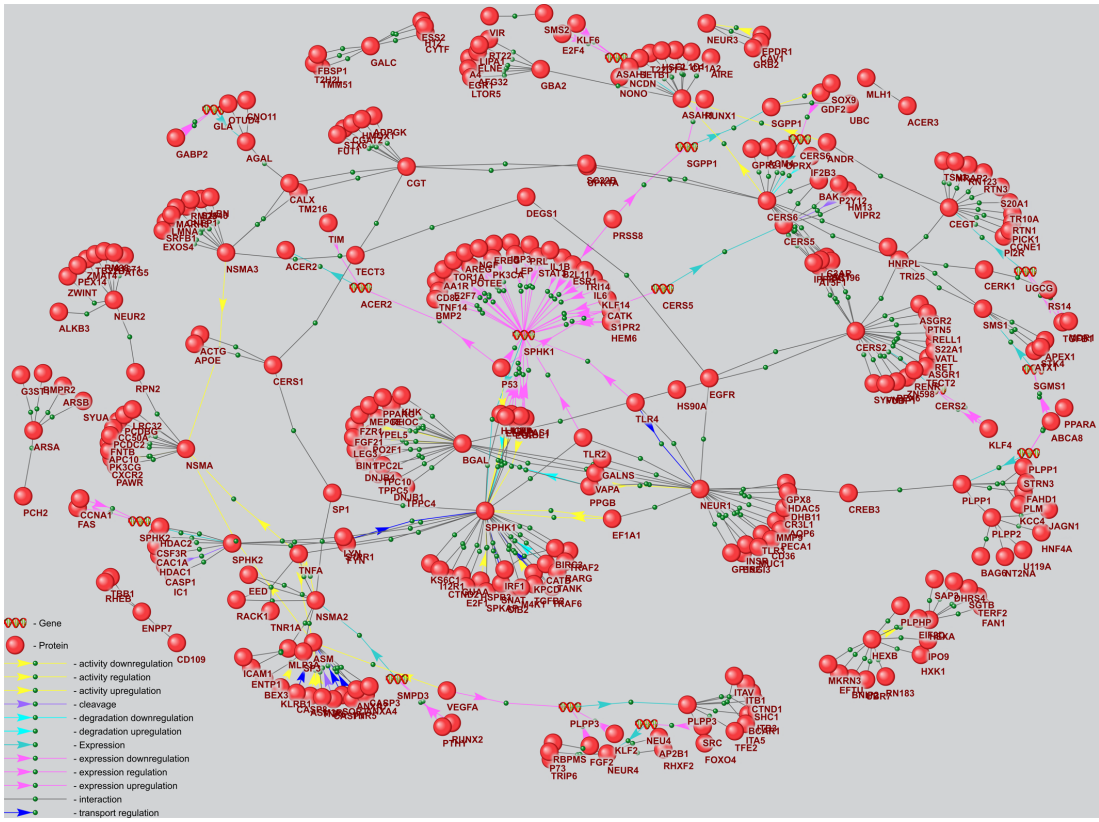


Рис. 3. Генная сеть регуляции ключевых ферментов метаболизма сфинголипидов, реконструированная путем объединения регуляторных путей, построенных по шаблонам трех типов.

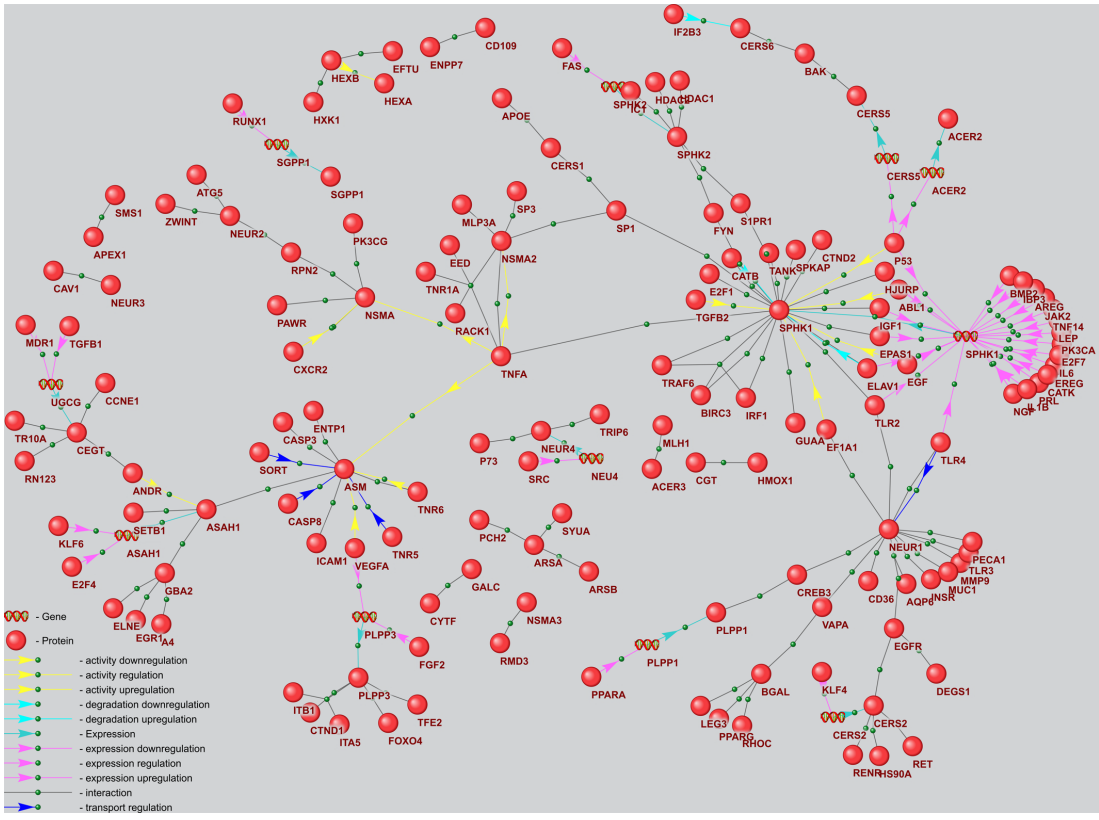


Рис. 4. Подсеть генной сети регуляции ключевых ферментов метаболизма сфинголипидов генетическими маркерами глиобластомы.

кислот в клетках глиобластомы человека этомоксиром, ингибитором карнитин-пальмитоилтрансферазы-1, заметно снижает уровень АТФ, НАДФН и восстановленного глутатиона, что оказывает влияние на их жизнеспособность (Pike et al., 2011). По всей видимости, β-окисление связано с устойчивостью к окислительному стрессу в клетках глиобластомы, а продемонстрированные нами результаты могут свидетельствовать в пользу этой гипотезы. Кроме того, имеются данные о связи содержания малонил-КоА с применением различных химиотерапевтических агентов. Так, в исследовании (Pizer et al., 2000) на клеточной модели рака молочной железы было обнаружено, что уровень малонил-КоА заметно повышался при ингибировании синтазы жирных кислот и снижался при ингибировании ацетил-КоА карбоксилазы. Ключевые метаболические пути, такие как гликолиз и цикл трикарбоновых кислот (ЦТК), характеризуются наличием множества микроРНК, которые участвуют в регуляции этапов этих путей. Раковые клетки в основном используют аэробный гликолиз вместо ЦТК, что позволяет им поддерживать высокий уровень АТФ для обеспечения биосинтетических потребностей (Chan et al., 2015).

В нашей работе было показано возрастание содержания 4-фосфопантотената, что согласуется с изменениями липидного профиля. Синтез этого метаболита осуществляется при катализе пантотенаткиназой – первым ферментом в биосинтетическом пути КоА. Связь пантотенаткиназы с изменениями в глиобластоме активно обсуждается в литературе. Так, в работе (Poli et al., 2010) установлено, что рост клеток линии глиомы U373 снижался при сайленсинге пантотенаткиназы-2. Содержание ацетил-КоА и липидов может контролироваться микроРНК miR-103 и miR-107 (Wilfred et al., 2007). В свою очередь, литературные данные свидетельствуют об ингибировании miR-103 пролиферации и миграции клеток глиобластомы (Chen L.P. et al., 2018) и ингибировании ангиогенеза глиобластомы при повышении экспрессии miR-107 (Chen L. et al., 2016).

Связь биосинтеза церамидов с опухолевым ростом

Церамиды представляют собой липидные медиаторы класса сфинголипидов, которые участвуют в сигнальных путях, контролирующих пролиферацию, дифференцировку и гибель клеток (Riboni et al., 2002). В нашем исследовании показано, что содержание церамидов (16:0), (16:0 OH), (16:2), (16:1 OH), (22:0) – производных сфингомиелина (18:1) – повышено в ткани опухоли в сравнении с тканью перитуморального пространства (см. рис. 1). Ткань перитуморального пространства использовалась нами как контроль ввиду возможности получения ее образцов от пациентов оперативным путем, без вреда для прогноза лечения. Исходя из повышения содержания церамидов в опухолевой ткани, можно предположить изменения на уровне ферментных систем биосинтеза и деградации церамидов, которые могут быть связаны с опухолевым ростом и уходом опухолевых клеток от апоптоза.

Существует три основных пути образования церамидов. Сфингомиелиназный путь включает действие фермента сфингомиелиназы, который расщепляет сфингомиелин в клеточной мембране для высвобождения церамидов. Церамиды могут также синтезироваться *de novo* путем синтеза из более простых молекул. Реутилизационный путь (salvage pathway) включает повторное использование сфинголипидов, которые расщепляются до сфингозина, через их реалкилирование для образования церамида.

Ключевой фермент сфингомиелиназного пути – сфингомиелиназа (SMase), катализирующая гидролиз сфингомиелина. Поскольку сфингомиелин является одним из наиболее представленных в клеточной мембране фосфолипидов, значимость этого пути синтеза церамидов заключается в том, что клеточная мембрана становится мишенью для внеклеточных сигналов, приводящих к программированной гибели клеток и клеточному стрессу (Haimovitz-Friedman et al., 1994). Различают три основных типа SMase: кислая (aSMase), нейтральная (nSMase) и щелочная (alk-SMase). Стимуляция активности сфингомиелиназы может быть вызвана разными факторами, в том числе действием противоопухолевых препаратов. Ингибиторы сфингомиелиназы, такие как перфеназин и флуфеназин, входящие в состав FIASMA (функционального ингибитора кислой сфингомиелиназы), могут применяться в терапии рака, хотя для подтверждения их эффективности необходимы дальнейшие исследования (Kornhuber et al., 2010). В недавних работах описаны ингибитор Arc39 (лабораторный код), блокирующий как лизосомальную, так и секреторную aSMase *in vitro* в клетках L929, HepG2 и B16 (Naser et al., 2020), и светоиндуцируемый ингибитор PCAI, способный ингибировать aSMase (Prause et al., 2020). Кроме того, сфингомиелиназа имеет важное значение в метаболизме сфинголипидов, что может оказывать влияние на развитие рака (Clarke C.J. et al., 2011). Ингибирование сфингомиелиназы и его влияние на экзосомы представляют интерес для онкологии и терапевтических подходов (Lin M. et al., 2018).

Семейство сфингомиелин-синтаз (SMS), состоящее из трех представителей: SMS1, SMS2 и SMS-родственного белка (SMSr) (Chen Y., Cao, 2017), катализирует синтез сфингомиелинов из церамидов (Cer) и фосфатидилхолина, тем самым высвобождая диацилглицерин. Продемонстрировано, что селективное ингибирование SMS приводит к увеличению концентрации церамидов в эндоплазматическом ретикулуме и запускает аутофагию в нейронах гиппокампа (Gulbins et al., 2018). В то же время при глиобластоме было замечено, что лечение 2-гидроксиолеиновой кислотой, противоопухолевым препаратом, повышало активность SMS. Активация SMS2 вызывает снижение уровня церамидов и стимулирует пролиферацию клеток посредством активации сигнального пути трансформирующего фактора роста-β (TGF-β)/Smad. Напротив, ингибирование SMS2 специфической миРНК приводило к накоплению церамидов и ускорению гибели клеток (Zheng et al., 2019). В недавних работах обнаружено, что SMS2 активируется при раке молочной железы и индуцирует поляризацию макрофагов и прогрессирование опухоли (Deng et al., 2021). Более того, ингибирование SMS2 с помощью нокдауна снижало высвобождение цитокинов, которые стимулируют поляризацию макрофагов в М2-макрофаги, уменьшая рост опухоли (Deng et al.,

2021). У пациентов с метастатической меланомой была обнаружена понижающая регуляция SMS1, связанная с ухудшением прогноза из-за дисбаланса между уровнями сфингомиелина и глюкозилцерамида (Bilal et al., 2019).

Серинопальмитоилтрансфераза (SPT) – гетеромерный фермент, состоящий из трех субъединиц, который катализирует первый этап синтеза церамидов *de novo*. SPT катализирует конденсацию L-серина и пальмитоил-кофермента А с образованием 3-кетосфинганина. В нескольких исследованиях отмечено повышение активности SPT в ответ на химиотерапию и лучевую терапию при различных видах рака. К настоящему времени описаны некоторые ингибиторы SPT, которые блокируют рост опухоли. В частности, мириоцин (также известный как ISP-1) является мощным ингибитором SPT (Glaros et al., 2007), способным вызывать подавление роста клеток рака молочной железы (Ogretmen, 2018) и меланомы B16F10 за счет остановки фазы G2/M (Lee et al., 2011). Существует положительная корреляция между ингибированием фермента SPT и подавлением роста клеточной линии аденокарциномы легкого человека (HCC4006) (Sano et al., 2017). Помимо того, ингибирование SPT мириоцином или снижение экспрессии фермента с помощью специфической миРНК ингибировало пролиферацию клеток глиобластомы человека U87MG – действие, которое было связано с подавлением внутриклеточных уровней S1P (Bernhart et al., 2015). Было высказано предположение, что противоопухолевая активность SPT обусловлена повышением уровня проапоптотических церамидов. В действительности у некоторых противоопухолевых препаратов терапевтическая активность по крайней мере частично обусловлена активацией SPT. В частности, фенретинид, синтетический ретиноид, оказывающий цитотоксический эффект для различных раковых клеток, повышал концентрации дезиксицерамидов и церамидов. Это приводило к апоптозу клеток нейробластомы, что было связано со взаимодействием фенретинида с SPT (Maurer et al., 1999).

Церамидсинтаза (CerS) участвует как в процессе синтеза церамидов *de novo*, так и в реутилизационном пути. CerS представляет собой семейство из шести разных изоформ, каждая из которых синтезирует церамиды с различной длиной цепи жирного ацил-КоА, что приводит к определенной активности. Этот факт важен, поскольку, например, CerS1 генерирует церамиды (18:0), для которых показано ингибирование роста опухоли (Wang Z. et al., 2017), тогда как CerS5/6 продуцирует церамиды (16:0) – антиапоптотические метаболиты при плоскоклеточном раке головы и шеи человека (Moro et al., 2019). Специфический ингибитор CerS1, называемый P053, снижает уровень церамида (18:0) в клетках HEK 293, полученных из эмбриональных почек человека (Turner et al., 2018). Интересно, что аналоги, полученные из финголимода (FTY720), специфически ингибируют изоформы CerS в клетках HCT-106 и HeLa. В частности, ингибиторы ST1058 и ST1074 преимущественно ингибируют CerS2 и CerS4, в то время как ST1072 блокирует активность CerS4 и CerS6, а ST1060 ингибирует CerS2 (Schiffmann et al., 2012).

Дигидроцерамиддесатураза (Des1, DEGS1) – последний фермент в процессе синтеза церамида *de novo*. Этот фермент катализирует превращение дигидроцерамида в церамид путем введения транс-двойной связи в положении (C_4-C_5). Нокдаун Des1 с помощью миРНК приводил к остановке клеточного цикла в клетках нейробластомы человека (Kraveka et al., 2007). Например, было показано, что ресвератрол, который является полифенолом с хорошо известными антиоксидантными свойствами, ингибирует Des1, вызывая аутофагию в клетках рака желудка HGC27 (Signorelli et al., 2009). Другие ингибиторы Des1, включая γ-токотриенол, феноксодиол или целекоксиб, индуцируют аутофагию путем накопления дигидроцерамида в клеточных линиях глиобластомы T98G и U87MG за счет ингибирования Des1 (Signorelli et al., 2009). N-[(1R,2S)-2-гидрокси-1-гидроксиметил-2-(2-тридецил-1-циклопропенил)этил]октанамид является специфическим ингибитором Des1, который эффективно активирует аутофагию и апоптоз клеточной линии глиомы человека U87MG. Обработка клеток тетрагидроканнабинолом вызывает изменение липидного состава эндоплазматического ретикулума и, наряду с этим, накопление дигидроцерамида из-за снижения экспрессии Des1, что приводит к стимуляции аутофагии и апоптоза в клетках глиомы U87MG человека (Hernández-Tiedra et al., 2016).

Глюкозилцерамидсинтаза (GCS) – лизосомальный фермент, который осуществляет гликозилирование церамидов с образованием гликозилцерамидов. Показано, что уровень GCS повышается при различных типах рака, что связано с устойчивостью к противоопухолевым методам лечения (Madigan et al., 2020).

Церамидаза – фермент, отщепляющий остатки жирных кислот от церамидов с образованием сфингозинов. Избыточная экспрессия кислой церамидазы (ASAH1) была выявлена в меланомах, что, вероятно, приводит к резистентности к химиотерапии. Обнаружено участие ASAH1 в функционировании митохондрий и клеточной аутофагии в клетках меланомы (Lai M. et al., 2021). Было описано, что alk-SMase тесно связана с ростом, миграцией и инвазией опухолевых клеток (Zhang et al., 2020). Структурные аналоги церамидов продемонстрировали свою эффективность в качестве селективных ингибиторов церамидсинтазы, блокирующих рост клеток и являющихся подходящими кандидатами для противоопухолевого лечения (Steiner et al., 2016).

Нарушение генетической регуляции метаболизма сфинголипидов при глиобластоме

Применение ANDSystem позволило реконструировать генную сеть, описывающую регуляцию ключевых ферментов метаболизма сфинголипидов (см. рис. 3). Анализ регуляторной генной сети показал, что 119 входящих в нее белков ассоциированы с глиобластомой. Это подтверждает связь реконструированной генной сети с данным заболеванием (p-value < 10^{-35}). Среди наиболее изученных белков, ассоциированных с глиобластомой и вошедших в регуляторную генную сеть, оказались такие, как p53, TNF-α, TGF-β, VEGF, KLF4, E2F4.

Известно, что p53 участвует во многих внутриклеточных процессах, а нарушение его функции наблюдается при многих онкологических заболеваниях. В частности,

p53 играет ключевую роль в метаболизме сфинголипидов. p53 регулирует активность пяти ключевых ферментов метаболизма сфинголипидов (CerS5, CerS6, SMPD3, ACER2, SPHK1) из 35, представленных в реконструированной генной сети. В статье (Lacroix et al., 2020) показано, что белок p53 в опухолевых клетках повышает экспрессию церамидсинтаз 5 (CerS5) и 6 (CerS6), нейтральной сфингомиелиназы 2 (SMPD3), трех ферментов, синтезирующих церамиды. В свою очередь, индукция транскрипции alk-SMase-2 белком p53 была исследована в работах (Wang Y. et al., 2017; Xu et al., 2018).

Согласно регуляторной генной сети, фактор некроза опухоли альфа (TNF-α) стимулирует активность сразу трех ферментов метаболизма сфинголипидов, среди которых кислая сфингомиелиназа (ASM), нейтральная сфингомиелиназа (NSMA) и нейтральная сфингомиелиназа 2 (NSMA2). Повышая активность этих ферментов, TNF-α может способствовать гидролизу сфингомиелина и образованию церамидов.

Трансформирующий фактор роста-бета (TGF-β) играет важную роль во многих типах клеток. Он инициирует клеточную передачу сигналов, приводя к активации нижестоящих субстратов и регуляторных белков, тем самым индуцируя транскрипцию различных генов-мишеней. В регуляторной генной сети TGF-β2 повышает активность сфингозинкиназы-1 (SPHK1). Данный факт подтвержден в статье (Ren et al., 2009).

Установлено, что VEGF является компонентом микроокружения опухоли, способным активировать эндотелиальные клетки. Передача сигналов VEGF активирует эндотелиальные клетки через тирозинкиназные рецепторы VEGFR1 и VEGFR2 и способствует миграции, выживанию, пролиферации и дифференцировке клеток, а также инициирует ангиогенез, рост и метастазирование опухоли. В регуляторной генной сети фактор роста эндотелия сосудов (VEGF) участвует в подавлении активности ASM и регуляции экспрессии фосфолипид фосфатазы 3 (PLPP3). При глиобластоме наблюдается высокая степень васкуляризации, а также сверхэкспрессия VEGF, что делает этот ген привлекательной мишенью для терапии глиобластом (Tea et al., 2020).

Круппель-подобный фактор 4 (KLF4) участвует в регуляции пролиферации, дифференцировки, апоптоза и в перепрограммировании соматических клеток. Имеющиеся данные свидетельствуют также о том, что KLF4 является супрессором опухолей при некоторых видах рака (El-Karim et al., 2013). В регуляторной генной сети KLF4 модулирует экспрессию гена церамидсинтазы 2 (CerS2). Действительно, анализ иммунопреципитации хроматина показал (Fan et al., 2015), что KLF4 напрямую связывается с промоторной областью *CerS2* и активирует его экспрессию.

Согласно реконструированным связям, представленным в генной сети, белок E2F4 регулирует экспрессию ASAH1. Литературные данные показывают, что E2F4 действует как репрессор транскрипции, играя ключевую роль в подавлении генов, связанных с пролиферацией. Мутации и сверхэкспрессия гена E2F4 могут быть связаны с раком человека. Связываясь с промотором гена ASAH1, E2F4 подавляет его экспрессию (Melland-Smith et al., 2015).

Значимые биологические процессы, ассоциированные с генной сетью

Перепредставленные биологические процессы, в которые вовлечены участники регуляторной генной сети (см. Приложение 5), можно разделить на несколько групп, включая программируемую клеточную гибель, клеточную мобильность, ангиогенез, пролиферацию, ассоциация которых с онкологическими заболеваниями описана в литературе (Hanahan, Weinberg, 2000). Особого интереса заслуживает такой процесс программируемой клеточной гибели, как пироптоз, роль которого в развитии и прогрессировании глиобластомы активно обсуждается в последние годы (Lin J. et al., 2022). Этот процесс в генной сети был представлен рядом каспаз (CASP1, CASP3 и CASP8), а также нейтрофильной эластазой (Neutrophil elastase), которая в определенных ситуациях расщепляет гасдермин D (GSDMD) и активирует пироптоз либо способствует расщеплению GSDMB, тем самым ингибируя пироптоз (Kambara et al., 2018; Oltra et al., 2023). Таким образом, белки генной сети, относящиеся к пироптозу, вносят значительный вклад в его функционирование (Rao et al., 2022) (см. Приложение 5).

Ангиогенез способствует обеспечению глиобластомы питательными веществами и кислородом, что необходимо для поддержания опухолевого роста (Lara-Velazquez et al., 2017). В процесс ангиогенеза оказались вовлечены участники генной сети, важные генетические маркеры глиобластомы: фактор роста сосудов A (VEGFA), эпидермальный фактор роста (EGF), каталитическая субъединица A фосфатидилинозитол-4,5-бифосфат-3-киназы (PIK3CA) и др. Данные гены играют ключевую роль в ангиогенезе (Danielsen, Rofstad, 1998). Глиобластома обладает высокой способностью к инвазии – ее клетки проникают в окружающие ткани мозга, что делает полное хирургическое удаление опухоли затруднительным, а иногда невозможным (Vollmann-Zwerenz et al., 2020). С клеточной миграцией в регуляторной генной сети были связаны такие белки, как TRIP6 (Thyroid receptor-interacting protein 6), сверхэкспрессирующийся при глиобластоме и способствующий инвазии опухолевых клеток (Lai Y.-J. et al., 2010), TGF-β1, ITAV (Integrin alpha-V), CREB3 (Cyclic AMP-responsive element-binding protein 3) и др.

Заключение

С использованием метода ВЭЖХ-МС/МС был проведен таргетированный метаболомный скрининг глиобластомы и ткани перитуморального пространства онкологических пациентов. Биоинформатический анализ полученных метаболомных профилей, основанный на статистических методах и реконструкции генных сетей, позволил получить новые данные о механизмах развития и прогрессирования глиобластомы. Было показано, что ткани глиобластомы обладают измененным метаболизмом кофермента А и связанных с ним метаболитов, что отличает их от клеток перитуморального пространства. Наши результаты продемонстрировали пониженное содержание КоА и малонил-КоА в тканях глиобластомы, что может быть связано с усилением β-окисления жирных кислот и устойчивостью клеток глиобластомы к окислительному стрессу. Дополнительно мы выявили повышение содержания церамидов

в опухолевой ткани, что указывает на возможное изменение активности ферментов, участвующих в их синтезе и деградации, и на их связь с опухолевым ростом. Данные подтверждают, что нарушение липидного метаболизма, включая метаболизм КоА и церамидов, играет важную роль в патогенезе глиобластомы. Эти изменения могут стать мишенями для разработки новых терапевтических стратегий, направленных на нарушенные метаболические пути опухолевых клеток. В частности, ингибирование ключевых ферментов, таких как серинопальмитоилтрансфераза и сфингомиелиназа, представляется многообещающим подходом для снижения жизнеспособности клеток глиобластомы и предотвращения их дальнейшего роста.

Таким образом, результаты настоящей работы расширяют понимание метаболических особенностей глиобластомы и открывают новые перспективы для разработки целевых терапевтических стратегий, направленных на нарушение метаболизма липидов в опухолевых клетках. Дальнейшие исследования специфических метаболических изменений в различных типах глиобластомы, а также разработка и тестирование ингибиторов ключевых ферментов могут привести к значительному прогрессу в лечении данного заболевания.

Список литературы / References

Антропова Е.А., Хлебодарова Т.М., Деменков П.С., Вензель А.С., Иванисенко Н.В., Гавриленко А.Д., Иванисенко Т.В., Адамовская А.В., Ревва П.М., Лаврик И.Н., Иванисенко В.А. Компьютерный анализ особенностей регуляции гиперметилированных маркерных генов гепатокарциномы вирусными белками гепатита С. *Вавиловский журнал генетики и селекции*. 2022;26(8):733-742. doi 10.18699/VJGB-22-89

[Antropova E.A., Khlebodarova T.M., Demenkov P.S., Venzel A.S., Ivanisenko N.V., Gavrilenko A.D., Ivanisenko T.V., Adamovskaya A.V., Revva P.M., Lavrik I.N., Ivanisenko V.A. Computer analysis of regulation of hepatocarcinoma marker genes hypermethylated by HCV proteins. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):733-742. doi 10.18699/VJGB-22-89]

Деменков П.С., Антропова Е.А., Адамовская А.В., Мищенко Е.Л., Хлебодарова Т.М., Иванисенко Т.В., Иванисенко Н.В., Вензель А.С., Лаврик И.Н., Иванисенко В.А. Приоритизация потенциальных фармакологических мишеней для создания лекарств против гепатокарциномы, модулирующих внешний путь апоптоза, на основе реконструкции и анализа ассоциативных генных сетей. *Вавиловский журнал генетики и селекции*. 2023;27(7):784-793. doi 10.18699/VJGB-23-91

[Demenkov P.S., Antropova E.A., Adamovskaya A.V., Mishchenko E.L., Khlebodarova T.M., Ivanisenko T.V., Ivanisenko N.V., Venzel A.S., Lavrik I.N., Ivanisenko V.A. Prioritization of potential pharmacological targets for the development of anti-hepatocarcinoma drugs modulating the extrinsic apoptosis pathway: the reconstruction and analysis of associative gene networks help. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):784-793. doi 10.18699/VJGB-23-91]

Иванисенко В.А., Басов Н.В., Макарова А.А., Вензель А.С., Рогачев А.Д., Деменков П.С., Иванисенко Т.В., Клещев М.А., Гайслер Е.В., Мороз Г.Б., Плеско В.В., Сотникова Ю.С., Патрушев Ю.В., Ломиворотов В.В., Колчанов Н.А., Покровский А.Г. Применение генных сетей к анализу результатов метаболомного скрининга плазмы крови пациентов с послеоперационным делирием. *Вавиловский журнал генетики и селекции*. 2023;27(7):768-775. doi 10.18699/VJGB-23-89

[Ivanisenko V.A., Basov N.V., Makarova A.A., Venzel A.S., Rogachev A.D., Demenkov P.S., Ivanisenko T.V., Kleshchev M.A., Gaisler E.V., Moroz G.B., Plesko V.V., Sotnikova Y.S., Patrushev Y.V., Lomivorotov V.V., Kolchanov N.A., Pokrovsky A.G. Gene networks for use in metabolomic data analysis of blood plasma from patients with postoperative delirium. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):768-775. doi 10.18699/VJGB-23-89]

Колчанов Н.А., Игнатьева Е.В., Подколодная О.А., Лихошвай В.А., Матушкин Ю.Г. Генные сети. *Вавиловский журнал генетики и селекции*. 2013;17(4/2):833-850

[Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/2):833-850 (in Russian)]

Adams K.J., Pratt B., Bose N., Dubois L.G., St. John-Williams L., Perrott K.M., Ky K., Kapahi P., Sharma V., Maccoss M.J., Moseley M.A., Colton C.A., Maclean B.X., Schilling B., Thompson J.W. Skyline for small molecules: a unifying software package for quantitative metabolomics. *J. Proteome Res*. 2020;19(4):1447-1458. doi 10.1021/acs.jproteome.9b00640

Basov N.V., Rogachev A.D., Aleshkova M.A., Gaisler E.V., Sotnikova Y.S., Patrushev Y.V., Tolstikova T.G., Yarovaya O.I., Pokrovsky A.G., Salakhutdinov N.F. Global LC-MS/MS targeted metabolomics using a combination of HILIC and RP LC separation modes on an organic monolithic column based on 1-vinyl-1,2,4-triazole. *Talanta*. 2024;267:125168. doi 10.1016/j.talanta.2023.125168

Bernhart E., Damm S., Wintersperger A., Nusshold C., Brunner A.M., Plastira I., Rechberger G., Reicher H., Wadsack C., Zimmer A., Malle E., Sattler W. Interference with distinct steps of sphingolipid synthesis and signaling attenuates proliferation of U87MG glioma cells. *Biochem. Pharmacol*. 2015;96(2):119-130. doi 10.1016/j.bcp.2015.05.007

Bilal F., Montfort A., Gilhodes J., Garcia V., Riond J., Carpentier S., Filleron T., Colacios C., Levade T., Daher A., Meyer N., Andrieu-Abadie N., Ségui B. Sphingomyelin synthase 1 (SMS1) downregulation is associated with sphingolipid reprogramming and a worse prognosis in melanoma. *Front. Pharmacol*. 2019;10:443. doi 10.3389/fphar.2019.00443

Binder H., Wirth H., Arakelyan A., Lembcke K., Tiys E.S., Ivanisenko V.A., Kolchanov N.A., Kononikhin A., Popov I., Nikolaev E.N., Pastushkova L.K., Larina I.M. Time-course human urine proteomics in space-flight simulation experiments. *BMC Genomics*. 2014;15(Suppl.12):S2. doi 10.1186/1471-2164-15-S12-S2

Bragina E.Y., Tiys E.S., Freidin M.B., Koneva L.A., Demenkov P.S., Ivanisenko V.A., Kolchanov N.A., Puzyrev V.P. Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis. *Immunogenetics*. 2014;66(7-8):457-465. doi 10.1007/s00251-014-0786-1

Bragina E.Y., Tiys E.S., Rudko A.A., Ivanisenko V.A., Freidin M.B. Novel tuberculosis susceptibility candidate genes revealed by the reconstruction and analysis of associative networks. *Infect. Genet. Evol*. 2016;46:118-123. doi 10.1016/j.meegid.2016.10.030

Bragina E.Y., Gomboeva D.E., Saik O.V., Ivanisenko V.A., Freidin M.B., Nazarenko M.S., Puzyrev V.P. Apoptosis genes as a key to identification of inverse comorbidity of Huntington's disease and cancer. *Int. J. Mol. Sci*. 2023;24(11):9385. doi 10.3390/ijms24119385

Chan B., Manley J., Lee J., Singh S.R. The emerging roles of microRNAs in cancer metabolism. *Cancer Lett*. 2015;356(2, Part A):301-308. doi 10.1016/j.canlet.2014.10.011

Chen L., Li Z.Y., Xu S.Y., Zhang X.J., Zhang Y., Luo K., Li W.P. Upregulation of miR-107 inhibits glioma angiogenesis and VEGF expression. *Cell. Mol. Neurobiol*. 2016;36(1):113-120. doi 10.1007/s10571-015-0225-3

Chen L.P., Zhang N.N., Ren X.Q., He J., Li Y. miR-103/miR-195/miR-15b regulate SALL4 and inhibit proliferation and migration in glioma. *Molecules*. 2018;23(11):2938. doi 10.3390/molecules23112938

Chen Y., Cao Y. The sphingomyelin synthase family: proteins, diseases, and inhibitors. *Biol. Chem*. 2017;398(12):1319-1325. doi 10.1515/hsz-2017-0148

Chinnaiyan P., Kensicki E., Bloom G., Prabhu A., Sarcar B., Kahali S., Eschrich S., Qu X., Forsyth P., Gillies R. The metabolomic signature of malignant glioma reflects accelerated anabolic metabolism. *Cancer Res*. 2012;72(22):5878-5888. doi 10.1158/0008-5472.CAN-12-1572-T

Clarke C.J., Cloessner E.A., Roddy P.L., Hannun Y.A. Neutral sphingomyelinase 2 (nSMase2) is the primary neutral sphingomyelinase isoform activated by tumour necrosis factor-α in MCF-7 cells. *Biochem. J*. 2011;435(2):381-390. doi 10.1042/BJ20101752

Clarke S.D., Nakamura M.T. Fatty acid synthesis and its regulation. In: Encyclopedia of Biological Chemistry. Acad. Press, 2004;99-103. doi 10.1016/B0-12-443710-9/00224-6

Comerford S.A., Huang Z., Du X., Wang Y., Cai L., Witkiewicz A.K., Walters H., Tantawy M.N., Fu A., Manning H.C., Horton J.D., Hammer R.E., Mcknight S.L., Tu B.P. Acetate dependence of tumors. *Cell*. 2014;159(7):1591-1602. doi 10.1016/j.cell.2014.11.020

Danielsen T., Rofstad E.K. VEGF, bFGF and EGF in the angiogenesis of human melanoma xenografts. *Int. J. Cancer*. 1998;76(6):836-841. doi 10.1002/(sici)1097-0215(19980610)76:6<836::aid-ijc12>3.0.co;2-0

Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol*. 2011;11(3-4):149-161. doi 10.3233/ISB-2012-0449

Deng Y., Hu J.C., He S.H., Lou B., Ding T.B., Yang J.T., Mo M.G., Ye D.Y., Zhou L., Jiang X.C., Yu K., Dong J.B. Sphingomyelin synthase 2 facilitates M2-like macrophage polarization and tumor progression in a mouse model of triple-negative breast cancer. *Acta Pharmacol. Sin*. 2021;42(1):149-159. doi 10.1038/s41401-020-0419-1

El-Karim E.A., Hagos E.G., Ghaleb A.M., Yu B., Yang V.W. Krüppel-like factor 4 regulates genetic stability in mouse embryonic fibroblasts. *Mol. Cancer*. 2013;12:89. doi 10.1186/1476-4598-12-89

Fan S.H., Wang Y.Y., Wu Z.Y., Zhang Z.F., Lu J., Li M.Q., Shan Q., Wu D.M., Sun C.H., Hu B., Zheng Y.L. AGPAT9 suppresses cell growth, invasion and metastasis by counteracting acidic tumor microenvironment through KLF4/LASS2/V-ATPase signaling pathway in breast cancer. *Oncotarget*. 2015;6(21):18406-18417. doi 10.18632/oncotarget.4074

Glaros E.N., Kim W.S., Wu B.J., Suarna C., Quinn C.M., Rye K.A., Stocker R., Jessup W., Garner B. Inhibition of atherosclerosis by the serine palmitoyl transferase inhibitor myriocin is associated with reduced plasma glycosphingolipid concentration. *Biochem. Pharmacol*. 2007;73(9):1340-1346. doi 10.1016/j.bcp.2006.12.023

Gulbins A., Schumacher F., Becker K.A., Wilker B., Soddemann M., Boldrin F., Müller C.P., Edwards M.J., Goodman M., Caldwell C.C., Kleuser B., Kornhuber J., Szabo I., Gulbins E. Antidepressants act by inducing autophagy controlled by sphingomyelin-ceramide. *Mol. Psychiatry*. 2018;23(12):2324-2346. doi 10.1038/s41380-018-0090-9

Haimovitz-Friedman A., Kan C.C., Ehleiter D., Persaud R.S., McLoughlin M., Fuks Z., Kolesnick R.N. Ionizing radiation acts on cellular membranes to generate ceramide and initiate apoptosis. *J. Exp. Med*. 1994;180(2):525-535. doi 10.1084/jem.180.2.525

Hanahan D., Weinberg R.A. The hallmarks of cancer. *Cell*. 2000;100(1):57-70. doi 10.1016/s0092-8674(00)81683-9

Heiden M.G.V., Cantley L.C., Thompson C.B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*. 2009;324(5930):1029-1033. doi 10.1126/science.1160809

Hernández-Tiedra S., Fabriàs G., Dávila D., Salanueva Í.J., Casas J., Montes L.R., Antón Z., García-Taboada E., Salazar-Roa M., Lorente M., Nylandsted J., Armstrong J., López-Valero I., McKee C.S., Serrano-Puebla A., García-López R., González-Martínez J., Abad J.L., Hanada K., Boya P., Goñi F., Guzmán M., Lovat P., Jäättelä M., Alonso A., Velasco G. Dihydroceramide accumulation mediates cytotoxic autophagy of cancer cells via autolysosome destabilization. *Autophagy*. 2016;12(11):2213-2229. doi 10.1080/15548627.2016.1213927

Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*. 2020;21(Suppl.11):228. doi 10.1186/s12859-020-03557-8

Ivanisenko T.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. The new version of the ANDDigest tool with improved ai-based short names recognition. *Int. J. Mol. Sci*. 2022;23(23):14934. doi 10.3390/ijms232314934

Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol*. 2015;9(Suppl.2):S2. doi 10.1186/1752-0509-9-S2-S2

Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl.1):34. doi 10.1186/s12859-018-2567-6

Ivanisenko V.A., Gaisler E.V., Basov N.V., Rogachev A.D., Cheresiz S.V., Ivanisenko T.V., Demenkov P.S., Mishchenko E.L., Khripko O.P., Khripko Yu.I., Voevoda S.M., Karpenko T.N., Velichko A.J., Voevoda M.I., Kolchanov N.A., Pokrovsky A.G. Plasma metabolomics and gene regulatory networks analysis reveal the role of nonstructural SARS-CoV-2 viral proteins in metabolic dysregulation in COVID-19 patients. *Sci. Rep*. 2022;12(1):19977. doi 10.1038/s41598-022-24170-0

Jaroch K., Modrakowska P., Bojko B. Glioblastoma metabolomics – in vitro studies. *Metabolites*. 2021;11(5):315. doi 10.3390/metabo11050315.

Kambara H., Liu F., Zhang X., Liu P., Bajrami B., Teng Y., Zhao L., Zhou S., Yu H., Zhou W., Silberstein L.E., Cheng T., Han M., Xu Y., Luo H.R. Gasdermin D exerts anti-inflammatory effects by promoting neutrophil death. *Cell Rep*. 2018;22(11):2924-2936. doi 10.1016/j.celrep.2018.02.067

Koppenol W.H., Bounds P.L., Dang C.V. Otto Warburg's contributions to current concepts of cancer metabolism. *Nat. Rev. Cancer*. 2011;11(5):325-337. doi 10.1038/nrc3038

Kornhuber J., Tripal P., Reichel M., Mühle C., Rhein C., Muehlbacher M., Groemer T.W., Gulbins E. Functional inhibitors of acid sphingomyelinase (FIASMAs): a novel pharmacological group of drugs with broad clinical applications. *Cell Physiol. Biochem*. 2010;26(1):9-20. doi 10.1159/000315101

Kraveka J.M., Li L., Szulc Z.M., Bielawski J., Ogretmen B., Hannun Y.A., Obeid L.M., Bielawska A. Involvement of dihydroceramide desaturase in cell cycle progression in human neuroblastoma cells. *J. Biol. Chem*. 2007;282(23):16718-16728. doi 10.1074/jbc.M700647200

Lacroix M., Riscal R., Arena G., Linares L.K., Le Cam L. Metabolic functions of the tumor suppressor p53: implications in normal physiology, metabolic disorders, and cancer. *Mol. Metab*. 2020;33:2-22. doi 10.1016/j.molmet.2019.10.002

Lai M., La Rocca V., Amato R., Freer G., Costa M., Spezia P.G., Quaranta P., Lombardo G., Piomelli D., Pistello M. Ablation of acid ceramidase impairs autophagy and mitochondria activity in melanoma cells. *Int. J. Mol. Sci*. 2021;22(6):3247. doi 10.3390/ijms22063247

Lai Y.-J., Lin V.T.G., Zheng Y., Benveniste E.N., Lin F.-T. The adaptor protein TRIP6 antagonizes fas-induced apoptosis but promotes its effect on cell migration. *Mol. Cell. Biol*. 2010;30(23):5582-5596. doi 10.1128/MCB.00134-10

Lara-Velazquez M., Al-Kharboosh R., Jeanneret S., Vazquez-Ramos C., Mahato D., Tavanaiepour D., Rahmathulla G., Quinone-Hinojosa A. Advances in brain tumor surgery for glioblastoma in adults. *Brain Sci*. 2017;7(12):166. doi 10.3390/brainsci7120166

Larina I.M., Pastushkova L.K., Tiys E.S., Kireev K.S., Kononikhin A.S., Starodubtseva N.L., Popov I.A., Custaud M.A., Dobrokhotov I.V., Nikolaev E.N., Kolchanov N.A., Ivanisenko V.A. Permanent proteins in the urine of healthy humans during the Mars-500 experiment. *J. Bioinform. Comput. Biol*. 2015;13(1):1540001. doi 10.1142/S0219720015400016

Lee Y.S., Choi K.M., Choi M.H., Ji S.Y., Lee S., Sin D.M., Oh K.W., Lee Y.M., Hong J.T., Yun Y.P., Yoo H.S. Serine palmitoyltransferase inhibitor myriocin induces growth inhibition of B16F10 melanoma cells through G_2/M phase arrest. *Cell Prolif*. 2011;44(4):320-329. doi 10.1111/j.1365-2184.2011.00761.x

Li K., Naviaux J.C., Bright A.T., Wang L., Naviaux R.K. A robust, single-injection method for targeted, broad-spectrum plasma metabolomics. *Metabolomics*. 2017;13(10):122. doi 10.1007/s11306-017-1264-1

Liberti M.V., Locasale J.W. The Warburg effect: how does it benefit cancer cells? *Trends Biochem. Sci*. 2016;41(3):211-218. doi 10.1016/j.tibs.2015.12.001

Lin J., Lai X., Liu X., Yan H., Wu C. Pyroptosis in glioblastoma: a crucial regulator of the tumour immune microenvironment and a predictor of prognosis. *J. Cell. Mol. Med*. 2022;26(5):1579-1593. doi 10.1111/jcmm.17200

Lin M., Liao W., Dong M., Zhu R., Xiao J., Sun T., Chen Z., Wu B., Jin J. Exosomal neutral sphingomyelinase 1 suppresses hepatocellular carcinoma via decreasing the ratio of sphingomyelin/ceramide. *FEBS J*. 2018;285(20):3835-3848. doi 10.1111/febs.14635

Louis D.N., Perry A., Wesseling P., Brat D.J., Cree I.A., Figarella-Branger D., Hawkins C., Ng H.K., Pfister S.M., Reifenberger G., Soffietti R., Von Deimling A., Ellison D.W. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncology*. 2021;23(8):1231-1251. doi 10.1093/neuonc/noab106

Madigan J.P., Robey R.W., Poprawski J.E., Huang H., Clarke C.J., Gottesman M.M., Cabot M.C., Rosenberg D.W. A role for ceramide glycosylation in resistance to oxaliplatin in colorectal cancer. *Exp. Cell Res*. 2020;388(2):111860. doi 10.1016/j.yexcr.2020.111860

Mashimo T., Pichumani K., Vemireddy V., Hatanpaa K.J., Singh D.K., Sirasanagandla S., Nannepaga S., Piccirillo S.G., Kovacs Z., Foong C., Huang Z., Barnett S., Mickey B.E., Deberardinis R.J., Tu B.P., Maher E.A., Bachoo R.M. Acetate is a bioenergetic substrate for human glioblastoma and brain metastases. *Cell*. 2014;159(7):1603-1614. doi 10.1016/j.cell.2014.11.025

Maurer B.J., Metelitsa L.S., Seeger R.C., Cabot M.C., Reynolds C.P. Increase of ceramide and induction of mixed apoptosis/necrosis by N-(4-hydroxyphenyl)-retinamide in neuroblastoma cell lines. *J. Natl. Cancer Inst*. 1999;91(13):1138-1146. doi 10.1093/jnci/91.13.1138

Melland-Smith M., Ermini L., Chauvin S., Craig-Barnes H., Tagliaferro A., Todros T., Post M., Caniggia I. Disruption of sphingolipid metabolism augments ceramide-induced autophagy in preeclampsia. *Autophagy*. 2015;11(4):653-669. doi 10.1080/15548627.2015.1034414

Moro K., Nagahashi M., Gabriel E., Takabe K., Wakai T. Clinical application of ceramide in cancer treatment. *Breast Cancer*. 2019;26(4):407-415. doi 10.1007/s12282-019-00953-8

Naser E., Kadow S., Schumacher F., Mohamed Z.H., Kappe C., Hessler G., Pollmeier B., Kleuser B., Arenz C., Becker K.A., Gulbins E., Carpinteiro A. Characterization of the small molecule ARC39, a direct and specific inhibitor of acid sphingomyelinase in vitro. *J. Lipid Res*. 2020;61(6):896-910. doi 10.1194/jlr.RA120000682

Ogretmen B. Sphingolipid metabolism in cancer signalling and therapy. *Nat. Rev. Cancer*. 2018;18(1):33-50. doi 10.1038/nrc.2017.96

Oltra S.S., Colomo S., Sin L., Pérez-López M., Lázaro S., Molina-Crespo A., Choi K.H., Ros-Pardo D., Martínez L., Morales S., González-Paramos C., Orantes A., Soriano M., Hernández A., Lluch A., Rojo F., Albanell J., Gómez-Puertas P., Ko J.K., Sarrió D., Moreno-Bueno G. Distinct GSDMB protein isoforms and protease cleavage processes differentially control pyroptotic cell death and mitochondrial damage in cancer cells. *Cell Death Differ*. 2023;30(5):1366-1381. doi 10.1038/s41418-023-01143-y

Omuro A., DeAngelis L.M. Glioblastoma and other malignant gliomas: a clinical review. *JAMA*. 2013;310(17):1842-1850. doi 10.1001/jama.2013.280319

Pandey R., Caflisch L., Lodi A., Brenner A.J., Tiziani S. Metabolomic signature of brain cancer. *Mol. Carcinog*. 2017;56(11):2355-2371. doi 10.1002/mc.22694

Pang Z., Chong J., Zhou G., de Lima Morais D.A., Chang L., Barrette M., Gauthier C., Jacques P.-É., Li S., Xia J. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res*. 2021;49(W1):W388-W396. doi 10.1093/nar/gkab382

Pastushkova L.K., Kireev K.S., Kononikhin A.S., Tiys E.S., Popov I.A., Starodubtseva N.L., Dobrokhotov I.V., Ivanisenko V.A., Larina I.M., Kolchanov N.A., Nikolaev E.N. Detection of renal tissue and urinary tract proteins in the human urine after space flight. *PLoS One*. 2013;8(8):e71652. doi 10.1371/journal.pone.0071652

Pastushkova L.K., Kashirina D.N., Brzhozovskiy A.G., Kononikhin A.S., Tiys E.S., Ivanisenko V.A., Koloteva M.I., Nikolaev E.N., Larina I.M. Evaluation of cardiovascular system state by urine proteome after manned space flight. *Acta Astronaut*. 2019;160:594-600. doi 10.1016/j.actaastro.2019.02.015

Patrushev Y., Yudina Y., Sidelnikov V. Monolithic rod columns for HPLC based on divinylbenzene-styrene copolymer with 1-vinylimidazole and 4-vinylpyridine. *J. Liq. Chromatogr. Relat. Technol*. 2018;41(8):458-466. doi 10.1080/10826076.2018.1455149

Pike L.S., Smift A.L., Croteau N.J., Ferrick D.A., Wu M. Inhibition of fatty acid oxidation by etomoxir impairs NADPH production and increases reactive oxygen species resulting in ATP depletion and cell death in human glioblastoma cells. *Biochim. Biophys. Acta*. 2011;1807(6):726-734. doi 10.1016/j.bbabio.2010.10.022

Pizer E.S., Thupari J., Han W.F., Pinn M.L., Chrest F.J., Frehywot G.L., Townsend C.A., Kuhajda F.P. Malonyl-coenzyme-A is a potential mediator of cytotoxicity induced by fatty-acid synthase inhibition in human breast cancer cells and xenografts. *Cancer Res*. 2000;60(2):213-218

Poli M., Derosas M., Luscieti S., Cavadini P., Campanella A., Verardi R., Finazzi D., Arosio P. Pantothenate kinase-2 (Pank2) silencing causes cell growth reduction, cell-specific ferroportin upregulation and iron deregulation. *Neurobiol. Dis.* 2010;39(2):204-210. doi 10.1016/j.nbd.2010.04.009

Popik O.V., Petrovskiy E.D., Mishchenko E.L., Lavrik I.N., Ivanisenko V.A. Mosaic gene network modelling identified new regulatory mechanisms in HCV infection. *Virus Res*. 2016;218:71-78. doi 10.1016/j.virusres.2015.10.004

Poteet E., Choudhury G.R., Winters A., Li W., Ryou M.G., Liu R., Tang L., Ghorpade A., Wen Y., Yuan F., Keir S.T., Yan H., Bigner D.D., Simpkins J.W., Yang S.H. Reversing the Warburg effect as a treatment for glioblastoma. *J. Biol. Chem*. 2013;288(13):9153-9164. doi 10.1074/jbc.M112.440354

Prause K., Naseri G., Schumacher F., Kappe C., Kleuser B., Arenz C. A photocaged inhibitor of acid sphingomyelinase. *Chem. Commun*. 2020;56(94):14885-14888. doi 10.1039/d0cc06661c

Rao Z., Zhu Y., Yang P., Chen Z., Xia Y., Qiao C., Liu W., Deng H., Li J., Ning P., Wang Z. Pyroptosis in inflammatory diseases and cancer. *Theranostics*. 2022;12(9):4310-4329. doi 10.7150/thno.71086

Ren S., Babelova A., Moreth K., Xin C., Eberhardt W., Doller A., Pavenstädt H., Schaefer L., Pfeilschifter J., Huwiler A. Transforming growth factor-B2 upregulates sphingosine kinase-1 activity, which in turn attenuates the fibrotic response to TGF-B2 by impeding CTGF expression. *Kidney Int*. 2009;76(8):857-867. doi 10.1038/ki.2009.297

Riboni L., Campanella R., Bassi R., Villani R., Gaini S.M., Martinelli-Boneschi F., Viani P., Tettamanti G. Ceramide levels are inversely associated with malignant progression of human glial tumors. *Glia*. 2002;39(2):105-113. doi 10.1002/glia.10087

Rogachev A.D., Alemasov N.A., Ivanisenko V.A., Ivanisenko N.V., Gaisler E.V., Oleshko O.S., Cheresiz S.V., Mishinov S.V., Stupak V.V., Pokrovsky A.G. Correlation of metabolic profiles of plasma and cerebrospinal fluid of high-grade glioma patients. *Metabolites*. 2021;11(3):133. doi 10.3390/metabo11030133

Saik O.V., Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Interactome of the hepatitis C virus: literature mining with ANDSystem. *Virus Res*. 2016;218:40-48. doi 10.1016/j.virusres.2015.12.003

Saik O.V., Demenkov P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Dosenko V.E., Zolotareva O.I., Choynzonov E.L., Hofestaedt R., Ivanisenko V.A. Search for new candidate genes involved in the comorbidity of asthma and hypertension based on automatic analysis of scientific literature. *J. Integr. Bioinform*. 2018a;15(4):20180054. doi 10.1515/jib-2018-0054

Saik O.V., Demenkov P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Goncharova I.A., Dosenko V.E., Zolotareva O.I., Hofestaedt R., Lavrik I.N., Rogaev E.I., Ivanisenko V.A. Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. *BMC Med. Genomics*. 2018b;11(Suppl.1): 15. doi 10.1186/s12920-018-0331-4

Saik O.V., Nimaev V.V., Usmonov D.B., Demenkov P.S., Ivanisenko T.V., Lavrik I.N., Ivanisenko V.A. Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema using an analysis of associative gene networks with ANDSystem. *BMC Med. Genomics*. 2019;12(Suppl.2):47. doi 10.1186/s12920-019-0492-9

Sano O., Kazetani K.I., Adachi R., Kurasawa O., Kawamoto T., Iwata H. Using a biologically annotated library to analyze the anticancer mechanism of serine palmitoyl transferase (SPT) inhibitors. *FEBS Open Bio*. 2017;7(4):495-503. doi 10.1002/2211-5463.12196

Santos C.R., Schulze A. Lipid metabolism in cancer. *FEBS J*. 2012; 279(15):2610-2623. doi 10.1111/j.1742-4658.2012.08644.x

Schiffmann S., Hartmann D., Fuchs S., Birod K., Ferreirs N., Schreiber Y., Zivkovic A., Geisslinger G., Grösch S., Stark H. Inhibitors of specific ceramide synthases. *Biochimie*. 2012;94(2):558-565. doi 10.1016/j.biochi.2011.09.007

Siegal T. Clinical impact of molecular biomarkers in gliomas. *J. Clin. Neurosci*. 2015;22(3):437-444. doi 10.1016/j.jocn.2014.10.004

Signorelli P., Munoz-Olaya J.M., Gagliostro V., Casas J., Ghidoni R., Fabriàs G. Dihydroceramide intracellular increase in response to resveratrol treatment mediates autophagy in gastric cancer cells. *Cancer Lett*. 2009;282(2):238-243. doi 10.1016/j.canlet.2009.03.020

Steiner R., Saied E.M., Othman A., Arenz C., Maccarone A.T., Poad B.L.J., Blanksby S.J., Von Eckardstein A., Hornemann T. Elucidating the chemical structure of native 1-deoxysphingosine. *J. Lipid Res*. 2016;57(7):1194-1203. doi 10.1194/jlr.M067033

Tea M.N., Poonnoose S.I., Pitson S.M. Targeting the sphingolipid system as a therapeutic direction for glioblastoma. *Cancers* (*Basel*). 2020;12(1):111. doi 10.3390/cancers12010111

Turner N., Lim X.Y., Toop H.D., Osborne B., Brandon A.E., Taylor E.N., Fiveash C.E., Govindaraju H., Teo J.D., McEwen H.P., Couttas T.A., Butler S.M., Das A., Kowalski G.M., Bruce C.R., Hoehn K.L., Fath T., Schmitz-Peiffer C., Cooney G.J., Montgomery M.K., Morris J.C., Don A.S. A selective inhibitor of ceramide synthase 1 reveals a novel role in fat metabolism. *Nat. Commun*. 2018;9(1):3165. doi 10.1038/s41467-018-05613-7

Vollmann-Zwerenz A., Leidgens V., Feliciello G., Klein C.A., Hau P. Tumor cell invasion in glioblastoma. *Int. J. Mol. Sci*. 2020;21(6): 1932. doi 10.3390/ijms21061932

Wang Y., Zhang C., Jin Y., Wang S., He Q., Liu Z., Ai Q., Lei Y., Li Y., Song F., Bu Y. Alkaline ceramidase 2 is a novel direct target of p53 and induces autophagy and apoptosis through ROS generation. *Sci. Rep*. 2017;7:44573. doi 10.1038/srep44573

Wang Z., Wen L., Zhu F., Wang Y., Xie Q., Chen Z., Li Y. Overexpression of ceramide synthase 1 increases C18-ceramide and leads to lethal autophagy in human glioma. *Oncotarget*. 2017;8(61):104022-104036. doi 10.18632/oncotarget.21955

Warburg O. On the origin of cancer cells. *Science*. 1956;123(3191): 309-314. doi 10.1126/science.123.3191.309

Wilfred B.R., Wang W.X., Nelson P.T. Energizing miRNA research: a review of the role of miRNAs in lipid metabolism, with a prediction that miR-103/107 regulates human metabolic pathways. *Mol. Genet. Metab*. 2007;91(3):209-217. doi 10.1016/j.ymgme.2007.03.011

Wolf A., Agnihotri S., Guha A. Targeting metabolic remodeling in glioblastoma multiforme. *Oncotarget*. 2010;1(7):552-562. doi 10.18632/oncotarget.190

Xu R., Garcia-Barros M., Wen S., Li F., Lin C.L., Hannun Y.A., Obeid L.M., Mao C. Tumor suppressor p53 links ceramide metabolism to DNA damage response through alkaline ceramidase 2. *Cell Death Differ*. 2018;25(5):841-856. doi 10.1038/s41418-017-0018-y

Youngblood M.W., Stupp R., Sonabend A.M. Role of resection in glioblastoma management. *Neurosurg. Clin. N. Am*. 2021;32(1):9-22. doi 10.1016/j.nec.2020.08.002

Yuan M., Breitkopf S.B., Yang X., Asara J.M. A positive/negative ion – switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nat. Protoc*. 2012;7(5):872-881. doi 10.1038/nprot.2012.024

Zhang S., Huang P., Dai H., Li Q., Hu L., Peng J., Jiang S., Xu Y., Wu Z., Nie H., Zhang Z., Yin W., Zhang X., Lu J. TIMELESS regulates sphingolipid metabolism and tumor cell growth through Sp1/ACER2/S1P axis in ER-positive breast cancer. *Cell Death Dis*. 2020;11(10):892. doi 10.1038/s41419-020-03106-4

Zheng K., Chen Z., Feng H., Chen Y., Zhang C., Yu J., Luo Y., Zhao L., Jiang X., Shi F. Sphingomyelin synthase 2 promotes an aggressive breast cancer phenotype by disrupting the homoeostasis of ceramide and sphingomyelin. *Cell Death Dis*. 2019;10(3):157. doi 10.1038/s41419-019-1303-0

Zhou W., Wahl D.R. Metabolic abnormalities in glioblastoma and metabolic strategies to overcome treatment resistance. *Cancers* (*Basel*). 2019;11(9):1231. doi 10.3390/cancers11091231

Zolotareva O., Saik O.V., Königs C., Bragina E.Y., Goncharova I.A., Freidin M.B., Dosenko V.E., Ivanisenko V.A., Hofestädt R. Comorbidity of asthma and hypertension may be mediated by shared genetic dysregulation and drug side effects. *Sci. Rep*. 2019;9(1):16302. doi 10.1038/s41598-019-52762-w

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 20.09.2024. После доработки 30.10.2024. Принята к публикации 02.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):897-903

doi 10.18699/vjgb-24-97

Программный модуль для оценки метаболического потенциала мутантных штаммов бактерии *Corynebacterium glutamicum*

Ф.В. Казанцев [1, 2, 3], М.Ф. Трофимова[2], Т.М. Хлебодарова[1, 2], Ю.Г. Матушкин [1, 2, 3], С.А. Лашин [1, 2, 3]

[1] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
kazfdr@bionet.nsc.ru

Аннотация. Технологии производства различных соединений с применением микроорганизмов приобретают все большую популярность в промышленном производстве. Создание современных высокопродуктивных штаммов, метаболизм которых ориентирован на синтез конкретного целевого продукта, невозможно без комплексной направленной модификации генома с применением методов математического и компьютерного моделирования. Одним из видов бактерий, активно используемых в биотехнологическом производстве, является *Corynebacterium glutamicum*. Для него существует уже пять полногеномных потоковых моделей, которые можно использовать для задач исследования и оптимизации метаболизма. В работе представлен программный модуль развиваемого в Институте цитологии и генетики СО РАН инструмента FluxMicrobiotech, в рамках которого реализована серия вычислительных протоколов, предназначенных для массового компьютерного анализа потоковых моделей *C. glutamicum* на высокопроизводительных вычислительных компьютерах. Программный модуль реализован на языке Python с применением библиотек Pandas, cobraPy и Escher и настроен на работу по принципу «файл на вход/файл на выход». Модель, условия среды и ограничения модели задаются как отдельные текстовые табличные файлы, что позволяет заготовить серию файлов для каждого из разделов, создавая базы доступных сценариев испытаний для вариаций модели. Или, наоборот, позволяет испытывать одну модель в серии разных условий культивирования. Настроены инструменты постобработки данных моделирования, обеспечивающие визуализацию сводных диаграмм и метаболических карт.
Ключевые слова: потоковые модели; метаболизм бактерии; оптимизация метаболизма; рациональная метаболическая инженерия.

Для цитирования: Казанцев Ф.В., Трофимова М.Ф., Хлебодарова Т.М., Матушкин Ю.Г., Лашин С.А. Программный модуль для оценки метаболического потенциала мутантных штаммов бактерии *Corynebacterium glutamicum*. *Вавиловский журнал генетики и селекции*. 2024;28(8):897-903. doi 10.18699/vjgb-24-97

Финансирование. Работа поддержана проектом Курчатовского геномного центра ИЦиГ СО РАН № 075-15-2019-1662.

A software module to assess the metabolic potential of mutant strains of the bacterium *Corynebacterium glutamicum*

F.V. Kazantsev [1, 2, 3], M.F. Trofimova[2], T.M. Khlebodarova[1, 2], Yu.G. Matushkin [1, 2, 3], S.A. Lashin [1, 2, 3]

[1] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
kazfdr@bionet.nsc.ru

Abstract. Technologies for the production of a range of compounds using microorganisms are becoming increasingly popular in industry. The creation of highly productive strains whose metabolism is aimed to the synthesis of a specific desired product is impossible without complex directed modifications of the genome using mathematical and computer modeling methods. One of the bacterial species actively used in biotechnological production is *Corynebacterium glutamicum*. There are already 5 whole-genome flux balance models for it, which can be used for metabolism research and optimization tasks. The paper presents fluxMicrobiotech, a software module developed at the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, which implements a series of computational protocols designed for high-performance computer analysis of *C. glutamicum* whole-genome flux balance models. The tool is based on libraries from the opencobra community (https://opencobra.github.io) within the Python programming language (https://www.python.org), using the Pandas (https://pandas.pydata.org) and Escher (https://escher.readthedocs.io) libraries . It is configured to operate on a 'file-in/file-out' basis. The model,

© Казанцев Ф.В., Трофимова М.Ф., Хлебодарова Т.М., Матушкин Ю.Г., Лашин С.А., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

environmental conditions, and model constraints are specified as separate text table files, which allows one to prepare a series of files for each section, creating databases of available test scenarios for variations of the model. Or vice versa, allowing a single model to be tested under a series of different cultivation conditions. Post-processing tools for modeling data are set up, providing visualization of summary charts and metabolic maps.
Key words: flux models; bacterial metabolism; metabolic optimization; rational metabolic engineering.

For citation: Kazantsev F.V., Trofimova M.F., Khlebodarova T.M., Matushkin Yu.G., Lashin S.A. A software module to assess the metabolic potential of mutant strains of the bacterium *Corynebacterium glutamicum*. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):897-903. doi 10.18699/vjgb-24-97

Введение

Технологии производства различных соединений с применением микроорганизмов приобретают все большую популярность в промышленном производстве. Создание современных высокопродуктивных штаммов микроорганизмов, метаболизм которых ориентирован на синтез конкретного целевого продукта, невозможно без комплексной направленной модификации генома. К настоящему времени разработан широкий спектр методов рациональной и системной метаболической инженерии для увеличения продукции целевых веществ (Sheremetieva et al., 2023; Шереметьева и др., 2024), использование которых совместно с подходами компьютерного моделирования позволит проводить более точные оценки влияния изменений генома на динамику системы и выход конечного продукта (Ananda et al., 2024). Реализация в рамках компьютерных платформ методов потокового математического моделирования молекулярно-генетических и метаболических систем (Mendoza et al., 2019; Mao et al., 2023) и создание полногеномных потоковых моделей позволяют *in silico* предсказывать генетические модификации, необходимые для увеличения скорости роста культуры, выхода целевого продукта при росте культуры в оптимальных условиях и в зависимости от различных субстратов (Gu et al., 2019; Mao et al., 2023).

Один из видов бактерий, активно используемых в биотехнологическом производстве, – *Corynebacterium glutamicum*. С момента открытия в 1956 г. (Kinoshita et al., 1957) и до сих пор основной сферой применения этого вида бактерий остается производство аминокислот и их производных (Tsuge, Matsuzawa, 2021), которое сегодня является вторым по экономической значимости процессом в промышленной биотехнологии (Barcelos et al., 2018). *C. glutamicum* – это непатогенные, GC-богатые грамположительные почвенные бактерии. Они не образуют спор, быстро растут, не требуют специальных условий для роста, не секретируют протеазы, имеют относительно стабильный геном и устойчивы к высоким концентрациям потенциально токсичных веществ, что делает этот микроорганизм идеальной платформой для разработки на его основе промышленно значимых штаммов-продуцентов (Wendisch et al., 2016).

Основные подходы по модификации генома биотехнологически значимых штаммов бактерий включают: 1) нокауты (выключение) генов; 2) встройки дополнительных генов, ведущих к созданию новых метаболических цепочек реакций; 3) введение мутаций как в регуляторные районы генов, так и в структуру генов с целью уменьшения/увеличения экспрессии генов и активности их продуктов соответственно; 4) другие современные методы редактирования генома *C. glutamicum*, без которых невозможно осуществление большого числа направленных модификаций, необходимых для реализации подходов рациональной и системной метаболической инженерии (Sheremetieva et al., 2023; Шереметьева и др., 2024). Эффективное планирование, осуществление и контроль проведения таких модификаций затруднительны без использования методов математического и компьютерного моделирования.

Настоящая работа посвящена разработке программного модуля в рамках развиваемого в Институте цитологии и генетики СО РАН инструментария FluxMicrobiotech для оценки метаболического потенциала бактерии методами потокового моделирования, включающего серию вычислительных протоколов, настроенных на массовый компьютерный анализ метаболизма целевых штаммов бактерий при культивировании на разных питательных средах и в различных условиях среды (аэробных/анаэробных).

Материалы и методы

Разработанные вычислительные протоколы основаны на открытой библиотеке методов потокового моделирования opencobra (https://opencobra.github.io) в рамках языка программирования Python (https://www.python.org/). Протоколы оформлены как «блокноты» в среде программирования Jupyter (https://jupyter.org/). Такая организация позволяет чередовать вычислительные блоки с этапами анализа результатов. Подход организации вычислений через «блокноты» стал привычным инструментом в методологии анализа больших данных, подразумевающим создание вычислительных конвейеров и регулярную подстройку их под меняющиеся условия задачи. Контроль корректности применения достигается развитым инструментарием по оформлению комментариев к блокам кода этапов вычислений. Для решения оптимизационных задач используются библиотека cobraPy (https://opencobra.github.io/cobrapy/) и библиотека Pandas (https://pandas.pydata.org/). Для первичной визуализации генных сетей – приложение yEd Graph Editor (https://yworks.com/products/yed). Создание метаболических карт и отображение решений на них при моделировании осуществляются в инструментарии Escher (https://escher.github.io/). Разработанные протоколы поддерживают методы высокопроизводительных вычислений и требуют памяти для хранения результатов. Таким образом, рекомендуется проводить работы на высокопроизводительных вычислительных компьютерах.

Методы потокового моделирования, или Flux Balance Analysis (FBA в англоязычной литературе), используемые в данной работе, относятся к разделу задач линейного программирования. Именно для решения задач исследова-

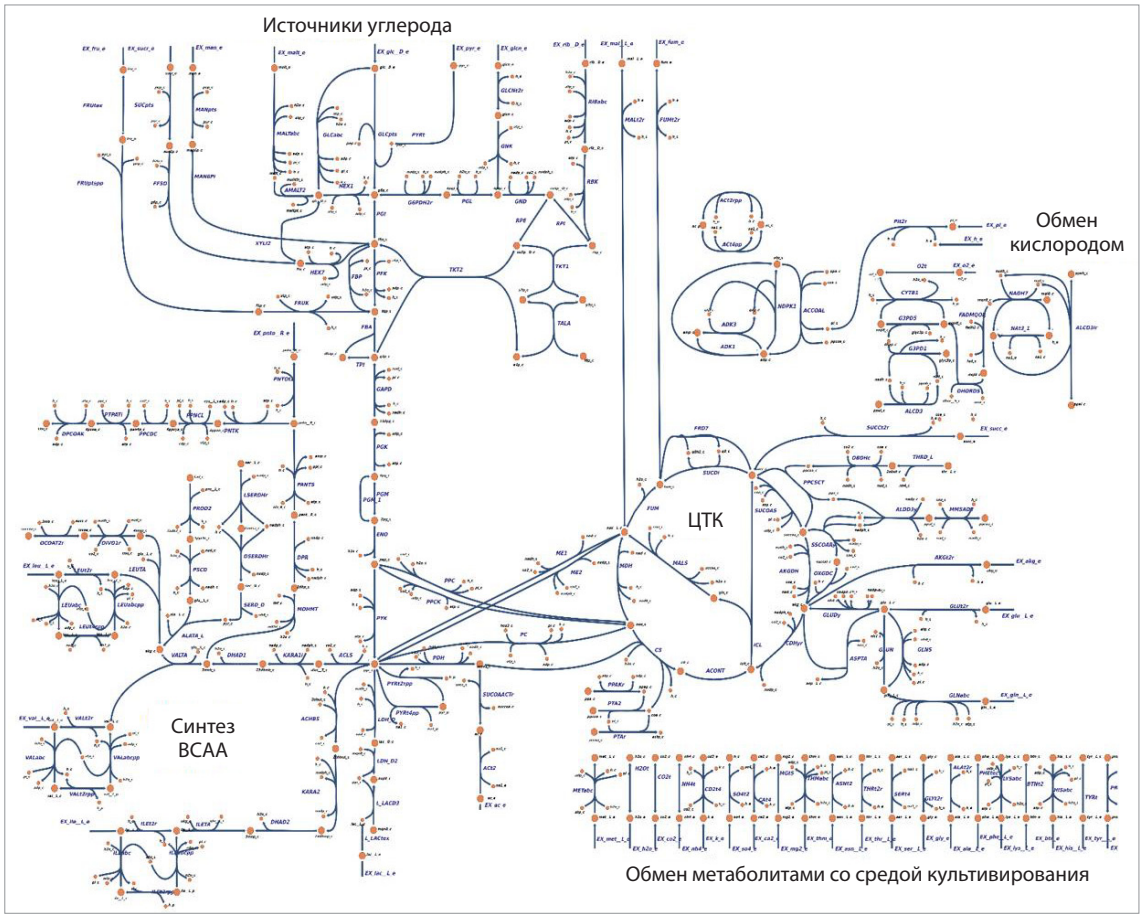


Рис. 1. Метаболическая карта, ориентированная на отображение путей синтеза аминокислот с разветвленной боковой цепью (BCAA).

Визуализация выполнена в инструменте Escher как расширенная сеть из модели iCGB21FR.

ния метаболизма в рамках сообщества opencobra (https://opencobra.github.io) развивается серия вычислительных библиотек методов FBA. Основа этой методологии – представление метаболического пути в виде графа, заданного матрицей смежности, строки которой соответствуют метаболитам, а столбцы – метаболическим реакциям и процессам. Элементы матрицы – стехиометрические коэффициенты, задающие пропорцию метаболита и его роль в выбранной реакции (реагент или продукт реакции). Подобные матрицы можно строить вручную, скрупулезно описывая целевые метаболические пути, или автоматически, генерируя матрицу по геномной информации. На базе хорошо аннотированного секвенированного генома бактерии с использованием различных биоинформатических инструментов можно выявить потенциальные метаболические пути и возможности бактерии к синтезу целевых метаболитов. Именно эта информация обрабатывается программными инструментами по генерации полногеномных потоковых моделей (GSM, genome-scale metabolic models в англоязычной литературе) (Machado et al., 2018; Kulyashov et al., 2023).

Построенная вышеупомянутым образом потоковая модель является отправной точкой в задачах оценки метаболизма бактерии и может содержать несколько тысяч реакций, описывающих полный набор функциональности, доступный в геноме. Существует база данных BiGG Models (http://bigg.ucsd.edu/), которая позиционируется как центральная точка хранения и повторного использования моделей. Данный ресурс содержит крупнейшую коллекцию полногеномных математических моделей, разработанных для разных организмов, и в дополнение развивается и как база эталонных биохимических реакций для такого рода моделей. В рамках BiGG параллельно развивается инструмент визуализации метаболических сетей Escher (King et al., 2015), что позволяет повторно применять одни и те же метаболические карты для моделей разных организмов. В базе BiGG содержатся 108 опубликованных и вручную проверенных полногеномных моделей метаболизма для 40 разных организмов (Norsigian et al., 2019).

Таким образом, связка из данных генома, инструментов построения и аннотации полногеномных потоковых моделей, интеграция их в рамках подхода BiGG дают основу для высокопроизводительного компьютерного анализа метаболизма бактерий. Несмотря на то что модель является полногеномной, при отображении метаболической карты в виде графа используют некоторое подмножество реакций – метаболические пути ключевых метаболитов (рис. 1), подразумевая, что не включенные в визуализацию пути также участвуют в анализе.

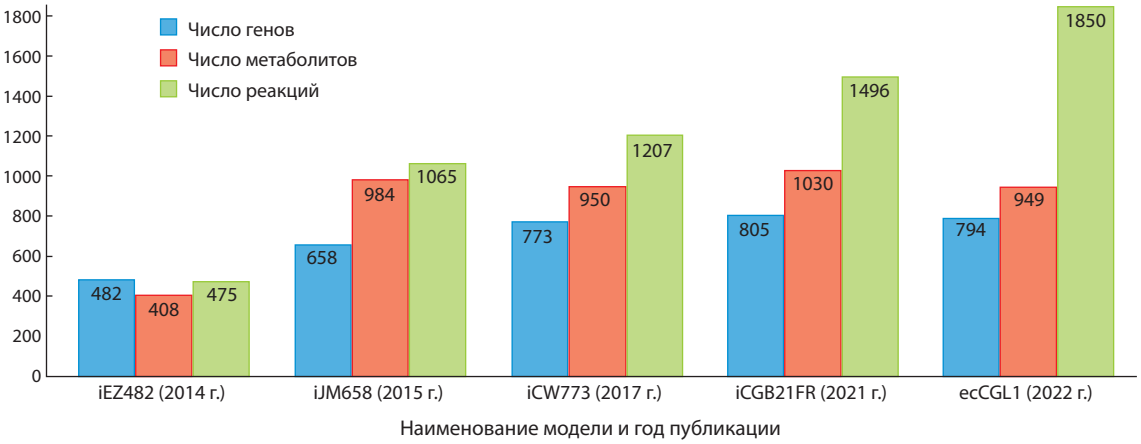


Рис. 2. Математические модели, описывающие метаболизм бактерии *C. glutamicum*, и их основные метрики.

Результаты

Потоковая модель *Corynebacterium glutamicum*

К настоящему времени создано и опубликовано несколько математических моделей, описывающих метаболизм бактерии *C. glutamicum*: iEZ482, iCW773, iCGB21FR, ecCGL1, iJM658 (Kjeldsen, Nielsen, 2009; Zelle et al., 2015; Mei et al., 2016; Zhang et al., 2017; Feierabend et al., 2021; Niu et al., 2022). Эти модели построены на основе полногеномных данных и были верифицированы на экспериментальных данных по росту бактерий, способности синтеза аминокислот на разных источниках углерода и при разных условиях среды культивирования. Модели использовались для анализа продукции глутамата (Mei et al., 2016; Feierabend et al., 2021), изолейцина (Zhang et al., 2017) и лизина (Kjeldsen, Nielsen, 2009; Zhang et al., 2017; Niu et al., 2022).

Модель iEZ482 была представлена в 2015 г. и описывает метаболизм штамма ATCC 13032. Она содержит 475 метаболических реакций и 408 метаболитов. Валидация модели проводилась авторами на основе экспериментальных данных по способности секреции 20 аминокислот. Модель iCW773, опубликованная в 2017 г., содержит уже 1207 реакций и 950 метаболитов. На базе iCW773 в 2022 г. была опубликована модель ecCGL1. Она представляет собой математическое описание метаболизма бактерии *C. glutamicum* штамма ATCC 13032 с ферментативными ограничениями, в котором задаются не только метаболиты и реакции, но также отдельно выносятся ограничения на максимальную концентрацию ферментов в бактерии. Модель iJM658 построена для штамма S9114. Опубликована в 2016 г. и содержит 658 генов, 984 метаболита и 1065 реакций. Дальнейшее развитие полногеномного моделирования для *C. glutamicum* ATCC 13032 привело к созданию модели iCGB21FR, вышедшей в 2021 г. Модель содержит 1496 реакций, 1030 метаболитов, 805 генов и три компартмента: внеклеточное пространство, цитозоль и периплазму. Валидация модели была осуществлена авторами на метаболизме L-глутамата, который, в свою очередь, является предшественником синтеза ряда аминокислот. Сводные характеристики найденных моделей представлены на рис. 2.

Именно модель iCGB21FR была выбрана нами в качестве базовой для настройки вычислительных протоколов, построения карт метаболизма и инструментов постобработки данных, так как она наиболее полно и актуально на данный момент описывает метаболизм бактерии *C. glutamicum*. Кроме того, она может служить эталоном аннотации моделей, поскольку закрывает большую часть рекомендательных пунктов в стандарте оформления моделей системной биологии, включая ссылки на существующие банки данных и онтологии. Модель iCGB21FR находится в свободном доступе в базе данных BioModels (https://www.ebi.ac.uk/biomodels, идентификатор MODEL2102050001). Демонстрирует возможность «роста биомассы» бактерии на разных источниках углерода при аэробных и анаэробных условиях на трех различных средах культивирования: минимальная среда M9, минимальная среда CGXII и полная среда лизогенного бульона (LB). Данные условия отличаются количеством и качеством (наличие дополнительных источников углерода или аминокислот) метаболитов, которые модель может потреблять из среды культивирования для переработки в продукты метаболизма.

Вычислительные протоколы

Разработанный программный модуль содержит в себе серию основных вычислительных сценариев, поток данных в которых схематично представлен на рис. 3. Это подготовленный блокнот Jupyter lab, в котором выставлены параметры вычислений.

Стартовые условия у всех протоколов одни. Необходимо задать:

1) потоковую модель (*.json файл), которая описывает базовую структуру и ограничения модели. Эту модель можно получить из баз данных BiGG или создать с помощью программного инструментария cobraPy;
2) параметры среды культивирования как табличный текстовый файл (*.csv);
3) дополнительные ограничения на потоки модели как табличный текстовый файл (*.csv).

Далее, в зависимости от решаемой задачи, настраиваются параметры вычислений. Jupyter lab блокнот как вычислительный протокол позволяет в случае необходимости

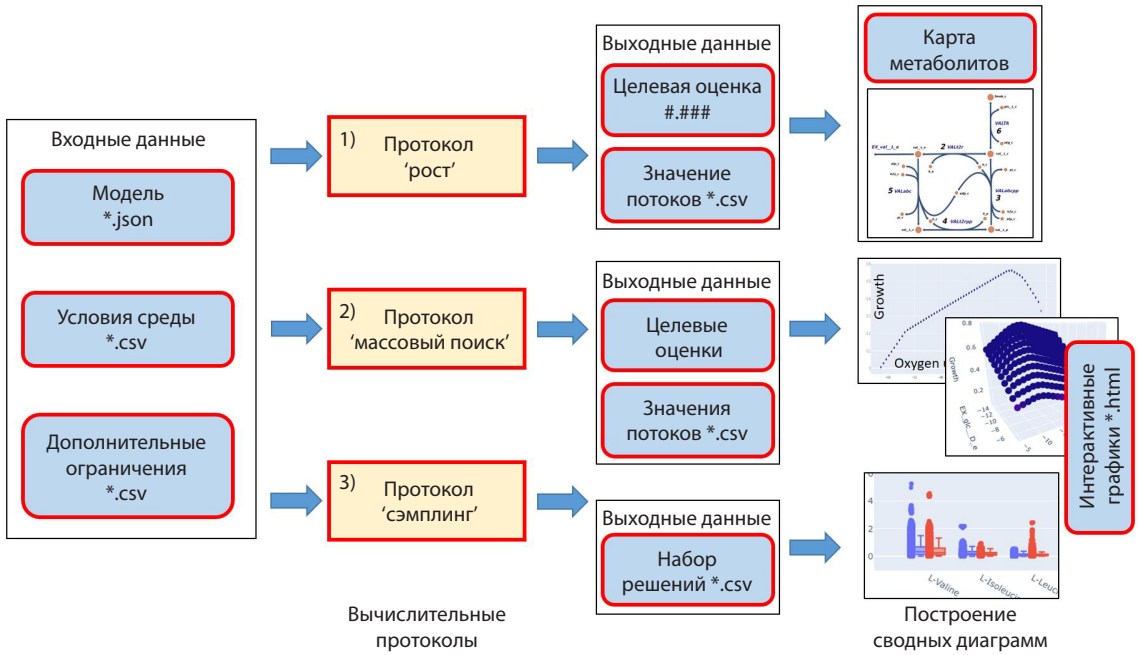


Рис. 3. Схема движения данных вычислительных протоколов.

оперативно модифицировать каждый блок вычислений. В итоге вычислительный протокол задается фактически через набор файлов: модель, среда культивирования, дополнительные ограничения. Это обеспечивает возможность заготовить серию файлов для каждого из разделов, создавая базы доступных сценариев испытаний для вариаций модели или, наоборот, испытывая одну модель в серии разных условий культивирования.

Результат работы протокола – вектор результирующих скоростей по всей структуре модели (или набор таких векторов в виде прямоугольной матрицы). Для задач постобработки данных настроен инструментарий по их отображению как в виде результирующих диаграмм, так и в виде визуализации потоков на метаболической карте (см. рис. 1). Экспорт в виде серии интерактивных метаболических карт сделан с помощью инструментария Escher (https://escher.readthedocs.io).

Сценарий оценки выхода биомассы

Большую роль в биотехнологическом производстве играет среда культивирования бактерий. Среды могут быть как минимального биохимического состава, так и богатые на аминокислоты, благодаря чему бактерия может не тратить внутренние ресурсы на синтез аминокислот и других метаболитов, а потреблять их из среды. Для оценки параметров метаболизма штаммов с помощью моделирования необходимо как можно более точно задавать условия культивирования.

Первой проверкой адекватности модели является ее способность предсказать рост биомассы на заданных субстратах в соответствии с экспериментальными данными. Как правило, этот показатель несложно исследовать экспериментально: имеется много данных о скоростях роста штаммов и скоростях потребления субстрата или об отсутствии роста на выбранных источниках углерода. Сравнение этих показателей – ключевой этап при базовой оценке модели на корректность. В частности, модель iCGB21FR тестировалась на полноту на нескольких средах на способность синтеза аминокислот и в аэробных, и в анаэробных условиях. Меняя условия среды культивирования, можно оценить лимитирующие субстраты в реакции наработки биомассы. Этот сценарий подходит и для оценки способности достичь выбранной/выбранных реакций при заданных условиях среды культивирования. То есть можно проверить достаточность метаболитов в среде для потенциального прохождения целевых метаболических реакций.

Сценарий оценки оптимизации пространства допустимых решений

Предыдущий сценарий проверял осуществление целенаправленных путей от точки поглощения субстратов до конкретных метаболических реакций. Следующим аспектом исследования таких моделей является оценка способности бактерии работать в заданных режимах, т. е. способность в принципе синтезировать серию метаболитов на заданном субстрате при условии наложенных ограничений. В этой работе помогают методы оценки допустимого пространства решений – сэмплинг. Решение в методе «сэмплинг» – это вектор скоростей потоков через все метаболические реакции, который удовлетворяет условиям баланса и наложенным пользователем ограничениям на границы выбранных скоростей реакций. В отличие от метода анализа баланса потоков, сэмплинг генерирует множество допустимых возможных решений системы реакций в модели без указания целевых характеристик, что делает этот метод удобным для оценки путей оптимизации реакций (Herrmann et al., 2019).

Для более точного представления о пространстве возможных решений требуется генерировать достаточно большое количество выборок, размерами в десятки/сотни

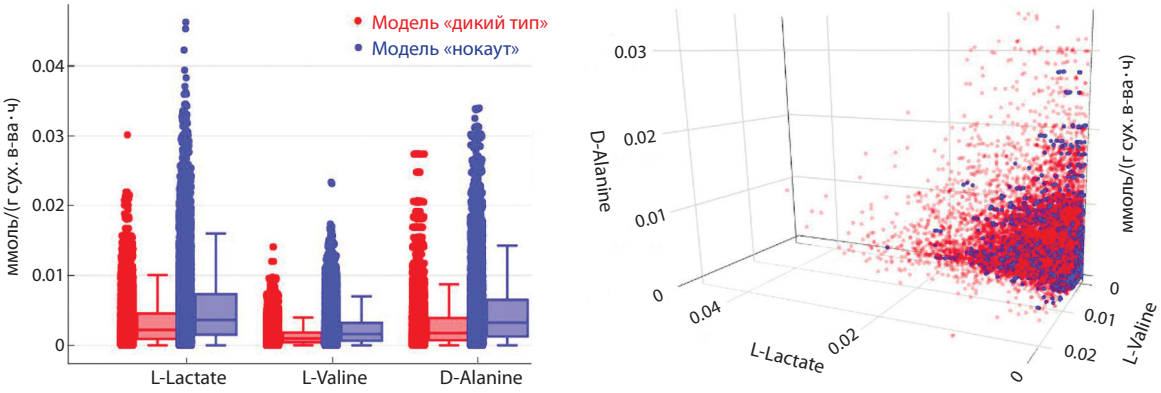


Рис. 4. Результат сравнения двух вариантов модели iCGB21FR: базовой («дикий тип») и модели «нокаут», где введен нокаут гена периплазматической АТФ-синтазы (*atpB*).

Слева – отображение значений скорости экскреции лактата, валина и аланина; справа – отображение этих же значений в одном трехмерном пространстве (проекции 10 тыс. точек решений на оси L-valine (валин), D-alanine (аланин) и L-lactate (лактат)). Скорости потоков реакций в модели выражаются в ммоль на грамм сухого вещества биомассы в час (ммоль/(г сух. в-ва·ч)).

тысяч точек пространства решений (принимая во внимание, что каждая точка в этом пространстве описывается сотнями или даже тысячами численных значений скоростей потоков). В результате можно получить множество точек в пространстве решений, которые могут указывать на наиболее часто встречающиеся решения при заданных условиях. Метод равномерно выбирает точки, покрывая пространство решений. Отображая точки на координаты целевых скоростей, можно получить предполагаемое распределение значений. Таким образом, мы получаем не конкретное распределение потоков на метаболической карте, а серию решений (серию результирующих потоков/облако точек). Каждую точку этой серии решений можно отобразить на оси скоростей выбранных реакций метаболической сети. Этот подход позволяет сравнивать распределение потоков как нескольких моделей в одних условиях, так и одной модели в разных условиях/ограничениях (рис. 4).

В частности, в рамках серии вычислительных экспериментов по влиянию нокаутов генов на экскрецию метаболитов был обнаружен ген *atpB*, продукт синтеза которого участвует в реакции фосфорилирования АТФ (см. рис. 4). Наименование гена в терминах базы данных KEGG – cgb:cg1362. Нокаут *atpB* обеспечивает потенциально большую экскрецию L-валина. Косвенным подтверждением важности этого гена служит работа, в которой показано (Jensen et al., 1993), что мутации в опероне АТФ-синтаз в *Escherichia coli* могут приводить к более высокой скорости роста на глюкозе.

Запуск расчетов на 10 тыс. решений/точек выдает около 200 Мб данных за один запуск. Вычисления и постобработка таких данных должны осуществляться на высокопроизводительных вычислительных машинах.

Заключение

В крупнейшей базе полногеномных моделей BiGG (http://bigg.ucsd.edu/models) выложено 108 моделей для 40 разных организмов. При этом мы нашли пять полногеномных математических моделей по *C. glutamicum*, что говорит о большом интересе к объекту. Сама методология находится на этапе становления, и требуется под каждый новый объект настраивать инструменты вручную, что дает большое пространство для развития математического и компьютерного аппарата моделирования в рамках сообщества системных биологов/рациональных метаболических инженеров. Сейчас идут исследования по внедрению в такие модели транскриптомных и протеомных данных, что ведет к более высокой предсказательной способности по сравнению с более простыми потоковыми моделями.

Несмотря на то что *C. glutamicum* изучают с 1956 г. (Kinoshita et al., 1957), собрать открытую информацию по штаммам бактерии – отдельная задача. Существует много штаммов, данные о которых являются коммерческой тайной и не могут находиться в открытом доступе. Развитие разработанных вычислительных конвейеров в перспективе позволит применить их для анализа метаболизма других штаммов.

Предложенный нами программный модуль в виде серии вычислительных протоколов настроен для массового компьютерного анализа моделей целевых штаммов *C. glutamicum* по культивированию на разных питательных средах и условиях среды (аэробных/анаэробных). Протоколы настроены на запуск по принципу «файл на вход/файл на выход», в рамках которого модель, условия среды и ограничения модели задаются как отдельные файлы. Настроены методы визуализации результатов моделирования, в частности по отображению данных на серии подготовленных пользователем метаболических карт. Особенности выполнения алгоритмов требуют применения высокопроизводительных вычислительных компьютеров и доступа к большим объемам хранилища данных. Модуль является частью развиваемого в ИЦиГ СО РАН инструмента FluxMicrobiotech.

Список литературы / References

Шереметьева М.Е., Хлебодарова Т.М., Дербиков Д.Д., Розанцева В.В., Колчанов Н.А., Яненко А.С. Системная метаболическая инженерия *Corynebacterium glutamicum* для продукции L-валина. *Биотехнология*. 2024;40(3):3-23. doi 10.56304/S0234275824030025

[Sheremetieva M.E., Khlebodarova T.M., Derbikov D.D., Rozantseva V.V., Kolchanov N.A., Yanenko A.S. Systems metabolic engineering of *Corynebacterium glutamicum* to create a producer of L-valine. *Biotekhnologiya = Biotechnology*. 2024;40(3):3-23. doi 10.56304/S0234275824030025 (in Russian)]

Ananda R., Daud K.M., Zainudin S. A review of advances in integrating gene regulatory networks and metabolic networks for designing strain optimization. *J. King Saud Univ. Comput. Inf. Sci.* 2024; 36(6):102120. doi 10.1016/j.jksuci.2024.102120

Barcelos M.C.S., Lupki F.B., Campolina G.A., Nelson D.L., Molina G. The colors of biotechnology: general overview and developments of white, green and blue areas. *FEMS Microbiol. Lett.* 2018;365(21): fny239. doi 10.1093/femsle/fny239

Feierabend M., Renz A., Zelle E., Nöh K., Wiechert W., Dräger A. High-quality genome-scale reconstruction of *Corynebacterium glutamicum* ATCC 13032. *Front. Microbiol.* 2021;12:750206. doi 10.3389/fmicb.2021.750206

Gu C., Kim G.B., Kim W.J., Kim H.U., Lee S.Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* 2019; 20(1):121. doi 10.1186/s13059-019-1730-3

Herrmann H.A., Dyson B.C., Vass L., Johnson G.N., Schwartz J.-M. Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *NPJ Syst. Biol. Appl.* 2019;5(1):32. doi 10.1038/s41540-019-0109-0

Jensen P.R., Michelsen O., Westerhoff H.V. Control analysis of the dependence of *Escherichia coli* physiology on the H^+-ATPase. *Proc. Natl. Acad. Sci. USA.* 1993;90(17):8068-8072. doi 10.1073/pnas.90.17.8068

King Z.A., Dräger A., Ebrahim A., Sonnenschein N., Lewis N.E., Palsson B.O. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol.* 2015;11(8):e1004321. doi 10.1371/journal.pcbi.1004321

Kinoshita S., Udaka S., Shimono M. Studies on the amino acid fermentation. *J. Gen. Appl. Microbiol.* 1957;3(3):193-205. doi 10.2323/jgam.3.193

Kjeldsen K.R., Nielsen J. In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol. Bioeng.* 2009;102(2):583-597. doi 10.1002/bit.22067

Kulyashov M.A., Kolmykov S.K., Khlebodarova T.M., Akberdin I.R. State-of the-art constraint-based modeling of microbial metabolism: from basics to context-specific models with a focus on methanotrophs. *Microorganisms*. 2023;11(12):2987. doi 10.3390/microorganisms11122987

Machado D., Andrejev S., Tramontano M., Patil K.R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* 2018;46(15):7542-7553. doi 10.1093/nar/gky537

Mao Z., Yuan Q., Li H., Zhang Y., Huang Y., Yang C., Wang R., Yang Y., Wu Y., Yang S., Liao X., Ma H. CAVE: a cloud-based platform for analysis and visualization of metabolic pathways. *Nucleic Acids Res.* 2023;51(W1):W70-W77. doi 10.1093/nar/gkad360

Mei J., Xu N., Ye C., Liu L., Wu J. Reconstruction and analysis of a genome-scale metabolic network of *Corynebacterium glutamicum* S9114. *Gene*. 2016;575(2):615-622. doi 10.1016/j.gene.2015.09.038

Mendoza S.N., Olivier B.G., Molenaar D., Teusink B. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.* 2019;20(1):158. doi 10.1186/s13059-019-1769-1

Niu J., Mao Z., Mao Y., Wu K., Shi Z., Yuan Q., Cai J., Ma H. Construction and analysis of an enzyme-constrained metabolic model of *Corynebacterium glutamicum*. *Biomolecules*. 2022;12(10):1499. doi 10.3390/biom12101499

Norsigian C.J., Pusarla N., McConn J.L., Yurkovich J.T., Dräger A., Palsson B.O., King Z. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* 2019;48(D1):D402-D406. doi 10.1093/nar/gkz1054

Sheremetieva M.E., Anufriev K.E., Khlebodarova T.M., Kolchanov N.A., Yanenko A.S. Rational metabolic engineering of *Corynebacterium glutamicum* to create a producer of L-valine. *Vavilov J. Genet. Breed.* 2023;26(8):743-757. doi 10.18699/VJGB-22-90

Tsuge Y., Matsuzawa H. Recent progress in production of amino acid-derived chemicals using *Corynebacterium glutamicum*. *World J. Microbiol. Biotechnol.* 2021;37(3):49. doi 10.1007/s11274-021-03007-4

Wendisch V.F., Jorge J.M.P., Pérez-García F., Sgobba E. Updates on industrial production of amino acids using *Corynebacterium glutamicum*. *World J. Microbiol. Biotechnol.* 2016;32(6):105. doi 10.1007/s11274-016-2060-1

Zelle E., Nööh K., Wiechert W. Growth and production capabilities of *Corynebacterium glutamicum*: interrogating a genome-scale metabolic network model. In: Burkovski A. (Ed.) *Corynebacterium glutamicum*: From Systems Biology to Biotechnological Applications. Caister Acad. Press, 2015;39-56. doi 10.21775/9781910190050.04

Zhang Yu, Cai J., Shang X., Wang B., Liu S., Chai X., Tan T., Zhang Yun, Wen T. A new genome-scale metabolic model of *Corynebacterium glutamicum* and its application. *Biotechnol. Biofuels*. 2017;10(1):169. doi 10.1186/s13068-017-0856-3

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 17.09.2024. После доработки 20.11.2024. Принята к публикации 21.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):904-917

DOI 10.18699/vjgb-24-98

Реконструкция и компьютерный анализ генной сети, отражающей роль микроРНК в регуляции ответа пшеницы на засуху

М.А. Клещев [1, 4], А.В. Мальцева [1, 3, 4], Е.А. Антропова [1, 4], П.С. Деменков [1, 2, 3, 4], Т.В. Иванисенко [1, 2, 3, 4], Ю.Л. Орлов [5, 6], Х. Чао [7], М. Чэнь [7], Н.А. Колчанов [1, 2, 3], В.А. Иванисенко [1, 2, 3, 4]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[4] Исследовательский центр в сфере искусственного интеллекта Новосибирского национального исследовательского государственного университета, Новосибирск, Россия
[5] Аграрно-технологический институт Российского университета дружбы народов им. Патриса Лумумбы, Москва, Россия
[6] Центр цифровой медицины, Первый Московский государственный медицинский университет им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия
[7] Отдел биоинформатики, Колледж естественных наук, Чжэцзянский университет, Ханчжоу, Китай

salix@bionet.nsc.ru

Аннотация. Недостаток влаги – критический фактор, ограничивающий продуктивность мягкой пшеницы (*Triticum aestivum* L.), одной из ключевых сельскохозяйственных культур. Адаптация пшеницы к водному дефициту обеспечивается комплексными молекулярно-генетическими механизмами, включающими согласованную работу множества генов, регулируемых транскрипционными факторами и сигнальными некодирующими РНК, в частности микроРНК. микроРНК – опосредованная регуляция экспрессии генов – рассматривается как один из основных механизмов устойчивости растений к абиотическим стрессам. Изучение этих сложных молекулярно-генетических механизмов требует применения методов компьютерной системной биологии. Цель данной работы – реконструкция и компьютерный анализ генной сети, связанной с микроРНК-регуляцией адаптации мягкой пшеницы к условиям недостаточного увлажнения. Для достижения этой цели использованы программно-информационная система ANDSystem и специализированная база знаний Smart crop, адаптированная для области генетики и селекции пшеницы. Нами была реконструирована генная сеть ответа пшеницы на водный дефицит, включающая 144 гена, 1017 белков и 21 микроРНК пшеницы. Анализ сети выявил, что микроРНК преимущественно регулируют гены, контролирующие процессы морфогенеза побегов и корней растений, что играет важную роль в морфологических адаптациях к засухе. Ключевыми компонентами генной сети, регулируемыми микроРНК, оказались транскрипционные факторы семейств MYB и WRKY, а также белок теплового шока HSP90 и белок RPM1. Эти белки связаны с сигнальными путями фитогормонов и кальций-зависимыми протеинкиназами, играющими существенную роль в адаптации растений к водному дефициту. Было идентифицировано несколько микроРНК (*MIR7757*, *MIR9653a*, *MIR9671*, *MIR9672b*), ранее не обсуждавшихся в контексте адаптации пшеницы к засухе, которые являются кандидатами для дальнейших экспериментальных исследований, направленных на усиление устойчивости пшеницы к недостатку влаги. Полученные результаты могут быть полезными для создания новых сортов пшеницы с повышенной устойчивостью к водному дефициту, что имеет существенное значение для сельского хозяйства в условиях изменения климата.

Ключевые слова: микроРНК; мягкая пшеница; дефицит влаги; гены; генетическая регуляция; ассоциативные генные сети; биоинформатика растений; база знаний Smart crop; программно-информационная система ANDSystem.

Для цитирования: Клещев М.А., Мальцева А.В., Антропова Е.А., Деменков П.С., Иванисенко Т.В., Орлов Ю.Л., Чао Х., Чэнь М., Колчанов Н.А., Иванисенко В.А. Реконструкция и компьютерный анализ генной сети, отражающей роль микроРНК в регуляции ответа пшеницы на засуху. *Вавиловский журнал генетики и селекции*. 2024;28(8):904-917. DOI 10.18699/vjgb-24-98

Финансирование. Работа МАК, АВМ, ЕАА, ПСД, ТВИ, ЮЛО, НАК и ВАИ поддержана российско-китайским грантом Российского научного фонда № 23-44-00030. Работа HCh и MCh поддержана Национальным фондом естественных наук Китая (№ 32261133526).

Reconstruction and computational analysis of the microRNA regulation gene network in wheat drought response mechanisms

M.A. Kleshchev [1, 4], A.V. Maltseva [1, 3, 4], E.A. Antropova [1, 4], P.S. Demenkov [1, 2, 3, 4], T.V. Ivanisenko [1, 2, 3, 4], Y.L. Orlov [5, 6], H. Chao [7], M. Chen [7], N.A. Kolchanov [1, 2, 3], V.A. Ivanisenko [1, 2, 3, 4]

© Клещев М.А., Мальцева А.В., Антропова Е.А., Деменков П.С., Иванисенко Т.В., Орлов Ю.Л., Чао Х., Чэнь М., Колчанов Н.А., Иванисенко В.А., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
[4] Research Center in the Field of Artificial Intelligence of Novosibirsk State University, Novosibirsk, Russia
[5] Agrarian and Technological Institute, Peoples' Friendship University of Russia named after Patrice Lumumba, Moscow, Russia
[6] Digital Health Center, I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, Russia
[7] Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China
salix@bionet.nsc.ru

Abstract. Drought is a critical factor limiting the productivity of bread wheat (*Triticum aestivum* L.), one of the key agricultural crops. Wheat adaptation to water deficit is ensured by complex molecular genetic mechanisms, including the coordinated work of multiple genes regulated by transcription factors and signaling non-coding RNAs, particularly microRNAs (miRNAs). miRNA-mediated regulation of gene expression is considered one of the main mechanisms of plant resistance to abiotic stresses. Studying these mechanisms necessitates computational systems biology methods. This work aims to reconstruct and analyze the gene network associated with miRNA regulation of wheat adaptation to drought. Using the ANDSystem software and the specialized Smart crop knowledge base adapted for wheat genetics and breeding, we reconstructed a wheat gene network responding to water deficit, comprising 144 genes, 1017 proteins, and 21 wheat miRNAs. Analysis revealed that miRNAs primarily regulate genes controlling the morphogenesis of shoots and roots, crucial for morphological adaptation to drought. The key network components regulated by miRNAs are the MYBa and WRKY41 family transcription factors, heat-shock protein HSP90, and the RPM1 protein. These proteins are associated with phytohormone signaling pathways and calcium-dependent protein kinases significant in plant water deficit adaptation. Several miRNAs (*MIR7757*, *MIR9653a*, *MIR9671* and *MIR9672b*) were identified that had not been previously discussed in wheat drought adaptation. These miRNAs regulate many network nodes and are promising candidates for experimental studies to enhance wheat resistance to water deficiency. The results obtained can find application in breeding for the development of new wheat varieties with increased resistance to water deficit, which is of substantial importance for agriculture in the context of climate change.

Key words: microRNA; bread wheat; drought; genes; genetic regulation; associative gene networks; plant bioinformatics; Smart crop knowledge base; ANDSystem computer tool.

For citation: Kleshchev M.A., Maltseva A.V., Antropova E.A., Demenkov P.S., Ivanisenko T.V., Orlov Y.L., Chao H., Chen M., Kolchanov N.A., Ivanisenko V.A. Reconstruction and computational analysis of the microRNA regulation gene network in wheat drought response mechanisms. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):904-917. DOI 10.18699/vjgb-24-98

Введение

Продуктивность мягкой пшеницы (*Triticum aestivum* L.) – важнейшей сельскохозяйственной культуры – зависит от многих экологических факторов, включая обеспечение микроэлементами, температуру, увлажнение, засоленность почвы. При этом недостаток влаги является важнейшим фактором, ограничивающим продуктивность пшеницы (Пакуль и др., 2018; Jeyasri et al., 2021). Следовательно, изучение физиологических и молекулярно-генетических механизмов адаптации пшеницы к недостатку влаги – актуальная задача, решение которой необходимо для создания новых сортов, устойчивых к засухе (Langridge, Reynolds, 2021), а также совершенствования агротехнологий.

Устойчивость растений к условиям недостаточного увлажнения обеспечивается рядом физиологических и морфологических адаптаций, которые включают в себя усиление апикального роста и ингибирование латерального роста корня, опадение листьев, изменение скорости развития, поддержание осмотического давления тканей, снижение транспирации за счет изменения функционирования устьичного аппарата, активацию антиоксидантной защиты клеток. Функционирование данных физиологических механизмов обеспечивается координированной работой целого ряда генов. Показано, что недостаток влаги вызывает изменения в экспрессии генов, активируемых абсцизовой кислотой, генов, кодирующих глутатион-S-трансферазу (GST), семейство белков дегидринов (Ferdous et al., 2015).

Восприятие сигнала рецепторами на клеточной стенке и клеточной мембране приводит к активации внутриклеточных сигнальных каскадов, главным образом из-за увеличения уровня активных форм кислорода (АФК) и изменения уровня ионов кальция. Кроме того, важными посредниками, координирующими запуск сигнальных каскадов, являются фитогормоны, такие как абсцизовая кислота (ABA), жасмоновая кислота (JA), салициловая кислота (SA) и этилен (ET). Сигнальные каскады, активирующиеся в ответ на стресс, включают в себя, в частности, сигнальные пути митоген-активируемой протеинкиназы (MAPK) и кальций-зависимой протеинкиназы (CDPK). Киназы и фосфатазы активируют или подавляют ряд транскрипционных факторов, которые, в свою очередь, регулируют активность генов, контролирующих адаптацию к неблагоприятным условиям (Baillo et al., 2019).

В настоящее время известно пять семейств генов, кодирующих транскрипционные факторы, регулирующие процессы адаптации к недостатку влаги: bZIP (главным образом AREB/ABF), DREB (AP2/EREBP), MYB/MYC, NAC и WRKY (Gahlaut et al., 2016). Анализ литературы показывает, что модификация данных транскрипционных факторов методами генной инженерии может усиливать устойчивость растений к неблагоприятным факторам окружающей среды. Например, трансгенные растения пшеницы, содержащие ген *DREB1A* арабидопсиса (*Arabidopsis thaliana*), показали повышенную устойчивость к засухе и солевому стрессу без снижения урожайности

(Pellegrineschi et al., 2004). C.F. Niu с коллегами (2012) получили трансгенные растения пшеницы с повышенной экспрессией генов *TaWRKY2* и *TaWRKY19*. Эти растения демонстрировали улучшенную устойчивость к засухе и оксидативному стрессу.

Помимо транскрипционных факторов, экспрессия генов может регулироваться и сигнальными некодирующими молекулами РНК. Они включают кольцевые молекулы РНК (circRNAs), а также линейные длинные (lncRNAs) РНК (lncRNAs) и микроРНК (Li N. et al., 2022). Данные сигнальные молекулы могут регулировать экспрессию любых генов, вовлеченных в реакцию на стресс, включая транскрипционные факторы, причем экспрессия генов, кодирующих сигнальные РНК, также может изменяться в ответ на стресс, предоставляя дополнительный уровень регуляции.

МикроРНК – одноцепочечные некодирующие молекулы РНК длиной 20–25 нуклеотидов, которые у растений регулируют активность генов, связываясь с матричной РНК гена мишени, что приводит к ее деградации и ингибированию трансляции (Ma, Hu, 2023). Выявлено, что экспрессия микроРНК изменяется у растений в ответ на недостаток влаги, это показано для многих видов растений, включая пшеницу. У видов семейства Triticeae в условиях засухи изменялась экспрессия генов микроРНК *miR159*, *miR1137*, *miR1318*, *miR168* и других, при этом направленность изменений в экспрессии зависела от ткани, стадии развития растения, продолжительности и интенсивности воздействия (Alptekin et al., 2017). В ответ на недостаток влаги в тканях корня пшеницы происходило изменение экспрессии микроРНК *miR1119*, ее мишени – транскрипционного фактора *MYC2*, а также изменение экспрессии целого ряда генов, регулирующих ответ на стресс, повышение содержания абсцизовой кислоты и активности антиоксидантной системы клеток (Shamloo-Dashtpagerdi et al., 2023).

Таким образом, воздействие микроРНК на транскрипционные факторы может приводить к изменению активности целого набора генов. Поэтому микроРНК можно рассматривать как мастер-регуляторы генных сетей, которые формируют вместе с транскрипционными факторами и их генами-мишенями регуляторные модули, в том числе и обеспечивающие адаптацию растений к абиотическому стрессу (Zhang et al., 2022), рост и развитие растений (Liebsch, Palatnik, 2020). Следовательно, воздействие на микроРНК и активность их регуляторных модулей может стать инструментом для генетического манипулирования сельскохозяйственными культурами с целью достижения оптимальных параметров роста и развития (Wang H., Wang H., 2015).

Особое значение в области маркер-ориентированной селекции имеют биоинформатические методы интеграции и анализа больших омиксных данных, включая методы реконструкции генных сетей (Chao et al., 2023). Биоинформатический анализ генных сетей может помочь идентифицировать регуляторные модули, участвующие в обеспечении адаптации растений к неблагоприятным факторам внешней среды, и понять ее молекулярные механизмы.

Ранее была разработана программно-информационная система ANDSystem, предназначенная для реконструкции генных сетей на основе информации, полученной из фактографических баз данных и собранной с помощью автоматического анализа текстов научных публикаций (Ivanisenko V.A. et al., 2015, 2019; Ivanisenko T.V. et al., 2020, 2022). ANDSystem применялась для решения задач в различных областях биологии и биомедицины, включая исследование молекулярно-генетических механизмов развития астмы (Bragina et al., 2014; Saik et al., 2018; Zolotareva et al., 2019), лимфедемы (Saik et al., 2019), туберкулеза (Bragina et al., 2016), гепатита С (Saik et al., 2016), коронавирусной инфекции (Ivanisenko V.A. et al., 2022), болезни Хантингтона (Bragina et al., 2023), глиомы (Rogachev et al., 2021), пост-оперативного делирия (Иванисенко В.А. и др., 2023), гепатоцеллюлярной карциномы (Antropova et al., 2023), исследования протеомного профиля космонавтов (Larina et al., 2015; Pastushkova et al., 2019).

В области биологии растений с использованием ANDSystem проведены реконструкция и анализ регуляторной генной сети функционирования клеточной стенки листьев *A. thaliana* L. при ответе на недостаточное увлажнение (Volyanskaya et al., 2023). На основе ANDSystem создана база знаний SOLANUM TUBEROSUM, содержащая информацию о генетической регуляции метаболических путей картофеля (Иванисенко Т.В. и др., 2018), и проведена приоритизация генов картофеля, вовлеченных в формирование агрономически ценных признаков растений (Деменков и др., 2019). Следует также отметить, что программно-информационная система ANDSystem ранее применялась для реконструкции генных сетей, описывающих микроРНК-регуляцию внешнего пути апоптоза (Khlebodarova et al., 2023).

Цель настоящей работы – провести реконструкцию и анализ генной сети, обеспечивающей регуляцию адаптации пшеницы к условиям недостаточного увлажнения с помощью микроРНК.

Материалы и методы

Поиск информации о генах ответа на засуху. Информация о генах мягкой пшеницы, связь которых с адаптацией растений к условиям засухи экспериментально доказана, извлекалась из полнотекстовых экспериментальных и обзорных статей, индексированных в PubMed (https://pubmed.ncbi.nlm.nih.gov/) по состоянию на сентябрь 2024 г. Поиск производили по ключевым словам “wheat”, “Triticum aestivum”, “drought”, “drought tolerance”, “gene”, “genetic”, “regulation” и их комбинациям.

Кроме того, информация о генах, связанных с ответом на водный дефицит, извлекалась из базы данных генных онтологий AmiGO для термина “response to water deprivation” (идентификатор термина GO:0009414). Дополнительно в список генов ответа на засуху были включены гены, ассоциированные с термином “response to water deficiency” в программно-информационной системе ANDSystem. В результате был сформирован список генов, для которых показано их участие в адаптации пшеницы к водному дефициту. Этот список использовался в качестве входных данных для реконструкции генных сетей.

База знаний Smart crop. В данной работе использована база знаний Smart crop, которая является специализированной версией программно-информационной системы ANDSystem, ориентированной на область генетики и селекции риса и пшеницы. Настройке на предметную область подверглись три ключевых модуля ANDSystem:

Модуль предметно-ориентированной онтологии. Этот модуль содержит расширенные словари, охватывающие различные объекты исследования, такие как гены, белки, метаболиты, некодирующие РНК/микроРНК, биологические процессы, генетические биомаркеры, QTL-полиморфизмы, сорта растений, селекционно значимые качества, фенотипические признаки, болезни, патогены, вредители, маркеры резистентности к средствам защиты растений, молекулярные мишени для химических средств защиты растений, биотические и абиотические факторы, средства защиты растений (гербициды) и другие. Для формирования словарей использованы различные базы данных и онтологии, включая NCBI Gene, ChEBI, MirBase, Gene Ontology, Wheat Ontology, Rice Ontology, Wheat Trait and Phenotype Ontology, The International Herbicide-Resistant Weed Database и другие. Словари дополнены синонимами и вариантами написания имен для повышения эффективности распознавания объектов в текстах.

Модуль экстракции информации из фактографических баз данных. Этот модуль осуществляет автоматизированное извлечение данных из различных источников, включая реляционные базы данных (например, ChEBI), онтологии в форматах OBO и OWL (с использованием инструмента ROBOT), текстовые файлы в табличных форматах (CSV, TSV) и форматы PSI-MI XML 2.5. Созданы специализированные программы-экстракторы для обработки информации из таких баз данных, как NCBI Gene, ToppGene, GrainGenes, IntAc и других.

Модуль text mining с использованием семантико-лингвистических шаблонов. Этот модуль предназначен для извлечения знаний из текстовых источников (научные статьи, патенты) с помощью семантико-лингвистических шаблонов.

Разработка новых и адаптация существующих в системе ANDSystem шаблонов позволили эффективно идентифицировать и извлекать разнообразные типы взаимодействий между объектами. Шаблоны охватывают такие типы взаимодействий, как ассоциации, регуляция экспрессии и активности генов и белков, физические взаимодействия, каталитические реакции, участие в биологических процессах, маркерные связи и другие. Всего было разработано и использовано более 2000 шаблонов, что значительно повысило точность и полноту извлечения информации.

Настройка ANDSystem на область селекции и генетики риса и пшеницы позволила интегрировать данные из разнообразных источников и обеспечить эффективное извлечение и анализ знаний, необходимых для исследований в данной предметной области.

Реконструкция и анализ генных сетей. Реконструкцию и анализ генных сетей производили с помощью Мастера запросов программного модуля ANDVisio (Demenkov et al., 2012), представляющего собой пользовательский интерфейс в системах ANDSystem и Smart crop.

Важные для функционирования реконструированной генной сети белки и гены определяли с использованием показателя “Network Connectivity”, который характеризует число связей данного узла сети с другими ее узлами. Далее гены и белки ранжировались по этому показателю для поиска наиболее значимых узлов в сети. Функциональная аннотация набора генов (анализ перепредставленности терминов генной онтологии и путей KEGG), представленных в сети, была проведена с использованием интернет-ресурса Database for Annotation, Visualization and Integrated Discovery (DAVID версия 2021; https://david.ncifcrf.gov/) с настройками по умолчанию (показателем статистической значимости считался p-value < 0.05 с поправкой Бонферрони).

Результаты и обсуждение

В результате анализа опубликованной литературы (Nagy et al., 2013; Gupta et al., 2014; Liu et al., 2015; Gahlaut et al., 2016; Shojaee et al., 2022) выявлено 130 генов, участвующих в адаптации пшеницы к недостатку влаги. Кроме того, 15 генов были ассоциированы с термином генной онтологии “response to water deprivation” (идентификатор термина GO:0009414). Дополнительно с использованием базы знаний Smart crop программно-информационной системы ANDSystem было обнаружено 59 генов, участвующих в адаптации пшеницы к недостатку влаги. Результирующий список 204 генов, для которых показано их участие в адаптации пшеницы к недостатку увлажнения (гены ответа на засуху), приведен в Приложении 1[1]. Используя список этих генов в качестве входных данных, мы реконструировали ассоциативную генную сеть, куда были добавлены микроРНК, которые, по данным базы знаний Smart crop, напрямую регулируют хотя бы один из компонентов сети. Эта ассоциативная сеть (рис. 1) включала в себя 75 генов, 98 белков и 14 микроРНК пшеницы, а также 695 взаимодействий между компонентами сети. Из них соответствовали типам связей: 594 связи – «ассоциация», 39 – «регуляция экспрессии», 21 – «взаимодействие», 18 – «экспрессия», 12 – «регуляция активности», 7 – «катализ», 2 – «усиление экспрессии», по 1 связи относились к типам «подавление экспрессии» и «коэкспрессия». Список микроРНК и их генов мишеней, входящих в сеть ответа на засуху, установленных по данным Smart crop, приведен в табл. 1.

Мишенями этих микроРНК являются главным образом гены, кодирующие транскрипционные факторы семейств GAMYB (*MybA*, *Ta-GAMyb*, *MYB3R1*), WRKY (*WRKY41*), фактор ответа на ауксин (*ARF22*, *LOC123121554*, *LOC123181091*), MADS-box транскрипционный фактор (*WM30*), транскрипционный фактор семейства SQUAMOSA (*LOC123151797*, *LOC123159884*, *SBP16*).

Транскрипционные факторы GAMYB, имеющие высококонсервативные сайты связывания с *MIR159a* (Millar et al., 2019), участвуют в активации экспрессии гена гидролазы в алейроновом слое семян, опосредованной гиббереллином (Woodger et al., 2003). В вегетативных

[1] Приложения 1–4 см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx32.xlsx

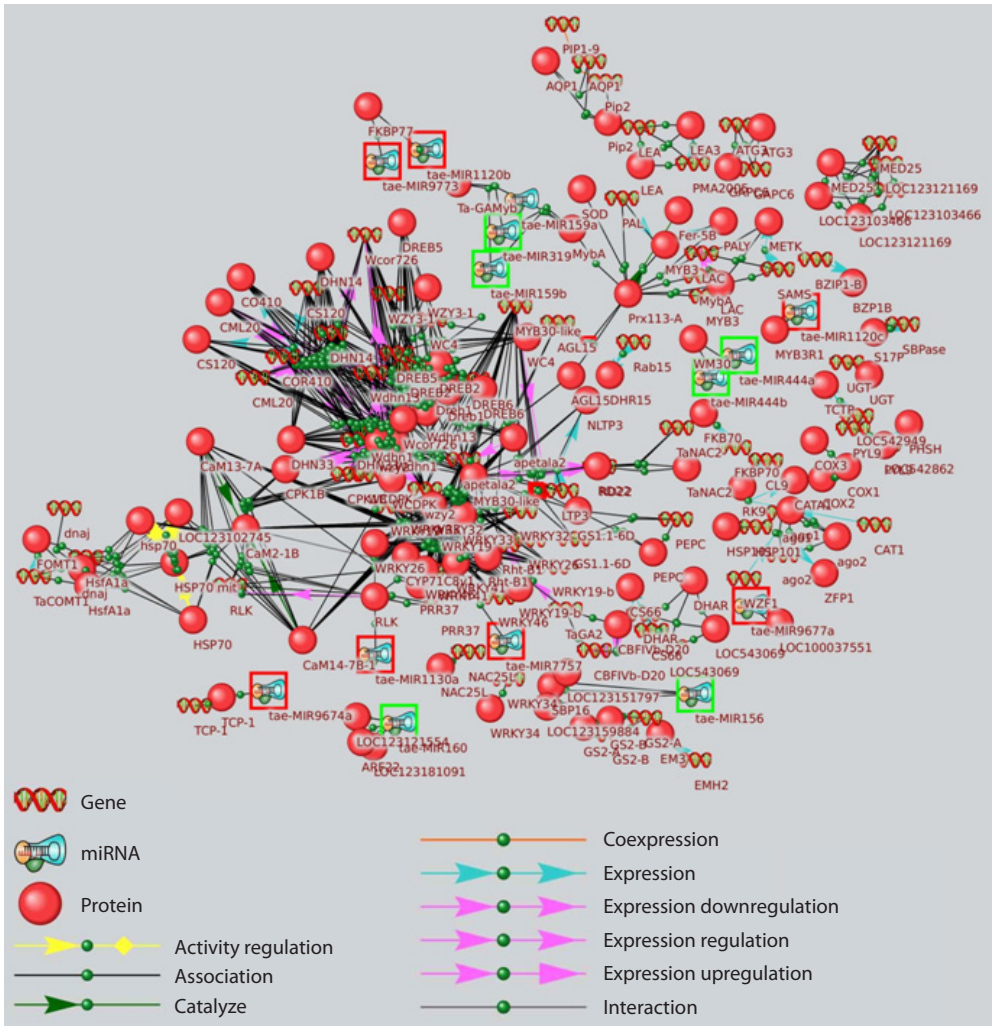


Рис. 1. Ассоциативная сеть генов и белков, для которых экспериментально доказано их участие в адаптации пшеницы к дефициту влаги, дополненная микроРНК, регулирующими их напрямую.

Зелеными рамками обозначены микроРНК, для которых имеются данные об их связи с засухой, красными рамками – микроРНК, для которых таких данных нет.

Таблица 1. Список микроРНК пшеницы, напрямую регулирующих гены ответа на засуху

микроРНК	Ген-мишень	Литературный источник
MIR160	*ARF22, LOC123121554, LOC123181091*	Kumar et al., 2015
MIR319	*MybA, Ta-GAMyb*	Li Y.-F. et al., 2013
MIR159a	*MybA, Ta-GAMyb*	Liu et al., 2015
MIR159b	*MybA, Ta-GAMyb*	Liu et al., 2015
MIR1120b	*FKBP77*	–
MIR1120c	*MYB3R1*	–
MIR1130a	*RLK*	–
MIR156	*LOC123151797, LOC123159884, SBP16*	Singroha et al., 2021
MIR444a	*WM30*	–
MIR444b	*WM30*	–
MIR7757	*WRKY41*	–
MIR9674a	*TCP-1*	–
MIR9677a	*LOC100037551*	–
MIR9773	*FKBP77*	–

Примечание. Жирным шрифтом выделены микроРНК, для которых экспериментально показано изменение в экспрессии в ответ на недостаток влаги. Для них приведена ссылка на соответствующий литературный источник.

частях растений микроРНК *MIR159* подавляет экспрессию транскрипционного фактора GAMYB, который является ингибитором роста, обеспечивая нормальное развитие растений (Millar et al., 2019). Экспрессия *MIR159* изменялась в ответ на засуху, наряду с изменением экспрессии гена *GAMyb* у картофеля (Yang J. et al., 2014) и мягкой пшеницы (Liu et al., 2015). Кроме того, продукт гена *MybA* регулирует экспрессию гена пероксидазы (Wei et al., 2021), способствующей адаптации растений к неблагоприятным факторам окружающей среды. Мишени микроРНК *MIR160* – гены, кодирующие транскрипционный фактор ARF, ключевой компонент, обеспечивающий реакцию растений на ауксины (Li Y. et al., 2023) – фитогормоны, которые, в частности, стимулируют апикальное доминирование, способствующее росту корня в длину, что является морфологической адаптацией растений к недостатку влаги.

Помимо транскрипционных факторов, также мишень микроРНК в генной сети ответа на засуху – ген серин/треонинкиназы RLK, который, взаимодействуя с кальмодулинами, участвует в адаптации растений к абиотическому стрессу (Virdi et al., 2015).

Таким образом, в результате анализа генной сети, включающей в себя гены и белки, роль которых в реакции на засуху показана экспериментально, установлен ряд микроРНК, регулирующих важные узлы этой генной сети (транскрипционные факторы), при этом некоторые микроРНК (*MIR1120*, *MIR1120c*, *MIR1130a*, *MIR444a*, *MIR444b*, *MIR7757*, *MIR9674a*, *MIR9677a*, *MIR9773*) ранее не обсуждались в литературе в связи с адаптацией пшеницы к засухе и могут быть перспективными для дальнейшего исследования.

Однако следует отметить, что микроРНК часто имеют много генов-мишеней, которые также могут быть компонентами генной сети ответа на засуху, хотя их роль в настоящее время экспериментально не установлена. Кроме того, микроРНК могут регулировать гены, контролирующие ответ на стресс не только напрямую, но и через посредников. Поэтому исходная генная сеть была дополнена с использованием базы знаний Smart crop, следующими компонентами: 1) всеми предсказанными, по данным Smart crop, мишенями тех 14 микроРНК, которые напрямую регулируют известные гены ответа на засуху и приведены в табл. 1; 2) генами и белками, которые непосредственно связаны с генами ответа на засуху, а также регулирующие их микроРНК.

Результирующая ассоциативная сеть представлена в Приложении 2. Список генов и белков, входящих в данную сеть, приведен в Приложении 3. Сеть включает 144 гена, 1017 белков и 21 микроРНК пшеницы, а также 5188 связей между компонентами сети. Из них 4158 связей соответствуют типу связей «ассоциация», 372 связи – «взаимодействие», 329 связей – «катализ», 180 связей – «регуляция экспрессии», 42 связи – «регуляция активности», 24 связи – «cleavage», 21 связь – «экспрессия», 15 связей – «подавление экспрессии», 12 связей – «усиление экспрессии», 7 связей – «коэкспрессия».

Функциональная аннотация всех компонентов (генов и белков) расширенной ассоциативной генной сети приведена в табл. 2. Как видно из табл. 2, компоненты генной сети оказались достоверно обогащены терминами, характеризующими биологические процессы, связанные со сборкой центриолей, а также морфогенезом побегов (regulation of morphogenesis of a branching structure), задержкой постэмбрионального развития, реакцией на абиотические и биотические стрессовые факторы, ответом на абсцизовую кислоту. Кроме того, компоненты генной сети вовлечены в сигнальные пути митоген-зависимой протеинкиназы, фосфатидилинозитола.

Интересно отметить, что расширенная генная сеть включает гены, участвующие не только в адаптации к водному дефициту (термин генной онтологии GO:0009414 response to water deprivation), но и в реакции растений к другим неблагоприятным факторам, включая адаптацию к холоду (термин генной онтологии GO:0009631 cold acclimation) и взаимодействие с патогенами (путь KEGG taes04626 plant-pathogen interaction). Вероятно, это обусловлено тем, что продукты одних и тех же генов могут участвовать в реакции растений на различные стрессовые факторы, обеспечивая их адаптацию к комплексу неблагоприятных факторов. В частности, гены, которые в нашей генной сети ассоциированы с термином “cold acclimation” (GO:0009631), принадлежат к семействам дегидринов и белков холодового шока (cold-shock proteins).

Известно, что белки семейства дегидринов, участвуя в стабилизации клеточной мембраны, способствуют адаптации растений к абиотическим стрессовым факторам различной природы, включая недостаток влаги, снижение температуры, засоление почвы (Szlachtowska, Rurek, 2023). С другой стороны, белки холодового шока, которые являются важнейшими участниками холодовой адаптации растений, также могут играть определенную роль в реакции растений на дефицит влаги, регулируя активность генов, продукты которых участвуют в антиоксидантной защите клеток (Yu T.F. et al., 2017; Li C. et al., 2021a). Кроме того, по данным литературы, такие компоненты генной сети ответа на засуху, как кальмодулины (Cheval et al., 2013) и транскрипционные факторы WRKY (Wani et al., 2021), могут участвовать также в регуляции иммунитета растений и защите их от патогенов.

В расширенной ассоциативной генной сети наибольшим числом связей с другими компонентами сети (Network Connectivity) обладали транскрипционный фактор MYB30-like, белки кальмодулины (CaM13-7A, CaM14-7B-1, CaM2-1B), белок APETALA2-like protein, являющийся членом подсемейства APETALA2 (AP2) семейства транскрипционных факторов AP2/Ethylene Responsive Factor (ERF), а также белок RHT1, транскрипционный фактор WRKY41, цитохром P450 (CYP71C8v1). Гены, кодирующие эти белки, уже обсуждались в литературе как контролирующие ответ растений на дефицит влаги.

Транскрипционные факторы MYB – одни из наиболее распространенных семейств факторов транскрипции в растениях, которые участвуют в развитии растений и ответе на различные неблагоприятные факторы окружающей среды, включая недостаток влаги. Транскрипционные факторы MYB, связываясь с MYB цис-элементами в промоторах множества генов-мишеней, регулируют ряд биологических процессов, в частности биосинтез флавоноидов, которые необходимы для защиты от оксида-

Таблица 2. Функциональная аннотация расширенной генной сети ответа на засуху пшеницы

Термин	Число генов	FE	*p*-value
Биологические процессы			
GO:0098534~centriole assembly	4	68.3	0.0065
GO:2000032~regulation of secondary shoot formation	10	23.0	0.0000
GO:0060688~regulation of morphogenesis of a branching structure	10	23.0	0.0000
GO:0048581~negative regulation of post-embryonic development	5	20.3	0.0390
GO:0009631~cold acclimation	10	12.6	0.0000
GO:0009414~response to water deprivation	14	11.4	0.0000
GO:0009737~response to abscisic acid	25	6.2	0.0000
GO:0009891~positive regulation of biosynthetic process	32	2.6	0.0013
GO:0045935~positive regulation of nucleobase-containing compound metabolic process	32	2.6	0.0018
GO:0098542~defense response to other organism	89	2.5	0.0000
GO:0051252~regulation of RNA metabolic process	139	1.7	0.0000
GO:0010556~regulation of macromolecule biosynthetic process	155	1.6	0.0000
Молекулярные функции			
GO:0043531~ADP binding	95	3.2	0.0000
GO:0043565~sequence-specific DNA binding	103	3.1	0.0000
GO:0000976~transcription cis-regulatory region binding	59	3.0	0.0000
GO:0003690~double-stranded DNA binding	60	2.2	0.0000
Пути KEGG			
taes04016:MAPK signaling pathway – plant	33	4.4	0.0000
taes04626:Plant-pathogen interaction	32	2.9	0.0000
taes04070:Phosphatidylinositol signaling system	12	4.0	0.0161

Примечание. FE – foldenrichment; *p*-value – показатель статистической значимости обогащения генов и белков ассоциативной сети с поправкой Бонферрони.

тивного стресса. Кроме того, транскрипционные факторы MYB активируют гены, контролирующие образование эпикутикулярного воска, снижающего испарение влаги листьями растений (Wang X. et al., 2021).

Известно, что кальций – это важнейший вторичный посредник, концентрация которого изменяется в ответ на неблагоприятные факторы различной природы, включая недостаток влаги. Кальмодулины и кальмодулин-подобные белки, связываясь с ионами кальция, изменяют свою конформацию и модулируют активность целого ряда других белков, включая киназы, транскрипционные факторы, транспортеры и ферменты различных метаболических путей, которые обеспечивают адаптацию растений к окружающей среде (Ranty et al., 2016). В частности, повышенная экспрессия гена, кодирующего один из белков семейства кальмодулинов, у пшеницы наблюдалась в ответ на недостаток влаги и повышенное засоление, а экспрессия этого гена у трансгенных растений арабидопсиса увеличивала их устойчивость к данным неблагоприятным факторам (Li Y. et al., 2022).

Белки подсемейства APETALA2 (AP2) принадлежат к семейству транскрипционных факторов AP2/Ethylene Responsive Factor (ERF), которые регулируют экспрессию генов, обеспечивающих адаптацию к неблагоприятным условиям среды, включая засуху (Park S.Y., Grabau, 2016; Srivastava, Kumar, 2018). Экспрессия генов, кодирующих белки подсемейства AP2, *TaAP2-1-1A*, *TaAP2-1-1D*, была повышена в ответ на засуху у пшеницы (Yu Y. et al., 2022).

Белки семейства цитохромов P450 – ферменты, вовлеченные во множество метаболических путей синтеза вторичных метаболитов растений, фитогормонов и антиоксидантов, играющих важную роль в адаптации растений к окружающей среде (Pandian et al., 2020). В исследовании (Li Y., Wei, 2020) показано, что у пшеницы в ответ на засуху изменялась экспрессия 77 генов, кодирующих цитохромы P450, которые участвуют в биосинтезе абсцизовой кислоты, являющейся важным посредником, активирующим различные сигнальные каскады в стресс-реакциях растений, а также цитохромы P450, участвующие в синтезе флавоноидов, играющих важную роль в антиоксидантной защите растительной клетки.

Среди белков-посредников, связанных с экспериментально найденными генами ответа на засуху, наибольшим числом связей с другими компонентами сети обладал ген *LOC123186119*, кодирующий белок устойчивости к заболеваниям RPM1, который связан со всеми транскрипционными факторами семейства WRKY, представленными в сети, а также с зависимой от кальция протеинкиназой 7 и 19. При этом белок RPM1 является мишенью микроРНК *MIR7757*. Список 21 микроРНК, связанной с компонен-

Таблица 3. Список микроРНК и их генов-мишеней с наиболее высоким числом связей в сети

микроРНК	MirBaseID	Число мишеней микроРНК, содержащихся в сети	Мишень с максимальным числом связей, представленных в сети	Число связей, которые имеет мишень в сети
MIR1120b	MI0030404	149	***FKBP77***	4
MIR1130a	MI0006192	148	*LOC123051594*, *LOC123091557*, *LOC123096508*	5
MIR159b	MI0006171	143	***MybA***	11
MIR7757	MI0030410	102	*LOC123186119* (*RPM1*) и *WRKY41*	49
MIR1120c	MI0030409	98	*LOC123078649*	3
MIR444a	MI0006178	58	*LOC123078649*	2
MIR444b	MI0016467	58	*LOC100037552*	2
MIR9773	MI0031525	54	***FKBP77***	4
MIR9674a	MI0030403	47	***TCP-1***	1
MIR159a	MI0006170	36	***MybA***	11
MIR156	MI0016450	24	***SBP16***	1
MIR319	MI0016453	22	***MybA***	11
MIR160	MI0006172	20	***ARF22***	1
MIR9677a	MI0030414	18	***LOC100037551***	1
MIR9668	MI0030392	2	*LOC543328*	7
MIR1121	MI0006183	1	*UCRIA*	3
MIR395b	MI0016464	1	*LOC123190485*	3
MIR9653a	MI0030370	1	*LOC543111*	25
MIR9671	MI0030395	1	*LOC543244*	36
MIR9672b	MI0031526	1	*LOC543244*	36
MIR9679	MI0030418	1	*LOC123114245*	5

Примечание. Жирным шрифтом выделены микроРНК и гены-мишени, для которых в литературе имеются сведения об их значении в адаптации к засухе у пшеницы. Подчеркнуты микроРНК, которые напрямую регулируют известные гены ответа на водный дефицит.

тами расширенной генной сети, приведен в табл. 3. Полный список 984 предсказанных мишеней микроРНК по данным БД Smart crop представлен в Приложении 4.

Результаты функциональной аннотации генов-мишеней микроРНК в ассоциативной сети приведены в табл. 4. Как видно из табл. 4, мишени микроРНК в генной сети, связанной с ответом на засуху, вовлечены в процессы морфогенеза боковых побегов и корней растений, регуляцию иммунитета, транспорт и метаболизм пуринов, а также функционирование транскрипционных факторов. К генам, контролирующим процессы морфогенеза побегов в расширенной генной сети (см. Приложение 2), относятся главным образом гены-мишени микроРНК *miR319*, кодирующие семейство транскрипционных факторов TEOSINTE BRANCHED/CYCLOIDEA/PCF (TCP), которые вовлечены в формирование архитектуры побегов и корней растений (Tokizawa et al., 2023), включая образование корневых волосков (Wang M.Y. et al., 2013), что является важной морфологической адаптацией растений к недостатку влаги.

Участие транскрипционных факторов семейства TEOSINTE BRANCHED/CYCLOIDEA/PCF (TCP) в ответе на недостаточное увлажнение обсуждается в литературе (Manna et al., 2021), хотя для пшеницы не показано участие этих транскрипционных факторов в реакции на дефицит влаги. Нокаут членов семейства микроРНК miR319 *IbmiR319a* и *IbmiR319c* в трансгенных растениях батата привел к повышению их чувствительности к недостатку влаги, увеличению числа устьиц, снижению содержания лигнина и нарушению гормональной регуляции роста растений (Ren et al., 2022). Авторы предполагают, что данные морфологические изменения вызваны изменениями в экспрессии транскрипционного фактора TCP11/17, который является мишенью *bmiR319a* и *IbmiR319c.*

Из 21 микроРНК в расширенной генной сети (см. табл. 3) 14 были напрямую связаны с генами, роль которых в адаптации пшеницы к недостатку влаги доказана экспериментально. Семь микроРНК (*MIR9668*, *MIR1121*, *MIR395b*, *MIR9653a*, *MIR9671*, *MIR9672b*, *MIR9679*) были связаны с генами ответа на засуху через посредника. Наибольшим числом связей с другими компонентами сети обладали микроРНК *MIR1120b*, *MIR1130a*, *MIR159b*, *MIR7757* и *MIR1120c.*

В частности, интересно отметить, что не только микроРНК *MIR7757* оказалась связанной с большим количеством узлов генной сети (102), но и ее мишень – ген *LOC123186119*, белок устойчивости к заболеваниям RPM1, был связан с наибольшим (49) числом других узлов сети. К этим узлам относятся набор транскрипционных факторов семейства WRKY, а также зависимые от кальция протеинкиназы 7 (WCDPK) и 19 (CPK 1B), белки SKP1 и SGT1, белок теплового шока HSP80 (рис. 2).

Таблица 4. Функциональная аннотация генов-мишеней микроРНК пшеницы в генной сети ответа на засуху

Термин	Число генов	FE	*p*-value
GO:2000032~regulation of secondary shoot formation	10	32.03	0.0000
GO:1905428~regulation of plant organ formation	10	32.03	0.0000
GO:0060688~regulation of morphogenesis of a branching structure	10	32.03	0.0000
GO:0098542~defense response to other organism	87	3.36	0.0000
GO:0015211~purine nucleoside transmembrane transporter activity	6	11.89	0.0216
GO:0043531~ADP binding	95	4.34	0.0000
GO:0043565~sequence-specific DNA binding	48	1.95	0.0027
GO:0003700~DNA-binding transcription factor activity	60	1.81	0.0006
GO:0032559~adenyl ribonucleotide binding	161	1.48	0.0000
GO:0017076~purine nucleotide binding	172	1.42	0.0000
GO:0032553~ribonucleotide binding	166	1.41	0.0000

Примечание. FE – foldenrichment, *p*-value – показатель статистической значимости обогащения с поправкой Бонферрони.

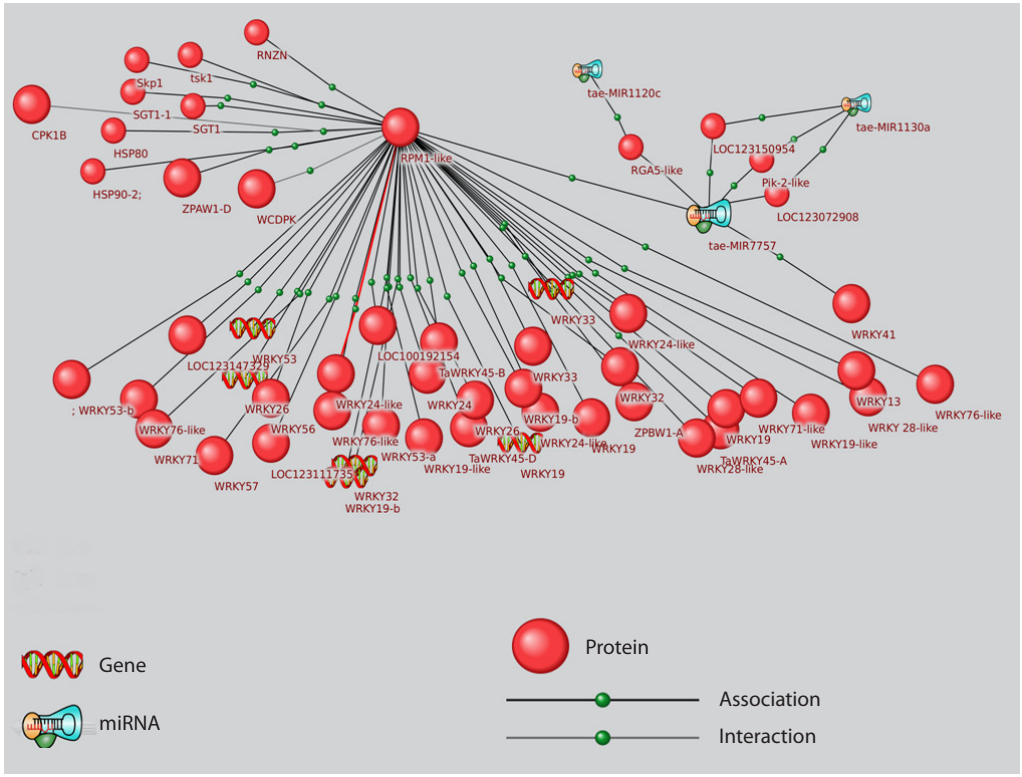


Рис. 2. Ассоциативная сеть микроРНК *MIR7757*, ее мишеней и связанных с мишенями посредников.
Крупными шарами отмечены белки, для которых экспериментально показано их участие в реакции на засуху.

Многочисленные данные, полученные на разных видах растений, свидетельствуют о том, что транскрипционные факторы семейства WRKY играют важнейшую роль в адаптации к самым различным стрессовым факторам, включая дефицит влаги. Повышение экспрессии транскрипционных факторов WRKY способствует уменьшению потери ионов, активации устьичного аппарата листьев, уменьшению потери влаги, снижению содержания активных форм кислорода (Khoso et al., 2022).

Известно, что транскрипционные факторы WRKY модулируют активность сигнальных путей фитогормонов – салициловой кислоты, этилена, абсцизовой кислоты, жасмоновой кислоты, митоген-активируемой протеинкиназы MAPK (Jiang et al., 2017), а также кальмодулинов, в том числе путем физического взаимодействия с доменом кальция в кальмодулинах (Park C.Y. et al., 2005). При этом активность транскрипционных факторов WRKY находится под контролем различных сигнальных путей и фито-

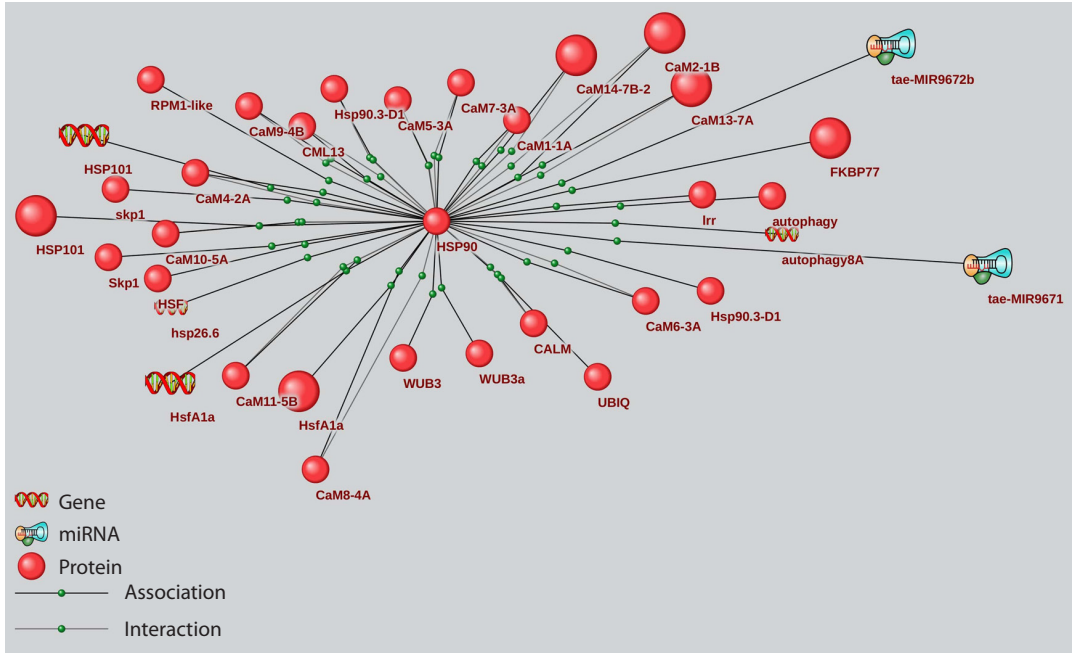


Рис. 3. Ассоциативная сеть микроРНК *MIR9671*, *MIR9672b*, ее мишеней и связанных с мишенями белков. Крупными шарами отмечены белки, для которых экспериментально показано их участие в реакции на засуху.

гормонов, включая этилен (Li J. et al., 2006), абсцизовую кислоту (Chen et al., 2010), сигнальный путь MAPK (Mao et al., 2011), что обеспечивает изменение активности WRKY в зависимости от условий окружающей среды. Таким образом, транскрипционные факторы WRKY – важнейшее регуляторное звено ответа растений на стресс, влияет на активность множества генов, регулирующих адаптацию, при этом активность WRKY может изменяться в зависимости от характера воздействия, что обеспечивает гибкую адаптацию растений к изменяющимся условиям среды.

Кальмодулины и зависимые от кальция протеинкиназы, связываясь с ионами кальция, концентрация которых возрастает в ответ на стрессовые факторы, изменяют функционирование сигнальных путей абсцизовой кислоты, что, в свою очередь, вызывает изменение скорости созревания семян, закрытие устьиц, снижение содержания активных форм кислорода (Asano et al., 2012).

Белок SKP1 входит в состав комплекса SCF (Skp1-Cullin 1-F-box), который является убиквитин-лигазой, играющей важную роль в передаче гормональных сигналов, регуляции циркадного ритма, роста и развития растений (Hong et al., 2012), адаптации к неблагоприятным факторам (Saxena et al., 2023). Таким образом, *MIR7757* может быть важнейшим мастер-регулятором генной сети ответа на дефицит влаги, действуя как напрямую на транскрипционный фактор WRKY41, так и через посредника – белок RPM1-like, координируя сигнальные пути фитогормонов, MAPK и кальций-зависимых протеинкиназ. Этот белок играет важную роль в иммунитете растений, однако его значение в ответе пшеницы на водный дефицит неизвестно, хотя сообщалось, что экспрессия гена *PRM1* была повышена в листьях винограда в ответ на недостаток влаги (Haider et al., 2017). Кроме того, нет данных об изменениях в экспрессии микроРНК *MIR7757* пшеницы при недостатке влаги, поэтому эта микроРНК, а также другие микроРНК, обладающие высоким числом связей с узлами сети, и их гены-мишени являются перспективными кандидатами для экспериментального исследования микроРНК регуляции ответа пшеницы на водный дефицит.

Мишень двух других микроРНК, *MIR9671*, *MIR9672b*, белок теплового шока 90, кодируемый геном *LOC543244*, также имеет обширные (n = 36) связи с другими узлами генной сети, а именно кальмодулинами (CaM14-7B-1, LOC123104984 и др.), белком теплового шока 101, белками SKP1 и RPM1 (которые обсуждались выше), транскрипционным фактором теплового стресса HSf1a, полиубиквитином UBIQ (рис. 3).

Известно, что белок HSP90, высококонсервативный шаперон, является важнейшим компонентом гомеостаза клеток эукариот и участвует в адаптации растений к различным видам абиотического стресса, модуляции роста и развития растений, взаимодействуя с сигнальными путями ауксина и жасмоновой кислоты. Белок HSP90 совместно со своими кошаперонами стабилизирует рецепторный комплекс ауксина в условиях повышения температуры воздуха (экологический фактор, который часто сопровождает недостаток влаги) и способствует физиологическим и морфологическим адаптациям, индуцируемым ауксином, в частности удлинению корня (di Donato, Geisler, 2019). Кроме того, HSP90, взаимодействуя с протеин-лигазами, содействует удалению поврежденных белков.

Следует отметить, что многочисленные белки-кальмодулины, связываясь с ионами кальция при стрессе, активируют не только сигнальные пути кальций-зависимых протеинкиназ, но и экспрессию HSP90 (Virdi et al., 2011), обеспечивая дополнительную, опосредованную белком теплового шока активацию гормональной системы растений. Таким образом, микроРНК *MIR9671*, *MIR9672b* через

их мишень HSP90 могут модулировать гормональную сигнализацию ауксина и жасмоновой кислоты, а также функционирование системы убиквитинирования белков при абиотическом стрессе.

Учитывая важную роль белка HSP90 в реакции на абиотический стресс, можно предположить, что усиление его экспрессии путем искусственного ослабления активности или экспрессии микроРНК *MIR9671*, *MIR9672b* может увеличить устойчивость растений пшеницы к недостатку влаги. Однако следует отметить, что HSP90 обладает плейотропным эффектом, воздействуя на значительное число сигнальных путей в клетке (di Donato, Geisler, 2019), поэтому опосредованное микроРНК ослабление его экспрессии может быть необходимым для адаптивного изменения некоторых сигнальных путей на определенной стадии развития растений или при изменениях условий окружающей среды.

Таким образом, микроРНК *MIR9671*, *MIR9672b* наряду с *MIR7757*, которые ранее не обсуждались в литературе в связи с реакцией пшеницы на засуху, могут быть перспективными для дальнейшего экспериментального исследования микроРНК регуляции ответа мягкой пшеницы на водный дефицит.

В ряде экспериментов, проведенных на различных видах растений, показано, что искусственная модуляция экспрессии микроРНК позволяет изменять функционирование регуляторных генных сетей, влияя на экспрессию генов, ответственных за адаптацию к неблагоприятным условиям среды или формирование определенных хозяйственно ценных признаков. Современные технологии генной инженерии – РНК-интерференция, создание специальных векторов, экспрессирующих определенную микроРНК, а также методы геномного редактирования, такие как CRISPR/Cas9 и Transcription activator-like effector nucleases (TALEN), дают возможность усиливать или ослаблять экспрессию и активность микроРНК в зависимости от того, какое влияние (стимулирующее или ослабляющее) оказывают продукты генов-мишеней микроРНК на целевые биологические процессы (Abbas et al., 2022; Raza et al., 2023). Например, искусственное повышение экспрессии микроРНК *miR319* с помощью специальных векторов, равно как и искусственное снижение экспрессии ее мишеней – транскрипционных факторов TEOSINTE BRANCHED/CYCLOIDEA/PCF (TCP), приводило к повышению устойчивости растений риса к холоду (Yang C. et al., 2013). В другом исследовании (Ni et al., 2013) подавление экспрессии микроРНК *miR169* увеличивало экспрессию ее мишени – транкрипционного фактора *GmNFYA3*, что улучшало устойчивость растений сои к засухе за счет изменения экспрессии генов, связанных с адаптацией к водному дефициту. Искусственное повышение экспрессии микроРНК *miR172b-3p* в трансгенном картофеле приводило к ослаблению экспрессии гена *ERFRAP2-7-like* и усилению фиксации углерода растениями (Raza et al., 2023).

Кроме того, получены патенты на трансгенные растения, созданные с применением микроРНК, которые обладают повышенной продуктивностью и устойчивостью к неблагоприятным факторам среды, что свидетельствует об успешном практическом использовании микроРНК для создания новых сортов растений. Например, получен трансгенный рис с повышенной экспрессией микроРНК *Osa-miR393* и усиленным кущением (патент CN102533760A (Wang S., Zhang, 2011)). Созданы трансгенные растения томата с подавленной экспрессией микроРНК *miR156e-3p* и повышенной устойчивостью к низким температурам (патент CN111705077B (Zhou et al., 2020)).

Таким образом, модуляция активности микроРНК с использованием генной инженерии может стать перспективным методом современной биотехнологии, направленным на увеличение устойчивости растений к неблагоприятным условиям окружающей среды, включая недостаток влаги, и в конечном итоге их продуктивности.

Заключение

С использованием базы знаний Smart crop программно-информационной системы ANDSystem произведена реконструкция генной сети микроРНК регуляции адаптации мягкой пшеницы к недостатку влаги. Гены, входящие в сеть, регулируют процессы морфогенеза корней и побегов, ответ на абиотические и биотические стрессовые факторы и вовлечены в сигнальные пути абсцизовой кислоты, кальций-зависимых протеинкиназ.

Выявлена 21 микроРНК, регулирующая генную сеть ответа пшеницы на засуху, мишени которых вовлечены главным образом в контроль процессов морфогенеза растений. Наиболее значимые узлы в этой сети, регулируемые микроРНК, – транскрипционные факторы MYBa, WRKY41, белок теплового шока HSP90, а также белок RPM1, связанный с белками транскрипционных факторов семейства WRKY, кальций-зависимыми протеинкиназами и сигнальными путями фитогормонов – ауксина, жасмоновой кислоты, абсцизовой кислоты, имеющими важнейшее значение в контроле адаптации растений к недостатку влаги. Ряд микроРНК, которые ранее не обсуждались в литературе в связи с адаптацией к засухе (*MIR7757*, *MIR9671*, *MIR9672b*), регулируют значительное число узлов сети и поэтому могут быть перспективными для дальнейшего экспериментального исследования микроРНК регуляции ответа мягкой пшеницы на водный дефицит.

Список литературы / References

Деменков П.С., Сайк О.В., Иванисенко Т.В., Колчанов Н.А., Кочетов А.В., Иванисенко В.А. Приоритизация генов картофеля, вовлеченных в формирование селекционно-значимых признаков, с использованием базы знаний SOLANUM TUBEROSUM. *Вавиловский журнал генетики и селекции.* 2019;23(3):312-319. doi 10.18699/VJ19.501

[Demenkov P.S., Saik O.V., Ivanisenko T.V., Kolchanov N.A., Kochetov A.V., Ivanisenko V.A. Prioritization of potato genes involved in the formation of agronomically valuable traits using the SOLANUM TUBEROSUM knowledge base. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2019;23(3):312-319. doi 10.18699/VJ19.501

Иванисенко В.А., Басов Н.В., Макарова А.А., Вензель А.С., Рогачев А.Д., Деменков П.С., Иванисенко Т.В., Клещев М.А., Гайслер Е.В., Мороз Г.Б., Плеско В.В., Сотникова Ю.С., Патрушев Ю.В., Ломиворотов В.В., Колчанов Н.А., Покровский А.Г. Применение генных сетей к анализу результатов метаболомного скрининга плазмы крови пациентов с послеоперационным делирием. *Вавиловский журнал генетики и селекции*. 2023; 27(7):768-775. doi 10.18699/VJGB-23-89

[Ivanisenko V.A., Basov N.V., Makarova A.A., Venzel A.S., Rogachev A.D., Demenkov P.S., Ivanisenko T.V., Kleshchev M.A., Gaisler E.V., Moroz G.B., Plesko V.V., Sotnikova Y.S., Patrushev Y.V., Lomivorotov V.V., Kolchanov N.A., Pokrovsky A.G. Gene networks for use in metabolomic data analysis of blood plasma from patients with postoperative delirium. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):768-775. doi 10.18699/VJGB-23-89]

Иванисенко Т.В., Сайк О.В., Деменков П.С., Хлесткин В.К., Хлесткина Е.К., Колчанов Н.А., Иванисенко В.А. База знаний SOLANUM TUBEROSUM: раздел по молекулярно-генетической регуляции метаболических путей. *Вавиловский журнал генетики и селекции*. 2018;22(1):8-17. doi 10.18699/VJ18.325
[Ivanisenko T.V., Sayk O.V., Demenkov P.S., Khlestkin V.K., Khlestkina E.K., Kolchanov N.A., Ivanisenko V.A. The SOLANUM TUBEROSUM knowledge base: the section on molecular-genetic regulation of metabolic pathways. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2018;22(1): 8-17. doi 10.18699/VJ18.325 (in Russian)]

Пакуль А.Л., Лапшинов Н.А., Божанова Г.В., Пакуль В.Н. Основные факторы, влияющие на продуктивность агроценозов яровой мягкой пшеницы. *Сиб. вестн. с.-х. науки*. 2018;48(6):21-29. doi 10.26898/0370-8799-2018-6-3
[Pakul A.L., Lapshinov N.A., Bozhanova G.V., Pakul V.N. The main factors influencing efficiency of spring common wheat agrocenosis. *Sibirskii Vestnik Sel'skokhozyajstvennoi Nauki = Siberian Herald of Agricultural Science*. 2018;48(6):21-29. doi 10.26898/0370-8799-2018-6-3 (in Russian)]

Abbas A., Shah A.N., Tanveer M., Ahmed W., Shah A.A., Fiaz S., Waqas M.M., Ullah S. MiRNA fine tuning for crop improvement: using advance computational models and biotechnological tools. *Mol. Biol. Rep*. 2022;49(6):5437-5450. doi 10.1007/s11033-022-07231-5

Alptekin B., Langridge P., Budak H. Abiotic stress miRNomes in the *Triticeae*. *Funct. Integr. Genomics*. 2017;17(2-3):145-170. doi 10.1007/s10142-016-0525-9

Antropova E.A., Khlebodarova T.M., Demenkov P.S., Volianskaia A.R., Venzel A.S., Ivanisenko N.V., Gavrilenko A.D., Ivanisenko T.V., Adamovskaya A.V., Revva P.M., Kolchanov N.A., Lavrik I.N., Ivanisenko V.A. Reconstruction of the regulatory hypermethylation network controlling hepatocellular carcinoma development during hepatitis C viral infection. *J. Integr. Bioinform*. 2023;20(3):20230013. doi 10.1515/jib-2023-0013

Asano T., Hayashi N., Kikuchi S., Ohsugi R. CDPK-mediated abiotic stress signaling. *Plant Signal Behav.* 2012;7(7):817-821. doi 10.4161/psb.20351

Baillo E.H., Kimotho R.N., Zhang Z., Xu P. Transcription factors associated with abiotic and biotic stress tolerance and their potential for crops improvement. *Genes* (*Basel*). 2019;10(10):771. doi 10.3390/genes10100771

Bragina E.Y., Tiys E.S., Freidin M.B., Koneva L.A., Demenkov P.S., Ivanisenko V.A., Kolchanov N.A., Puzyrev V.P. Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis. *Immunogenetics*. 2014;66(7-8):457-465. doi 10.1007/s00251-014-0786-1

Bragina E.Y., Tiys E.S., Rudko A.A., Ivanisenko V.A., Freidin M.B. Novel tuberculosis susceptibility candidate genes revealed by the reconstruction and analysis of associative networks. *Infect. Genet. Evol*. 2016;46:118-123. doi 10.1016/j.meegid.2016.10.030

Bragina E.Y., Gomboeva D.E., Saik O.V., Ivanisenko V.A., Freidin M.B., Nazarenko M.S., Puzyrev V.P. Apoptosis genes as a key to identification of inverse comorbidity of huntington's disease and cancer. *Int. J. Mol. Sci*. 2023;24(11):9385. doi 10.3390/ijms24119385

Chao H., Zhang S., Hu Y., Ni Q., Xin S., Zhao L., Ivanisenko V.A., Orlov Y.L., Chen M. Integrating omics databases for enhanced crop breeding. *J. Integr. Bioinform*. 2023;20(4):20230012. doi 10.1515/jib-2023-0012

Chen H., Lai Z., Shi J., Xiao Y., Chen Z., Xu X. Roles of arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biol*. 2010;10:281. doi 10.1186/1471-2229-10-281

Cheval C., Aldon D., Galaud J.P., Ranty B. Calcium/calmodulin-mediated regulation of plant immunity. *Biochim. Biophys. Acta*. 2013;1833(7):1766-1771. doi 10.1016/j.bbamcr.2013.01.031

Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol*. 2012;11(3-4):149-161. doi 10.3233/ISB-2012-0449

di Donato M., Geisler M. HSP90 and co-chaperones: a multitaskers' view on plant hormone biology. *FEBS Lett*. 2019;593(13):1415-1430. doi 10.1002/1873-3468.13499

Ferdous J., Hussain S.S., Shi B.J. Role of microRNAs in plant drought tolerance. *Plant Biotechnol. J*. 2015;13(3):293-305. doi 10.1111/pbi.12318

Gahlaut V., Jaiswal V., Kumar A., Gupta P.K. Transcription factors involved in drought tolerance and their possible role in developing drought tolerant cultivars with emphasis on wheat (*Triticum aestivum* L.). *Theor. Appl. Genet*. 2016;129(11):2019-2042. doi 10.1007/s00122-016-2794-z

Gupta O.P., Meena N.L., Sharma I., Sharma P. Differential regulation of microRNAs in response to osmotic, salt and cold stresses in wheat. *Mol. Biol. Rep*. 2014;41(7):4623-4629. doi 10.1007/s11033-014-3333-0

Haider M.S., Kurjogi M.M., Khalil-Ur-Rehman M., Fiaz M., Pervaiz T., Jiu S., Haifeng J., Chen W., Fang J. Grapevine immune signaling network in response to drought stress as revealed by transcriptomic analysis. *Plant Physiol. Biochem*. 2017;121:187-195. doi 10.1016/j.plaphy.2017.10.026

Hong M.J., Kim D.Y., Kang S.Y., Kim D.S., Kim J.B., Seo Y.W. Wheat F-box protein recruits proteins and regulates their abundance during wheat spike development. *Mol. Biol. Rep*. 2012;39(10):9681-9696. doi 10.1007/s11033-012-1833-3

Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*. 2020;21(Suppl. 11):228. doi 10.1186/s12859-020-03557-8

Ivanisenko T.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. The new version of the ANDDigest tool with improved ai-based short names recognition. *Int. J. Mol. Sci*. 2022;23(23):14934. doi 10.3390/ijms232314934

Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an associative network discovery system for automated literature mining in the field of biology. *BMC Syst. Biol*. 2015;9(Suppl. 2):S2. doi 10.1186/1752-0509-9-S2-S2

Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl. 1):34. doi 10.1186/s12859-018-2567-6

Ivanisenko V.A., Gaisler E.V., Basov N.V., Rogachev A.D., Cheresiz S.V., Ivanisenko T.V., Demenkov P.S., Mishchenko E.L., Khripko O.P., Khripko Y.I., Voevoda S.M., Karpenko T.N., Velichko A.J., Voevoda M.I., Kolchanov N.A., Pokrovsky A.G. Plasma metabolomics and gene regulatory networks analysis reveal the role of nonstructural SARS-CoV-2 viral proteins in metabolic dysregulation in COVID-19 patients. *Sci. Rep.* 2022;12(1):19977. doi 10.1038/s41598-022-24170-0

Jeyasri R., Muthuramalingam P., Satish L., Pandian S.K., Chen J.T., Ahmar S., Wang X., Mora-Poblete F., Ramesh M. An overview of abiotic stress in cereal crops: negative impacts, regulation, biotechnology and integrated omics. *Plants* (*Basel*). 2021;10(7):1472. doi 10.3390/plants10071472

Jiang J., Ma S., Ye N., Jiang M., Cao J., Zhang J. WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol*. 2017;59(2):86-101. doi 10.1111/jipb.12513

Khlebodarova T.M., Demenkov P.S., Ivanisenko T.V., Antropova E.A., Lavrik I.N., Ivanisenko V.A. Primary and secondary micro-RNA modulation the extrinsic pathway of apoptosis in hepatocellular carcinoma. *Mol. Biol*. 2023;57(2):165-175. doi 10.1134/S0026893323020103

Khoso M.A., Hussain A., Ritonga F.N., Ali Q., Channa M.M., Alshegaihi R.M., Meng Q., Ali M., Zaman W., Brohi R.D., Liu F., Manghwar H. WRKY transcription factors (TFs): Molecular switches to regulate drought, temperature, and salinity stresses in plants. *Front. Plant Sci*. 2022;13:1039329. doi 10.3389/fpls.2022.1039329

Kumar R.R., Pathak H., Sharma S.K., Kala Y.K., Nirjal M.K., Singh G.P., Goswami S., Rai R.D. Novel and conserved heat-responsive microRNAs in wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics.* 2015;15(3):323-348. doi 10.1007/s10142-014-0421-0

Langridge P., Reynolds M. Breeding for drought and heat tolerance in wheat. *Theor. Appl. Genet*. 2021;134(6):1753-1769. doi 10.1007/s00122-021-03795-1

Larina I.M., Pastushkova L.Kh., Tiys E.S., Kireev K.S., Kononikhin A.S., Starodubtseva N.L., Popov I.A., Custaud M.A., Dobrokhotov I.V., Nikolaev E.N., Kolchanov N.A., Ivanisenko V.A. Permanent proteins in the urine of healthy humans during the Mars-500 experiment. *J. Bioinform. Comput. Biol.* 2015;13(1):1540001. doi 10.1142/S0219720015400016

Li C., Hou N., Fang N., He J., Ma Z., Ma F., Guan Q., Li X. Cold shock protein 3 plays a negative role in apple drought tolerance by regulating oxidative stress response. *Plant Physiol. Biochem*. 2021a;168:83-92. doi 10.1016/j.plaphy.2021.10.003

Li C., Li L., Reynolds M.P., Wang J., Chang X., Mao X., Jing R. Recognizing the hidden half in wheat: root system attributes associated with drought tolerance. *J. Exp. Bot*. 2021b;72(14):5117-5133. doi 10.1093/jxb/erab124

Li J., Brader G., Kariola T., Palva E.T. WRKY70 modulates the selection of signaling pathways in plant defense. *Plant J*. 2006;46(3):477-491. doi 10.1111/j.1365-313X.2006.02712.x

Li N., Liu T., Guo F., Yang J., Shi Y., Wang S., Sun D. Identification of long non-coding RNA-microRNA-mRNA regulatory modules and their potential roles in drought stress response in wheat (*Triticum aestivum* L.). *Front. Plant Sci*. 2022;13:1011064. doi 10.3389/fpls.2022.1011064

Li Y., Wei K. Comparative functional genomics analysis of cytochrome P450 gene superfamily in wheat and maize. *BMC Plant Biol*. 2020;20(1):93. doi 10.1186/s12870-020-2288-7

Li Y., Zhang H., Dong F., Zou J., Gao C., Zhu Z., Liu Y. Multiple roles of wheat calmodulin genes during stress treatment and TaCAM2-D as a positive regulator in response to drought and salt tolerance. *Int. J. Biol. Macromol*. 2022;220:985-997. doi 10.1016/j.ijbiomac.2022.08.124

Li Y., Han S., Qi Y. Advances in structure and function of auxin response factor in plants. *J. Integr. Plant Biol*. 2023;65(3):617-632. doi 10.1111/jipb.13392

Li Y.-F., Zheng Y., Jagadeeswaran G., Sunkar R. Characterization of small RNAs and their target genes in wheat seedlings using sequencing-based approaches. *Plant Sci*. 2013;203-204:17-24. doi 10.1016/j.plantsci.2012.12.014

Liebsch D., Palatnik J.F. MicroRNA miR396, GRF transcription factors and GIF co-regulators: a conserved plant growth regulatory module with potential for breeding and biotechnology. *Curr. Opin. Plant Biol*. 2020;53:31-42. doi 10.1016/j.pbi.2019.09.008

Liu J., Feng L., Li J., He Z. Genetic and epigenetic control of plant heat responses. *Front. Plant Sci*. 2015;6:267. doi 10.3389/fpls.2015.00267

Ma Z., Hu L. MicroRNA: a dynamic player from signalling to abiotic tolerance in plants. *Int. J. Mol. Sci*. 2023;24(14):11364. doi 10.3390/ijms241411364

Manna M., Thakur T., Chirom O., Mandlik R., Deshmukh R., Salvi P. Transcription factors as key molecular target to strengthen the drought stress tolerance in plants. *Physiol. Plant.* 2021;172(2):847-868. doi 10.1111/ppl.13268

Mao G., Meng X., Liu Y., Zheng Z., Chen Z., Zhang S. Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in *Arabidopsis*. *Plant Cell*. 2011;23(4):1639-1653. doi 10.1105/tpc.111.084996

Millar A.A., Lohe A., Wong G. Biology and function of miR159 in plants. *Plants* (*Basel*). 2019;8(8):255. doi 10.3390/plants8080255

Nagy Z., Németh E., Guóth A., Bona L., Wodala B., Pécsváradi A. Metabolic indicators of drought stress tolerance in wheat: glutamine synthetase isoenzymes and Rubisco. *Plant Physiol. Biochem*. 2013;67:48-54. doi 10.1016/j.plaphy.2013.03.001

Ni Z., Hu Z., Jiang Q., Zhang H. *GmNFYA3*, a target gene of miR169, is a positive regulator of plant tolerance to drought stress. *Plant Mol. Biol*. 2013;82(1-2):113-129. doi 10.1007/s11103-013-0040-5

Niu C.F., Wei W., Zhou Q.Y., Tian A.G., Hao Y.J., Zhang W.K., Ma B., Lin Q., Zhang Z.B., Zhang J.S., Chen S.Y. Wheat *WRKY* genes *TaWRKY2* and *TaWRKY19* regulate abiotic stress tolerance in transgenic *Arabidopsis* plants. *Plant Cell Environ*. 2012;35(6):1156-1170. doi 10.1111/j.1365-3040.2012.02480.x

Pandian B.A., Sathishraj R., Djanaguiraman M., Prasad P.V.V., Jugulam M. Role of cytochrome P450 enzymes in plant stress response. *Antioxidants* (*Basel*). 2020;9(5):454. doi 10.3390/antiox9050454

Park C.Y., Lee J.H., Yoo J.H., Moon B.C., Choi M.S., Kang Y.H., Lee S.M., Kim H.S., Kang K.Y., Chung W.S., Lim C.O., Cho M.J. WRKY group IId transcription factors interact with calmodulin. *FEBS Lett*. 2005;579(6):1545-1550. doi 10.1016/j.febslet.2005.01.057

Park S.Y., Grabau E. Differential isoform expression and protein localization from alternatively spliced Apetala2 in peanut under drought stress. *J. Plant Physiol*. 2016;206:98-102. doi 10.1016/j.jplph.2016.09.007

Pastushkova L., Kashirina D.N., Brzhozovskiy A.G., Kononikhin A.S., Tiys E.S., Ivanisenko V.A., Koloteva M.I., Nikolaev E.N., Larina I.M. Evaluation of cardiovascular system state by urine proteome after manned space flight. *Acta Astronaut.* 2019;160:594-600. doi 10.1016/j.actaastro.2019.02.015

Pellegrineschi A., Reynolds M., Pacheco M., Brito R.M., Almeraya R., Yamaguchi-Shinozaki K., Hoisington D. Stress-induced expression in wheat of the *Arabidopsis thaliana* DREB1A gene delays water stress symptoms under greenhouse conditions. *Genome*. 2004;47(3):493-500. doi 10.1139/g03-140

Ranty B., Aldon D., Cotelle V., Galaud J.P., Thuleau P., Mazars C. Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. *Front. Plant Sci*. 2016;7:327. doi 10.3389/fpls.2016.00327

Raza A., Charagh S., Karikari B., Sharif R., Yadav V., Mubarik M.S., Habib M., Zhuang Y., Zhang C., Chen H., Varshney R.K., Zhuang W. miRNAs for crop improvement. *Plant Physiol. Biochem*. 2023;201:107857. doi 10.1016/j.plaphy.2023.107857

Ren L., Zhang T., Wu H., Ge X., Wan H., Chen S., Li Z., Ma D., Wang A. Blocking IbmiR319a impacts plant architecture and reduces drought tolerance in sweet potato. *Genes* (*Basel*). 2022;13(3):404. doi 10.3390/genes13030404

Rogachev A.D., Alemasov N.A., Ivanisenko V.A., Ivanisenko N.V., Gaisler E.V., Oleshko O.S., Cheresiz S.V., Mishinov S.V., Stupak V.V., Pokrovsky A.G. Correlation of metabolic profiles of plasma and cerebrospinal fluid of high-grade glioma patients. *Metabolites.* 2021;11(3):133. doi 10.3390/metabo11030133

Saik O.V., Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Interactome of the hepatitis C virus: Literature mining with ANDSystem. *Virus Res.* 2016;218:40-48. doi 10.1016/j.virusres.2015.12.003

Saik O.V., Demenkov P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Dosenko V.E., Zolotareva O.I., Choynzonov E.L., Hofestaedt R., Ivanisenko V.A. Search for new candidate genes involved in the comorbidity of asthma and hypertension based on automatic analysis

of scientific literature. *J. Integr. Bioinform.* 2018;15(4):20180054. doi 10.1515/jib-2018-0054

Saik O.V., Nimaev V.V., Usmonov D.B., Demenkov P.S., Ivanisenko T.V., Lavrik I.N., Ivanisenko V.A. Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema using an analysis of associative gene networks with ANDSystem. *BMC Med. Genomics*. 2019;12(Suppl. 2):47. doi 10.1186/s12920-019-0492-9

Saxena H., Negi H., Sharma B. Role of F-box E3-ubiquitin ligases in plant development and stress responses. *Plant Cell Rep*. 2023; 42(7):1133-1146. doi 10.1007/s00299-023-03023-8

Shamloo-Dashtpagerdi R., Shahriari A.G., Tahmasebi A., Vetukuri R.R. Potential role of the regulatory miR1119-MYC2 module in wheat (*Triticum aestivum* L.) drought tolerance. *Front. Plant Sci*. 2023;14: 1161245. doi 10.3389/fpls.2023.1161245

Shojaee S., Ravash R., Shiran B., Ebrahimie E. Meta-analysis highlights the key drought responsive genes in genes: *PEPC* and *TaSAG7* are hubs response networks. *J. Genet. Eng. Biotechnol.* 2022;20(1): 127. doi 10.1186/s43141-022-00395-4

Srivastava R., Kumar R. The expanding roles of APETALA2/Ethylene Responsive Factors and their potential applications in crop improvement. *Brief. Funct. Genomics*. 2018;18(4):240-254. doi 10.1093/bfgp/elz001

Szlachtowska Z., Rurek M. Plant dehydrins and dehydrin-like proteins: characterization and participation in abiotic stress response. *Front. Plant Sci*. 2023;14:1213188. doi 10.3389/fpls.2023.1213188

Tokizawa M., Enomoto T., Chandnani R., Mora-Macías J., Burbridge C., Armenta-Medina A., Kobayashi Y., Yamamoto Y.Y., Koyama H., Kochian L.V. The transcription factors, STOP1 and TCP20, are required for root system architecture alterations in response to nitrate deficiency. *Proc. Natl. Acad. Sci. USA*. 2023;120(35):e2300446120. doi 10.1073/pnas.2300446120

Virdi A.S., Pareek A., Singh P. Evidence for the possible involvement of calmodulin in regulation of steady state levels of Hsp90 family members (Hsp87 and Hsp85) in response to heat shock in sorghum. *Plant Signal. Behav*. 2011;6(3):393-399. doi 10.4161/psb.6.3.13867

Virdi A.S., Singh S., Singh P. Abiotic stress responses in plants: roles of calmodulin-regulated proteins. *Front. Plant Sci.* 2015;6:809. doi 10.3389/fpls.2015.00809

Volyanskaya A.R., Antropova E.A., Zubairova U.S., Demenkov P.S., Venzel A.S., Orlov Y.L., Makarova A.A., Ivanisenko T.V., Gorshkova T.A., Aglyamova A.R., Kolchanov N.A., Chen M., Ivanisenko V.A. Reconstruction and analysis of the gene regulatory network for cell wall function in *Arabidopsis thaliana* L. leaves in response to water deficit. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(8):1031-1041. doi 10.18699/VJGB-23-118

Wang H., Wang H. The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits. *Mol. Plant.* 2015;8(5):677-688. doi 10.1016/j.molp.2015.01.008

Wang M.Y., Zhao P.M., Cheng H.Q., Han L.B., Wu X.M., Gao P., Wang H.Y., Yang C.L., Zhong N.Q., Zuo J.R., Xia G.X. The cotton transcription factor TCP14 functions in auxin-mediated epidermal cell differentiation and elongation. *Plant Physiol.* 2013;162(3): 1669-1680. doi 10.1104/pp.113.215673

Wang S., Zhang M. Small-molecule ribonucleic acid (RNA) OsamiR393 for improving rice tillering and application. 2011. https://patents.google.com/patent/CN102533760A/en?oq=CN102533760A

Wang X., Niu Y., Zheng Y. Multiple functions of MYB transcription factors in abiotic stress responses. *Int. J. Mol. Sci*. 2021;22(11): 6125. doi 10.3390/ijms22116125

Wani S.H., Anand S., Singh B., Bohra A., Joshi R. WRKY transcription factors and plant defense responses: latest discoveries and future prospects. *Plant Cell Rep*. 2021;40(7):1071-1085. doi 10.1007/s00299-021-02691-8

Wei T., Guo D., Liu J. PtrMYB3, a R2R3-MYB transcription factor from *Poncirus trifoliata*, negatively regulates salt tolerance and hydrogen peroxide scavenging. *Antioxidants* (*Basel*). 2021;10(9): 1388. doi 10.3390/antiox10091388

Woodger F.J., Gubler F., Pogson B.J., Jacobsen J.V. A Mak-like kinase is a repressor of GAMYB in barley aleurone. *Plant J*. 2003;33(4): 707-717. doi 10.1046/j.1365-313x.2003.01663.x

Yang C., Li D., Mao D., Liu X., Ji C., Li X., Zhao X., Cheng Z., Chen C., Zhu L. Overexpression of microRNA319 impacts leaf morphogenesis and leads to enhanced cold tolerance in rice (*Oryza sativa* L.). *Plant Cell Environ*. 2013;36(12):2207-2218. doi 10.1111/pce.12130

Yang J., Zhang N., Mi X., Wu L., Ma R., Zhu X., Yao L., Jin X., Si H., Wang D. Identification of miR159s and their target genes and expression analysis under drought stress in potato. *Comput. Biol. Chem*. 2014;53(Part B):204-213. doi 10.1016/j.compbiolchem.2014.09.009

Yu T.F., Xu Z.S., Guo J.K., Wang Y.X., Abernathy B., Fu J.D., Chen X., Zhou Y.B., Chen M., Ye X.G., Ma Y.Z. Improved drought tolerance in wheat plants overexpressing a synthetic bacterial cold shock protein gene *SeCspA*. *Sci Rep*. 2017;7:44050. doi 10.1038/srep44050

Yu Y., Yu M., Zhang S., Song T., Zhang M., Zhou H., Wang Y., Xiang J., Zhang X. Transcriptomic identification of wheat AP2/ERF transcription factors and functional characterization of *TaERF-6-3A* in response to drought and salinity stresses. *Int. J. Mol. Sci*. 2022; 23(6):3272. doi 10.3390/ijms23063272

Zhang F., Yang J., Zhang N., Wu J., Si H. Roles of microRNAs in abiotic stress response and characteristics regulation of plant. *Front. Plant Sci*. 2022;13:919243. doi 10.3389/fpls.2022.919243

Zhou Y., Zhang L., Yu J. Application of tomato miR156e-3p gene in improvement of tomato low-temperature resistance and plant overexpression vector. 2020. https://patents.google.com/patent/CN111705077B/en?oq=CN111705077B

Zolotareva O., Saik O.V., Königs C., Bragina E.Y., Goncharova I.A., Freidin M.B., Dosenko V.E., Ivanisenko V.A., Hofestädt R. Comorbidity of asthma and hypertension may be mediated by shared genetic dysregulation and drug side effects. *Sci. Rep*. 2019;9(1):16302. doi 10.1038/s41598-019-52762-w

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 22.09.2024. После доработки 31.10.2024. Принята к публикации 02.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):918-926

doi 10.18699/vjgb-24-99

Пограничные клетки корневого чехлика как регулятор ризосферной микробиоты

Н.А. Омельянчук[1], В.А. Черенко[1, 2], Е.В. Землянская[1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

ezemlyanskaya@bionet.nsc.ru

Аннотация. Ризосфера (почва, окружающая корни растения) – это экологическая ниша, внутри которой полезные микроорганизмы и патогены конкурируют друг с другом за органические углеродные соединения и возможность колонизации корней. Для взаимодействия с микробиотой корни выделяют в ризосферу ризодепозиты, к которым относят пограничные клетки, продукты гибели клеток корня и секретируемые живыми клетками жидкости (корневые экссудаты). Пограничные клетки, получившие свое название ввиду их локализации в почве рядом с корнем (на границе корня и почвы), представляют собой конечный этап дифференцировки клеток корневого чехлика. Слущивание пограничных клеток с поверхности корневого чехлика может происходить как одиночными клетками, так и рядами клеток. Пограничные клетки постоянно поставляются в почву на протяжении всей жизни растения, а тип и интенсивность слущивания пограничных клеток определяются как видом растений, так и почвенными условиями. В настоящее время появились данные о факторах, контролирующих тип слущивания, а также исследования этого процесса и его регуляции у разных видов растений. Пограничные клетки специализированы для взаимодействия с внешней средой, в частности, они служат живым барьером между корнем и почвенной микробиотой. После отделения от кончика корня в пограничных клетках снижается уровень первичного метаболизма и повышается число транскриптов генов вторичного метаболизма, усиливаются синтез компонентов и выделение слизи, содержащей вторичные метаболиты, внеклеточную ДНК, протеогликаны и другие вещества. Слизь, в которую пограничные клетки оказываются погруженными, служит как для привлечения микроорганизмов, способствующих росту растения, так и для защиты корня от патогенов. В настоящем обзоре описаны взаимодействия пограничных клеток с различными видами микроорганизмов и продемонстрирована их важность для роста растений и их устойчивости к болезням. Эти аспекты могут быть использованы в генной инженерии и селекции для усиления полезных функций пограничных клеток, что, в свою очередь, откроет новые горизонты для повышения урожайности и устойчивости сельскохозяйственных культур.
Ключевые слова: корень; пограничные клетки; биотический стресс; защита растений от патогенов; почвенные симбионты.

Для цитирования: Омельянчук Н.А., Черенко В.А., Землянская Е.В. Пограничные клетки корневого чехлика как регулятор ризосферной микробиоты. *Вавиловский журнал генетики и селекции*. 2024;28(8):918-926. doi 10.18699/vjgb-24-99

Финансирование. Работа выполнена в рамках бюджетного проекта FWNR-2022-0005.

Root cap border cells as regulators of rhizosphere microbiota

N.A. Omelyanchuk[1], V.A. Cherenko[1, 2], E.V. Zemlyanskaya[1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia

ezemlyanskaya@bionet.nsc.ru

Abstract. A rhizosphere (a narrow area of soil around plant roots) is an ecological niche, within which beneficial microorganisms and pathogens compete with each other for organic carbon compounds and for the opportunity to colonize roots. The roots secrete rhizodeposits into the rhizosphere, which include border cells, products of root cell death and liquids secreted by living cells (root exudates). Border cells, which have their name due to their location in the soil next to the root (at the border of the root and soil), represent terminal differentiation of columella and adjacent lateral root cap cells. Border cells can detach from the root cap surface both as single cells and as cell layers. Border cells are constantly supplied to the soil throughout plant life, and the type and intensity of border cells' sloughing depend on both plant species and soil conditions. Currently, data on the factors that control the type of border cells' release and its regulation have been described in different plant species. Border cells are specialized for interaction with the environment, in particular, they are a living barrier between soil microbiota and roots. After separation of border cells from the root tip, transcription of primary metabolism genes decreases, whereas transcription of secondary metabo-

© Омельянчук Н.А., Черенко В.А., Землянская Е.В., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

lism genes as well as the synthesis and secretion of mucilage containing these metabolites along with extracellular DNA, proteoglycans and other substances increase. The mucilage that the border cells are embedded in serves both to attract microorganisms promoting plant growth and to protect plants from pathogens. In this review, we describe interactions of border cells with various types of microorganisms and demonstrate their importance for plant growth and disease resistance.

Key words: root; border cells; biotic stress; plant defense against pathogens; soil symbionts.

For citation: Omelyanchuk N.A., Cherenko V.A., Zemlyanskaya E.V. Root cap border cells as regulators of rhizosphere microbiota. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):918-926. doi 10.18699/vjgb-24-99

Введение

Корни растений окружены большим количеством микроорганизмов: в ризосфере (почве, непосредственно контактирующей с корневой системой растения) на 1 г почвы содержится ~10^8–10^9 бактерий, 10^5–10^6 грибов, 10^3–10^5 водорослей и простейших (Mendes et al., 2013). Эта метаболически активная микробиота изменяет свойства почвы и влияет на рост корня и всего растения. Корневая система растения, в свою очередь, проникая глубоко в почву, изменяет ее, обогащая ризодепозитами, живыми и мертвыми клетками и различными органическими соединениями, влияющими на состав и численность микробных популяций. Большую часть ризодепозита составляют регулярно слущивающиеся в почву клетки поверхности корневого чехлика, миниатюрного органа, расположенного на самом кончике корня (Hawes et al., 2011). Слущенные клетки корневого чехлика называются пограничными клетками из-за их расположения на границе корня и почвы (Hawes, Lin, 1990). Это живые клетки, секретирующие полисахариды, белки и ряд других веществ (Driouich et al., 2021). Из этого секрета формируется матрикс, в который пограничные клетки оказываются погруженными. Пограничные клетки по мере роста корня вступают во взаимодействие и с расположенными выше корневого чехлика клетками, их можно найти на довольно большом расстоянии от кончика корня, из которого они произошли (Hawes, Lin, 1990; Driouich et al., 2019).

Пограничные клетки описаны у папоротников, голосеменных и покрытосеменных растений (Vermeer, McCully, 1982; Hawes et al., 2003; Forino et al., 2012). Число живых пограничных клеток на корень зависит от семейства растений, а также меняется по мере роста корня, варьируя в молодых (до 2 см) корнях от 800 у *Bromus carinatus* и 11 000 у *Cucumis sativus* до 17 000 у *Zea mays* со значимым уменьшением в корнях с длиной более 9 см до 70, 300 и 150 клеток соответственно (Odell et al., 2008; Darshan et al., 2020). Число пограничных клеток может различаться даже у разных экотипов одного вида и зависит от условий выращивания (Zhao et al., 2000; Iijima et al., 2003; Pankievicz et al., 2022). Например, когда растения гороха помещают в условия с высоким уровнем углекислого газа, пограничных клеток производится в два раза больше по сравнению с нормальными условиями (Zhao et al., 2000).

Пограничные клетки являются «возобновляемыми», т. е. постоянно поставляемыми в почву на протяжении всей жизни корня и имеющими определенный срок жизни (Driouich et al., 2019). Например, корневая система одного растения гороха продуцирует примерно 3000–4000 пограничных клеток в день. Время, в течение которого пограничные клетки остаются живыми после слущивания с поверхности корневого чехлика, зависит от вида растений и составляет от нескольких дней у арабидопсиса (Vicré et al., 2005; Plancot et al., 2013) до нескольких недель у кукурузы (Vermeer, McCully, 1982). Во многих семействах покрытосеменных растений (Злаковых, Бобовых, Тыквенных) клетки крайнего слоя корневого чехлика отделяются без сохранения контактов друг с другом, т. е. в почве оказываются отдельные жизнеспособные пограничные клетки (Driouich et al., 2007). У других семейств, например у Крестоцветных (в том числе и у модельного вида растений *Arabidopsis thaliana* L.), наблюдается слущивание живых клеток единым слоем (Fendrych et al., 2014). Некоторые исследователи выделяют такие клетки в отдельную группу и называют их «клетки, подобные пограничным» (Vicré et al., 2005; Driouich et al., 2007; Plancot et al., 2013). Также был предложен альтернативный термин, объединяющий пограничные клетки и «клетки, подобные пограничным», – «ассоциированные с корнем клетки, произошедшие из чехлика» (root AC-DC – root associated cap-derived cells) (Driouich et al., 2019).

В настоящее время появились данные о факторах, контролирующих тип слущивания наружных клеток корневого чехлика, исследования этого процесса и его регуляции у разных видов растений, описания функционирования пограничных клеток в почве и их важной роли в жизни растений. Согласно этим данным, пограничные клетки можно определить как живые клетки, слущиваемые корневым чехликом во внешнюю среду как отдельные клетки, слоем клеток или агрегатами в несколько слоев для выполнения специальных функций по поддержанию роста растений и их защите от болезней (Darshan et al., 2020). В связи с этим далее мы будем пользоваться общим термином «пограничные клетки», независимо от типа их слущивания.

В этом обзоре мы подробно разбираем факторы, определяющие тип слущивания пограничных клеток, описываем отличия пограничных клеток от других клеток кончика корня, их секреторную функцию и формирование микробиоты ризосферы под влиянием секрета пограничных клеток.

Дифференцировка пограничных клеток и типы их слущивания у разных видов растений

У *A. thaliana* корневой чехлик состоит из двух отдельных частей: центрально расположенной колумеллы и бокового корневого чехлика (БКЧ), который окружает колумеллу и расположенные выше клетки меристемы корня (Dolan et al., 1993). У клеток наружного слоя БКЧ происходит программируемая клеточная смерть в транзитной зоне корня с

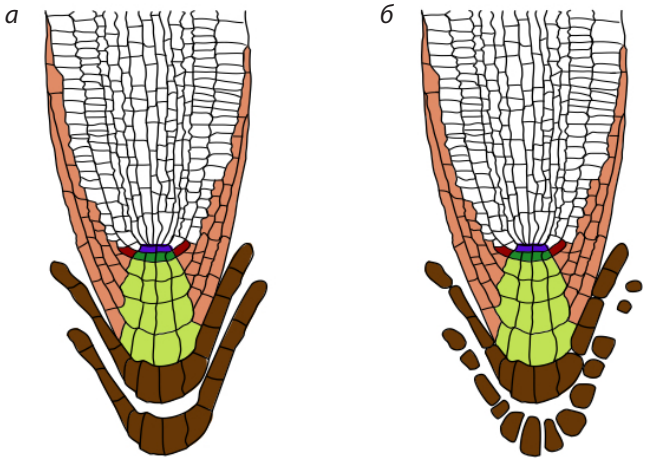


Рис. 1. Слущивание пограничных клеток единым слоем (*а*) и отдельными клетками (*б*) у проростков *A. thaliana*.

а – кончик корня проростка дикого типа; *б* – кончик корня мутанта *nlp7*. Синим цветом обозначены клетки покоящегося центра, темно-зеленым – инициали колумеллы, светло-зеленым – колумелла, светло-коричневым – боковой корневой чехлик, красным – инициали эпиблемы и БКЧ, темно-коричневым – пограничные клетки. Схематичные изображения сделаны в соответствии с данными R. Karve с коллегами (2016).

последующим быстрым автолизом, в то время как клетки наружного слоя колумеллы слущиваются единым слоем с сохранением контактов между ними (Vicré et al., 2005; Durand et al., 2009; Fendrych et al., 2014). На первом этапе слущивания появляется разрыв в наружном слое БКЧ на уровне чуть выше покоящегося центра с последующим отделением клеток этого слоя сверху вниз, за которым следует слущивание внешнего слоя клеток колумеллы (рис. 1, *а*) (Shi et al., 2018). Весь процесс, от возникновения разрыва и до конца слущивания, занимает в среднем 18 ч, и примерно 18 ч проходит до начала слущивания следующего клеточного слоя. Необходимо отметить, что отделяющиеся от корня клетки корневого чехлика отвечают первоначальному определению пограничных клеток – находятся на границе между корнем и почвой (Hawes, Lin, 1990). Более того, у *A. thaliana* до 12 % корней проростков экотипа Columbia продуцируют отдельные, изолированные друг от друга пограничные клетки (Karve et al., 2016).

Основными компонентами срединных пластинок – частей клеточных стенок, благодаря которым соседние клетки «склеены» друг с другом, являются пектины (полигалактуронаны, представленные гомогалактуронанами, рамногалактуронанами и замещенными галактуронанами) (Caffall, Mohnen, 2009; Albersheim et al., 2010). Пектины синтезируются в клетке, затем секретируются в клеточную стенку преимущественно в метилэтерифицированной форме (Atmodjo et al., 2013). В клеточной стенке ферменты пектинметилэстеразы отщепляют метильные группы, образуя свободные карбоксильные группы на остатках галактуроновой кислоты полигалактуронанов, что локально сдвигает pH в кислую сторону и способствует активности полигалактуроназ, гидролизующих полигалактуронаны (Moustacas et al., 1991; Micheli, 2001). Этим объясняется разделение слоя пограничных клеток на отдельные клетки у *A. thaliana* при повышенной кислотности среды, в которой растут корни (Karve et al., 2016). Роль пектинов в слущивании пограничных клеток установлена у гороха (Wen et al., 1999). При ингибировании экспрессии гена, кодирующего пектинметилэстеразу, пограничные клетки не отделяются от корня. У мутантов *A. thaliana* по генам *QUASIMODO 1/2* с низкой продукцией одного из пектинов – гомогалактуронана (линейного полимера галактуроновой кислоты) – клетки корневого чехлика слущиваются отделенными друг от друга (Durand et al., 2009).

У арабидопсиса выявлен также транскрипционный фактор NIN-LIKE PROTEIN7 (NLP7), контролирующий слущивание пограничных клеток целым слоем (Karve et al., 2016). При мутации *nlp7* увеличивается раздельное слущивание пограничных клеток (см. рис. 1, *б*). В то время как в диком типе только в 12 % корней наблюдалось слущивание отдельных пограничных клеток, в мутантах это было зарегистрировано в 44 % корней. В мутантах с потерей функции *nlp7* также снижен уровень целлюлозы и пектина и активированы гены, кодирующие целлюлазу (CEL5) и пектинлиазы – ферменты, ослабляющие клеточную стенку. У арабидопсиса при отсутствии экспрессии гена *CEL5* скорость слущивания пограничных клеток замедлялась (Del Campillo et al., 2004). Также отдельными клетками происходит слущивание пограничных клеток при потере функции *AUTOPHAGY 5* (*ATG5*), одного из ключевых генов, контролирующих аутофагию (Goh et al., 2022). У *atg5* мутантов в пограничных клетках не происходит образования аутофагосом и центральной вакуоли.

В отношении типов слущивания пограничных клеток существует большое разнообразие. Так, у акации *Acacia mangium*, тропического дерева семейства Бобовых, пограничные клетки БКЧ слущиваются единым листом, состоящим из нескольких рядов клеток, от транзитной зоны корня к его апексу, при этом клетки колумеллы слущиваются отдельными пограничными клетками (Endo et al., 2011). Из трех древесных видов Бобовых, растущих к югу от Сахары, у *Balanites aegyptiaca* клетки корневого чехлика слущиваются отдельно, в то время как у *Acacia raddiana* и *Tamarindus indica* – и отдельными клетками, и рядами (Carreras et al., 2020). У *Pinus densiflora* с центральной части корневого чехлика слущиваются одиночные продолговатые пограничные клетки, а с боковых сторон слущивание идет длинными одиночными слоями (Shirakawa et al., 2023).

У сои выделено три морфотипа пограничных клеток: сферические, промежуточные и удлиненные (Ropitaux et al., 2020). Сферические пограничные клетки локализованы преимущественно рядом с корневым чехликом, промежуточные окружают корень в области его меристематической зоны, а удлиненные – в зонах растяжения и дифференцировки (рис. 2). Удлиненные клетки составляют более 30 % пограничных клеток и могут быть как единичными, так и в составе группы из десятка или нескольких десятков плотно прикрепленных друг к другу клеток. Примерно 80 % удлиненных и 50 % сферических пограничных клеток – это живые клетки. У кукурузы сферические клетки отделяются от колумеллы, в то время как БКЧ производит удлиненные клетки (Guinel, McCully, 1987). У банана удлиненные (эллипсоидные) клетки со-

ставляют 92 % пограничных клеток, оставшиеся 8 % – это сферические клетки с амилопластами (Wuyts et al., 2006). У картофеля сферические пограничные клетки малого размера наблюдались в области корневого чехлика, в то время как удлиненные клетки в основном были локализованы в зоне растяжения (Koroney et al., 2016). Клетки обоих типов содержали крахмал.

Таким образом, наружные клетки корневого чехлика могут удаляться с поверхности корневого чехлика после программируемой клеточной смерти с последующим быстрым автолизом, а также слущиваться отдельными живыми клетками или рядами живых клеток. В последующем гибель пограничных клеток, слущенных живыми, в почве дает клеточные останки, которые, как и другие мертвые клетки, служат для питания микробиоты. В пограничных клетках, по сравнению с клетками кончика корня, снижен уровень первичного метаболизма и повышен уровень транскриптов генов вторичного метаболизма, кодирующих белки для синтеза воска, фенилпропаноидов, лигнина, фенольных соединений и флавоноидов (Watson et al., 2015).

Источником энергии и углерода для синтеза вторичных метаболитов в пограничных клетках служат большие запасы крахмала. Пограничные клетки также синтезируют уникальный набор белков: 13 % белков, продуцируемых в пограничных клетках, не детектируются в кончике корня (Brigham et al., 1995). Таким образом, пограничные клетки представляют собой последний этап дифференцировки клеток корневого чехлика. Из вышесказанного понятно, что дифференцировка и слущивание пограничных клеток – достаточно энергозатратный процесс. Тогда возникает вопрос, для каких значимых целей растения регулируемым образом высвобождают большое количество здоровых клеток с периферии корневого чехлика? Несомненно, это предполагает их ключевую функцию во взаимоотношениях со средой, окружающей корень.

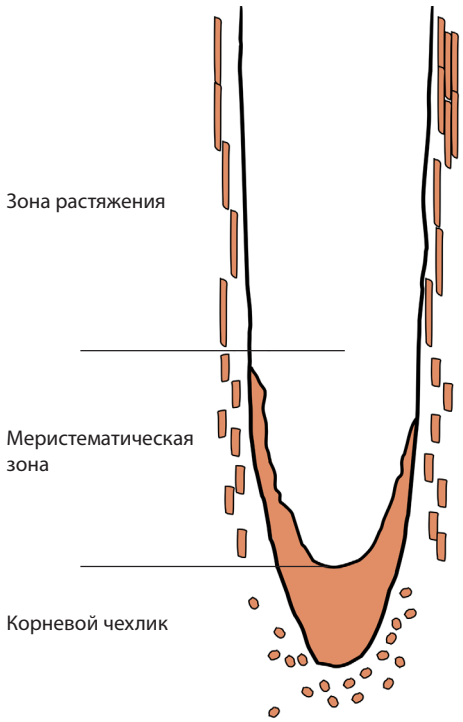


Рис. 2. Три морфотипа пограничных клеток у сои.

Корневой чехлик, а также сферические, промежуточные и удлиненные пограничные клетки обозначены коричневым цветом. Для подготовки схемы использованы результаты M. Ropitaux с коллегами (2020).

Состав и функции слизи, секретируемой пограничными клетками

Процесс приобретения предшественниками пограничных клеток способности секретировать подробно описан для клеток колумеллы арабидопсиса (Maeda et al., 2019). Если инициали колумеллы считать слоем c1, то при переходе клеток из c5 в c6 слизь начинает накапливаться у боковых клеточных стенок, а клеточная стенка на стороне, обращенной к побегу, начинает деградировать (рис. 3). В клетках c7 большая часть слизи выделяется в межклеточное пространство между слоями c6 и c7. Параллельно с этим идут развитие вакуолей и деградация амилопластов. После отделения пограничных клеток слизь из межклеточного пространства переходит в ризосферу, а пограничные клетки продолжают секрецию. Таким образом, пограничные клетки оказываются погруженными в плотную слизь с фибриллярной структурой (Ropitaux et al., 2020). Аппарат Гольджи, необходимый для секреции, формируется в периферических клетках колумеллы до их отделения и превращения в пограничные клетки (Poulsen et al., 2008). Пузырьки аппарата Гольджи, в том числе сливающиеся с плазматической мембраной, характерны для пограничных клеток (Driouich et al., 2007; Wang et al., 2017). У сои наибольшее количество слизи производят сферические пограничные клетки, наименьшее – удлиненные (Ropitaux et al., 2020).

У большинства видов растений около 94 % растворимой фракции слизи составляют нейтральные и кислые полисахариды, а остальные 6 % – белки (Carminati, Vetterlein, 2013); 25 % белков, синтезируемых пограничными клетками, сразу выделяются в окружающую среду (Brigham et al., 1995). Большинство метаболитов, вырабатываемых в пограничных клетках, также секретируются сразу после синтеза. В слизи корневого чехлика 3–4-дневных проростков кукурузы было обнаружено 2848 различных белков, среди которых существенную часть (25 %) представляли белки, задействованные в метаболизме, остальные белки функционально

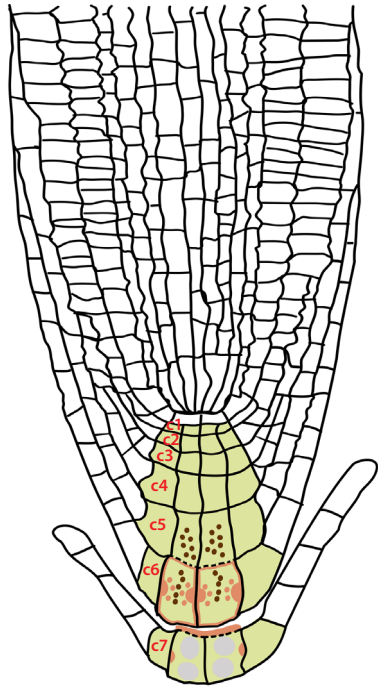


Рис. 3. Дифференцировка пограничных клеток в колумелле у *A. thaliana*.

Клетки колумеллы обозначены светло-зеленым цветом. Клеточные слои колумеллы пронумерованы по порядку, от c1 (инициали колумеллы) до c7. Коричневыми точками обозначены крахмальные гранулы, красным цветом – слизь, серым – вакуоли.

имели отношение к клеточной стенке, активным формам кислорода, получению питательных веществ и ответу на стресс (Ma et al., 2010). Для 85–94 % белков слизи, ранее найденных у арабидопсиса и рапса, выявлены гомологи, представленные в слизи у кукурузы, что говорит о достаточной консервативности белкового состава слизи у одно- и двудольных растений.

Кислые (пектиновые) полисахариды придают желеобразные свойства слизи, т. е. делают ее гелем с пористой структурой. Выделяемая пограничными клетками слизь может удерживать воду в количестве в 1000 раз больше своего веса (Guinel, McCully, 1986). У сои основной компонент фибриллярной структуры в слизи – нейтральный полисахарид ксилоглюкан (Ropitaux et al., 2019). Ксилоглюкан и целлюлоза образуют молекулярные поперечные мостики, которые держат пограничные клетки в контакте друг с другом. Известно, что первичная клеточная стенка двудольных растений представляет собой целлюлозные и ксилоглюкановые полисахариды, встроенные в матрикс из пектинов, гликопротеинов и протеогликанов (Driouich et al., 2012). Таким образом, пограничные клетки секретируют полисахариды и протеогликаны клеточной стенки, из которых формируются матрикс и внутренняя структура слизи (Castilleux et al., 2018; Driouich et al., 2019).

Среди белковых компонентов секрета пограничных клеток выделяют гидроксипролин-богатые гликопротеины – экстенсины и арабиногалактановые белки (Vicré et al., 2005; Plancot et al., 2013). Арабиногалактановые белки были найдены в слизи пограничных клеток гороха, арабидопсиса, рапса и картофеля (Knee et al., 2001; Durand et al., 2009; Cannesan et al., 2012; Koroney et al., 2016). Слизь также содержит фенольные кислоты, фосфолипиды, антимикробные пептиды/белки (дефенсины, белки, связанные с патогенезом и др.), фитоалексины, гистон H4, ферменты, внеклеточную ДНК, активные формы кислорода, токсичные для патогенов, и ферменты, их производящие (Wen et al., 2007, 2017; Carminati, Vetterlein, 2013; Plancot et al., 2013; Weiller et al., 2017).

Выделяемая пограничными клетками слизь и сами пограничные клетки образуют комплекс, называемый «корневая внеклеточная ловушка, КВЛ» (root extracellular trap, RET) (Driouich et al., 2013). Он имеет много общего с внеклеточными ловушками животных, образуемыми фагоцитарными иммунными клетками (нейтрофилами, макрофагами, тучными клетками, эозинофилами, гетерофилами) при стимуляции (Driouich et al., 2019, 2021). И у растений, и у животных внеклеточные ловушки неспецифически активны против широкого спектра микробных и грибковых патогенов. Ловушки содержат одинаковые защитные компоненты (антимикробные белки и внеклеточную ДНК) и имеют одинаковые функции – улавливать, иммобилизовывать и уничтожать патогены, тем самым ограничивая распространение микробов на другие ткани.

Механизм действия внеклеточной ДНК, секретируемой пограничными клетками, неясен (Monticolo et al., 2020). Тем не менее деградация внеклеточной ДНК в секрете пограничных клеток с помощью ДНКазы приводила к потере устойчивости корней к патогенным грибам (Wen et al., 2009). Мутации генов, кодирующих секретируемые ДНКазы у фитопатогенных бактерий и грибов, обуславливали снижение заражаемости этими патогенами корней растений (Hawes et al., 2016; Tran et al., 2016). Секреция ДНКаз описана для многих почвенных патогенных видов грибов и некоторых видов бактерий (Darshan et al., 2020). Пограничные клетки гороха и томата секретируют внеклеточную ДНК в ответ на патогенные бактерии, тогда как непатогенные бактерии не вызывают секрецию ДНК (Tran et al., 2016).

Гистон H4 человека, на 97 % гомологичный гистону H4 гороха, секретируемому пограничными клетками, летален для бактерии *Ralstonia solanacearum*, инфицирующей корни гороха, и его токсическая активность устраняется при обработке корней антителами против него (Tran et al., 2016).

Пограничные клетки формируют микробиоту в ризосфере

Пограничные клетки могут защищать растения и способствовать их росту, препятствуя инфицированию корня патогенами или стимулируя ассоциацию с полезной микробиотой. Совместное культивирование на поверхности агара пограничных клеток в окружающей их слизи с разными видами бактерий показало различную реакцию бактерий на пограничные клетки и их секрет (Gochnauer et al., 1990). Наблюдались как сильное торможение роста (*Rhizobium* sp. и *Escherichia coli*), так и сильная стимуляция (*Pseudomonas fluorescens*), а также отсутствие эффекта (*Streptomyces* sp. и *Cytophaga* sp.) или раннее ингибирование с последующей сильной стимуляцией, а затем образованием спор (*Bacillus* spp.).

Таким образом, состав бактериального сообщества в ризосфере определяется способностью разных видов бактерий реагировать на вещества, составляющие секрет пограничных клеток. Можно предположить, что за счет этого пограничные клетки активно контролируют формирование и другой части ризосферной микробиоты (грибов, протистов и др.). Секрет пограничных клеток также обеспечивает различные источники углерода для микроорганизмов, тем самым влияя на состав микробиома (Knee et al., 2001; Benizri et al., 2007).

Полезные для растений ризосферные бактерии выделяют в особую группу – ризобактерии, способствующие росту растений (PCPP) (Hasan et al., 2024). Они разнообразны по видовому составу, к ним относятся представители *Agrobacterium*, *Arthrobacter*, *Azotobacter*, *Azospirillum*, *Burkholderia*, *Caulobacter*, *Chromobacterium*, *Erwinia*, *Flavobacterium*, *Micrococcous*, *Pseudomonas*, *Rhizobium*, *Serratia* и других родов. Взаимодействуя с корнем, они повышают устойчивость растений к биотическому и абиотическому стрессу, доступность различных элементов (железа, калия, фосфора и др.) из почвы, синтезируют фитогормоны и другие метаболиты, влияющие на рост растений, способствуют очищению почвы от многих вредных примесей. Многие PCPP подавляют рост патогенных организмов, продуцируя антибиотики (Ulloa-Ogaz et al., 2015).

Актиномицеты не только способствуют росту растений сами по себе, некоторые их изоляты усиливают рост грибов арбускулярной микоризы и прорастание их спор, что приумножает эффект актиномицетов на рост рас-

тений и усвоение ими азота (Franco-Correa et al., 2010). Другие изоляты актиномицетов показали сильную активность против грибов-возбудителей болезней растений (Lee, Hwang, 2002). Бактерия *Herbaspirillum seropedicae* образует азотфиксирующие ассоциации с кукурузой и другими злаками (Chubatsu et al., 2012). Примечательно, что гуминовые кислоты увеличивают слущивание пограничных клеток и плотность этих бактерий в области кончика корня (Canellas, Olivares, 2017).

Живые пограничные клетки являются основными производителями слизи, которая содержит вещества, способствующие привлечению благоприятных для растения микроорганизмов (Hawes et al., 1998). Пограничные клетки выделяют соединения, стимулирующие ветвление микоризных гиф, и арабиногалактаны, запускающие образование биопленок ряда полезных бактерий (Nagahashi, Douds, 2004; Beauregard et al., 2013). Разрушение арабиногалактановых белков с помощью специальных агентов приводит к ослаблению колонизации бактериями рода *Rhizobium* пограничных клеток и кончика корня (Vicré et al., 2005). У *Pinus densiflora* на ранних стадиях развития корней (до образования микоризы) ризобактерии, контактирующие с пограничными клетками и их секретом, способствуют защите корней хозяина, оказывая замедляющее воздействие на рост патогенов (Shirakawa et al., 2023).

Повсеместно распространенные почвенные грибы *Arbuscular mycorrhizae* формируют микоризу со многими видами покрытосеменных растений, включая большинство сельскохозяйственных культур (Khaliq et al., 2022). Микориза улучшает поглощение растениями воды и питательных веществ, особенно фосфора, в то время как растения отдают грибам 10–20 % продуктов своего фотосинтеза. При этом количество пограничных клеток, продуцируемых разными видами растений, положительно коррелирует со способностью этих видов образовывать микоризные ассоциации (Niemira et al., 1996; Arriola et al., 1997). Один из штаммов гриба-аскомицета *Trichoderma* при колонизации пограничных клеток проростков пшеницы вызывал примерно 40 % увеличение веса стебля и подавлял более чем на 90 % рост фитопатогенных видов рода *Fusarium* (Jaroszuk-Ściseł et al., 2019).

В настоящее время можно говорить о том, что формируется новое направление в сельскохозяйственной биотехнологии – биоинженерия микробиома ризосферы, которое нацелено на заселение ризосферы предпочтительно полезными для растений микроорганизмами (Mohanram, Kumar, 2019). Например, такие роды бактерий, как *Bacillus* и *Pseudomonas*, используются в качестве биоудобрений и для биологической защиты растений (для производства биопрепаратов против патогенов, в качестве их естественных врагов или индукторов системной устойчивости к ним растений) (Hasan et al., 2024). Еще одним перспективным подходом для инженерии микробиома ризосферы является модификация пограничных клеток (Mohanram, Kumar, 2019). Успешность такого подхода была продемонстрирована при трансформации растений арабидопсиса и картофеля геном, кодирующим пептидный противонематодозный репеллент, под промотором гена арабидопсиса *MDK4-20* (Lilley et al., 2011). Последний, специфически активируя транскрипцию в клетках корневого чехлика и пограничных клетках, вызвал в них экспрессию гена, кодирующего репеллент, и тем самым обеспечил устойчивость трансгенных растений к нематоде.

Пограничные клетки взаимодействуют с почвенными патогенами

Отделение от корня пограничных клеток и выделение ими в почву различных соединений представляют один из механизмов, с помощью которых растения справляются с патогенами (Hawes et al., 2000). Выше мы уже упоминали об антимикробных функциях слизи, которые осуществляют некоторые белки, вторичные метаболиты и внеклеточная ДНК, работающая для защиты от некоторых грибов и бактерий (Wen et al., 2009; Cannesan et al., 2011; Koroney et al., 2016; Tran et al., 2016). Однако взаимодействие пограничных клеток с патогенами не ограничивается бактерицидными и фунгицидными свойствами секретируемой ими слизи. Пограничные клетки способны воспринимать специфические сигналы патогенов, так называемые связанные с патогенами молекулярные паттерны (molecular/pathogen associated molecular patterns, MAMP/PAMP), и реагировать на них типичными MAMP-индуцируемыми процессами первичного иммунного ответа, продукцией активных форм кислорода и укреплением клеточной стенки путем накопления и модификации экстенсинов и отложением каллозы (Plancot et al., 2013).

Атака патогена может привести к усилению слущивания пограничных клеток, стимулированию производства слизи этими клетками или изменению состава слизи (Cannesan et al., 2011; Koroney et al., 2016). Например, обработка корней элиситором, полученным из *Pectobacterium atrosepticum*, почвенного патогена картофеля, изменяет состав слизи, включая состав арабиногалактановых белков в ней (Koroney et al., 2016). Оомицет *Aphanomyces euteiches* вызывает до 80 % потерь урожая гороха, проникая в его корни, что приводит к остановке роста корней и гибели растения (Cannesan et al., 2011). Инокуляция *A. euteiches* корней гороха увеличивает количество пограничных клеток, и это увеличение зависит от количества инокулированных ооспор. Пограничные клетки реагируют на инокуляцию увеличением синтеза фенольного соединения фитоалексина пизатина, которое *in vitro* при определенном количестве ингибирует рост гиф и производство зооспор.

Таким образом, увеличение синтеза пизатина пограничными клетками может служить одним из способов повышения устойчивости корней гороха к данной инфекции. Более того, пограничные клетки привлекают оомицет посредством хемотаксиса и затем нейтрализуют их с помощью антимикробных компонентов слизи (Hawes et al., 2016). В частности, было показано, что арабиногалактановые белки, компоненты слизи и клеточных стенок пограничных клеток, способны вызывать инцистирование и предотвращать прорастание зооспор этого патогена (Cannesan et al., 2012). Таким образом, пограничные клетки и их секрет предотвращают заселение зооспорами кончиков корней, блокируя их проникновение в ткани корня и вызывая их лизис (Ropitaux et al., 2020).

Пограничные клетки проростков ржи нейтрализуют действие патогенного штамма гриба *Fusarium culmorum*, стимулируя прорастание спор в макроконидии и образуя с последними вокруг корневого чехлика компактные кластеры, так называемые мантиеподобные структуры, в то время как непатогенные штаммы таких структур не образуют (Jaroszuk-Ściseł et al., 2009). В дополнение к общеизвестным механизмам подавления грибной инфекции – ингибированию прорастания спор, подавлению активности генов патогенеза гриба, усилению экспрессии генов защиты растений – образование мантиеподобных структур на кончике корня представляет собой еще одну форму взаимоотношений корень–патоген, когда секрет пограничных клеток вызывает, наоборот, быстрое прорастание спор с последующей гибелью пограничных клеток и подавлением роста гриба (Gunawardena et al., 2005).

Образование мантиеподобных структур на кончике корня наблюдалось также при инокуляции корней гороха с патогенным грибом *Nectria haematococca*, при этом под мантиеподобной структурой сами кончики корня большей частью остаются интактными (Gunawardena, Hawes, 2002). При этой инфекции повреждается только около 4 % кончиков корней, в случае же протеолитической деградации секрета пограничных клеток повреждаются все кончики корней (Wen et al., 2007).

Заключение

Пограничные клетки представляют собой жизнеспособные компоненты корневой системы, играющие ключевую роль во взаимодействии корней с микроорганизмами ризосферы. После отделения от кончика корня пограничные клетки меняют свой метаболизм, синтезируют и выделяют гидратированную слизь, содержащую протеогликаны, вторичные метаболиты, антимикробные белки и внеклеточную ДНК. Эта слизь служит активным компонентом для привлечения полезных микроорганизмов, усиливающих рост растений. В то же время пограничные клетки служат барьером для патогенов. Они секретируют различные антимикробные вещества, и в них индуцируется первичный иммунный ответ различными элиситорами. Таким образом, усиление полезных для растения функций пограничных клеток может стать одним из актуальных и эффективных направлений в генной инженерии и селекции растений.

Список литературы / References

Albersheim P., Darvill A., Roberts K., Sederoff R., Staehelin A. Plant Cell Walls. From Chemistry to Biology. New York: Garland Science, 2010

Arriola L., Niemira B.A., Safir G.R. Border cells and arbuscular mycorrhizae in four Amaranthaceae species. *Phytopathology*. 1997; 87(12):1240-1242. doi 10.1094/PHYTO.1997.87.12.1240

Atmodjo M.A., Hao Z., Mohnen D. Evolving views of pectin biosynthesis. *Annu. Rev. Plant Biol.* 2013;64:747-779. doi 10.1146/annurev-arplant-042811-105534

Beauregard P.B., Chai Y., Vlamakis H., Losick R., Kolter R. *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proc. Natl. Acad. Sci. USA*. 2013;110(17):1621-1630. doi 10.1073/pnas.1218984110

Benizri E., Nguyen C., Piutti S., Slezack-Deschaumes S., Philippot L. Additions of maize root mucilage to soil changed the structure of the bacterial community. *Soil Biol. Biochem*. 2007;39(5):1230-1233. doi 10.1016/j.soilbio.2006.12.026

Brigham L.A., Woo H.H., Nicoll S.M., Hawes M.C. Differential expression of proteins and mRNAs from border cells and root tips of pea. *Plant Physiol*. 1995;109(2):457-463. doi 10.1104/pp.109.2.457

Caffall K.H., Mohnen D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr. Res*. 2009;344: 1879-1900. doi 10.1016/j.carres.2009.05.021

Canellas L.P., Olivares F.L. Production of border cells and colonization of maize root tips by *Herbaspirillum seropedicae* are modulated by humic acid. *Plant Soil*. 2017;417:403-413. doi 10.1007/s11104-017-3267-0

Cannesan M.A., Gangneux C., Lanoue A., Giron D., Laval K., Hawes M., Driouich A., Vicré-Gibouin M. Association between border cell responses and localized root infection by pathogenic *Aphanomyces euteiches*. *Ann. Bot.* 2011;108(3):459-469. doi 10.1093/aob/mcr177

Cannesan M.A., Durand C., Burel C., Gangneux C., Lerouge P., Ishii T., Laval K., Follet-Gueye M.L., Driouich A., Vicré-Gibouin M. Effect of arabinogalactan proteins from the root caps of pea and *Brassica napus* on *Aphanomyces euteiches* zoospore chemotaxis and germination. *Plant Physiol.* 2012;159(4):1658-1670. doi 10.1104/pp.112.198507

Carminati A., Vetterlein D. Plasticity of rhizosphere hydraulic properties as a key for efficient utilization of scarce resources. *Ann. Bot.* 2013;112(2):277-290. doi 10.1093/aob/mcs262

Carreras A., Bernard S., Durambur G., Gügi B., Loutelier C., Pawlak B., Boulogne I., Vicré M., Driouich A., Goffner D., Follet-Gueye M.L. In vitro characterization of root extracellular trap and exudates of three Sahelian woody plant species. *Planta*. 2020;251(1):19. doi 10.1007/s00425-019-03302-3

Castilleux R., Plancot B., Ropitaux M., Carreras A., Leprince J., Boulogne I., Follet-Gueye M.L., Popper Z.A., Driouich A., Vicré M. Cell wall extensins in root – microbe interactions and root secretions. *J. Exp. Bot*. 2018;69(18):4235-4247. doi 10.1093/jxb/ery238

Chubatsu L.S., Monteiro R.A., de Souza E.M., de Oliveira M.A.S., Yates M.G., Wassem R., Bonatto A.C., Huergo L.F., Steffens M.B.R., Rigo L.U., Pedrosa F.D.O. Nitrogen fixation control in *Herbaspirillum seropedicae*. *Plant Soil*. 2012;356:197-207. doi 10.1007/s11104-011-0819-6

Darshan K., Singh J., Yadav S., Venugopala K.M., Aggarwal R. Root border cells: A pioneer's of plant defence in rhizosphere. *Indian J. Agric. Sci*. 2020;90(10):1850-1855. doi 10.56093/ijas.v90i10.107884

Del Campillo E.D., Abdel-Aziz A., Crawford D., Patterson S.E. Root cap specific expression of an endo-β-1,4-D-glucanase (cellulase): a new marker to study root development in *Arabidopsis*. *Plant Mol. Biol*. 2004;56(2):309-323. doi 10.1007/s11103-004-3380-3

Dolan L., Janmaat K., Willemsen V., Linstead P., Poethig S., Roberts K., Scheres B. Cellular organisation of the *Arabidopsis thaliana* root. *Development*. 1993;119(1):71-84. doi 10.1242/dev.119.1.71

Driouich A., Durand C., Vicre-Gibouin M. Formation and separation of root border cells. *Trends Plant Sci.* 2007;12:14-19. doi 10.1016/j.tplants.2006.11.003

Driouich A., Follet-Gueye M.L., Bernard S., Kousar S., Chevalier L., Vicré-Gibouin M., Lerouxel O. Golgi-mediated synthesis and secretion of matrix polysaccharides of the primary cell wall of higher plants. *Front Plant Sci*. 2012;3:79. doi 10.3389/fpls.2012.00079

Driouich A., Follet-Gueye M.L., Vicré-Gibouin M., Hawes M. Root border cells and secretions as critical elements in plant host defense. *Curr. Opin. Plant Biol*. 2013;16(4):489-495. doi 10.1016/j.pbi.2013.06.010

Driouich A., Smith C., Ropitaux M., Chambard M., Boulogne I., Bernard S., Follet-Gueye M.L., Vicré M., Moore J. Root extracellular traps versus neutrophil extracellular traps in host defence, a case of functional convergence? *Biol. Rev. Camb. Philos. Soc.* 2019;94(5): 1685-1700. doi 10.1111/brv.12522

Driouich A., Gaudry A., Pawlak B., Moore J.P. Root cap-derived cells and mucilage: a protective network at the root tip. *Protoplasma*. 2021;258(6):1179-1185. doi 10.1007/s00709-021-01660-y

Durand C., Vicré-Gibouin M., Follet-Gueye M.L., Duponchel L., Moreau M., Lerouge P., Driouich A. The organization pattern of

root border-like cells of Arabidopsis is dependent on cell wall homogalacturonan. *Plant Physiol*. 2009;150(3):1411-1421. doi 10.1104/pp.109.136382

Endo I., Tange T., Osawa H. A cell-type-specific defect in border cell formation in the *Acacia mangium* root cap developing an extraordinary sheath of sloughed-off cells. *Ann. Bot*. 2011;108(2):279-290. doi 10.1093/aob/mcr139

Fendrych M., Hautegem T.V., Durme M.V., Olvera-Carrillo Y., Huysmans M., Karimi M., Lippens S., Guérin C.J., Krebs M., Schumacher K., Nowack M.K. Programmed cell death controlled by ANAC033/SOMBRERO determines root cap organ size in *Arabidopsis*. *Curr. Biol*. 2014;24:931. doi 10.1016/j.cub.2014.03.025

Forino L.M.C., Castiglione M.R., Bartoli G., Balestri M., Andreucci A., Tagliasacchi A.M. Arsenic-induced morphogenic response in roots of arsenic hyperaccumulator fern *Pteris vittata*. *J. Hazard. Mater*. 2012;235-236:271-278. doi 10.1016/j.jhazmat.2012.07.051

Franco-Correa M., Quintana A., Duque C., Suarez C., Rodríguez M.X., Barea J.M. Evaluation of actinomycete strains for key traits related with plant growth promotion and mycorrhiza helping activities. *Appl. Soil Ecol*. 2010;45(3):209-217. doi 10.1016/j.apsoil.2010.04.007

Gochnauer M.B., Sealey L.J., McCully M.E. Do detached root-cap cells influence bacteria associated with maize roots? *Plant Cell Environ.* 1990;13(8):793-801. doi 10.1111/j.1365-3040.1990.tb01095.x

Goh T., Sakamoto K., Wang P., Kozono S., Ueno K., Miyashima S., Toyokura K., Fukaki H., Kang B.H., Nakajima K. Autophagy promotes organelle clearance and organized cell separation of living root-ap cells in *Arabidopsis thaliana*. *Development*. 2022;149(11): dev200593. doi 10.1242/dev.200593

Guinel F.C., McCully M.E. Some water-related physical properties of maize root-cap mucilage. *Plant Cell Environ*. 1986;9(8):657-666. doi 10.1111/J.1365-3040.1986.TB01624.X

Guinel F.C., McCully M.E. The cells shed by the root cap of *Zea*: their origin and some structural and physiological properties. *Plant Cell Environ*. 1987;10(7):565-578. doi 10.1111/1365-3040.EP11604101

Gunawardena U., Hawes M.C. Tissue specific localization of root infection by fungal pathogens: role of root border cells. *Mol. Plant Microbe Interact*. 2002;15(11):1128-1136. doi 10.1094/MPMI.2002.15.11.1128

Gunawardena U., Rodriguez M., Straney D., Romeo J.T., VanEtten H.D., Hawes M.C. Tissue-specific localization of pea root infection by *Nectria haematococca*. Mechanisms and consequences. *Plant Physiol*. 2005;137(4):1363-1374. doi 10.1104/pp.104.056366

Hasan A., Tabassum B., Hashim M., Khan N. Role of plant growth promoting rhizobacteria (PGPR) as a plant growth enhancer for sustainable agriculture: A review. *Bacteria*. 2024;3(2):59-75. doi 10.20944/preprints202310.1504.v1

Hawes M., Allen C., Turgeon B.G., Curlango-Rivera G., Minh Tran T., Huskey D.A., Xiong Z. Root border cells and their role in plant defense. *Annu. Rev. Phytopathol*. 2016;54:143-161. doi 10.1146/annurev-phyto-080615-100140

Hawes M.C., Lin H.J. Correlation of pectolytic enzyme activity with the programmed release of cells from root caps of pea (*Pisum sativum*). *Plant Physiol*. 1990;94(4):1855-1859. doi 10.1104/pp.94.4.1855

Hawes M.C., Brigham L.A., Wen F., Woo H.H., Zhu Y. Function of root border cells in plant health: Pioneers in the rhizosphere. *Annu. Rev. Phytopathol*. 1998;36:311-327. doi 10.1146/annurev.phyto.36.1.311

Hawes M.C., Gunawardena U., Miyasaka S., Zhao X. The role of root border cells in plant defense. *Trends Plant Sci*. 2000;5(3):128-133. doi 10.1016/s1360-1385(00)01556-9

Hawes M.C., Bengough G., Cassab G., Ponce G. Root caps and rhizosphere. *J. Plant Growth Regul*. 2003;21:352-367. doi 10.1007/s00344-002-0035-y

Hawes M.C., Curlango-Rivera G., Wen F., White G.J., VanEtten H.D., Xiong Z. Extracellular DNA: the tip of root defenses? *Plant Sci*. 2011;180(6):741-745. doi 10.1016/j.plantsci.2011.02.007

Iijima M., Barlow P.W., Bengough A.G. Root cap structure and cell production rates of maize (*Zea mays*) roots in compacted sand. *New Phytol*. 2003;160(1):127-134. doi 10.1046/j.1469-8137.2003.00860.x

Jaroszuk-Ściseł J., Kurek E., Rodzik B., Winiarczyk K. Interactions between rye (*Secale cereale*) root border cells (RBCs) and pathogenic and nonpathogenic rhizosphere strains of *Fusarium culmorum*. *Mycol. Res*. 2009;113(10):1053-1061. doi 10.1016/j.mycres.2009.07.001

Jaroszuk-Ściseł J., Tyśkiewicz R., Nowak A., Ozimek E., Majewska M., Hanaka A., Tyśkiewicz K., Pawlik A., Janusz G. Phytohormones (auxin, gibberellin) and ACC deaminase in vitro synthesized by the mycoparasitic *Trichoderma* DEMTkZ3A0 strain and changes in the level of auxin and plant resistance markers in wheat seedlings inoculated with this strain conidia. *Int. J. Mol. Sci*. 2019;20(19):4923. doi 10.3390/ijms20194923

Karve R., Suárez-Román F., Iyer-Pascuzzi A.S. The transcription factor NIN-LIKE PROTEIN7 controls border-like cell release. *Plant Physiol*. 2016;171(3):2101-2111. doi 10.1104/pp.16.00453

Khaliq A., Perveen S., Alamer K.H., Zia Ul Haq M., Rafique Z., Alsudays I.M., Althobaiti A.T., Saleh M.A., Hussain S., Attia H. Arbuscular mycorrhizal fungi symbiosis to enhance plant – soil interaction. *Sustainability*. 2022;14(13):7840. doi 10.3390/su14137840

Knee E.M., Gong F.C., Gao M., Teplitski M., Jones A.R., Foxworthy A., Mort A.J., Bauer W.D. Root mucilage from pea and its utilization by rhizosphere bacteria as a sole carbon source. *Mol. Plant Microbe Interact*. 2001;14(6):775-784. doi 10.1094/MPMI.2001.14.6.775

Koroney A.S., Plasson C., Pawlak B., Sidikou R., Driouich A., Menu-Bouaouiche L., Vicré-Gibouin M. Root exudate of *Solanum tuberosum* is enriched in galactose-containing molecules and impacts the growth of *Pectobacterium atrosepticum*. *Ann. Bot*. 2016;118(4):797-808. doi 10.1093/aob/mcw128

Lee J.Y., Hwang B.K. Diversity of antifungal actinomycetes in various vegetative soils of Korea. *Can. J. Microbiol*. 2002;48(5):407-417. doi 10.1139/w02-025

Lilley C.J., Wang D., Atkinson H.J., Urwin P.E. Effective delivery of a nematode-repellent peptide using a root-cap-specific promoter. *Plant Biotechnol. J*. 2011;9(2):151-161. doi 10.1111/j.1467-7652.2010.00542.x

Ma W., Muthreich N., Liao C., Franz-Wachtel M., Schütz W., Zhang F., Hochholdinger F., Li C. The mucilage proteome of maize (*Zea mays* L.) primary roots. *J. Proteome Res*. 2010;9(6):2968-2976. doi 10.1021/pr901168v

Maeda K., Kunieda T., Tamura K., Hatano K., Hara-Nishimura I., Shimada T. Identification of periplasmic root-cap mucilage in developing columella cells of *Arabidopsis thaliana*. *Plant Cell Physiol*. 2019;60(6):1296-1303. doi 10.1093/pcp/pcz047

Mendes R., Garbeva P., Raaijmakers J.M. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol. Rev*. 2013;37(5):634-663. doi 10.1111/1574-6976.12028

Micheli F. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci*. 2001;6(9):414-419. doi 10.1016/s1360-1385(01)02045-3

Mohanram S., Kumar P. Rhizosphere microbiome: revisiting the synergy of plant-microbe interactions. *Ann. Microbiol*. 2019;69(3):307-320. doi 10.1007/s13213-019-01448-9

Monticolo F., Palomba E., Termolino P., Chiaiese P., De Alteriis E., Mazzoleni S., Chiusano M.L. The role of DNA in the extracellular environment: a focus on NETs, RETs and biofilms. *Front. Plant Sci*. 2020;11:589837. doi 10.3389/fpls.2020.589837

Moustacas A.M., Nari J., Borel M., Noat G., Ricard J. Pectin methylesterase, metal ions and plant cell-wall extension. The role of metal ions in plant cell-wall extension. *Biochem. J*. 1991;279(2):351-354. doi 10.1042/bj2790343

Nagahashi G., Douds D.D. Isolated root caps, border cells, and mucilage from host roots stimulate hyphal branching of the arbuscular mycorrhizal fungus, *Gigaspora gigantea*. *Mycol. Res*. 2004;108(9):1079-1088. doi 10.1017/s0953756204000693

Niemira B.A., Safir G.R., Hawes M.C. Arbuscular mycorrhizal colonization and border cell production: a possible correlation. *Phytopathology*. 1996;86(6):563-565

Odell R.E., Dumlao M.R., Samar D., Silk W.K. Stage-dependent border cell and carbon flow from roots to rhizosphere. *Am. J. Bot*. 2008; 95(4):441-446. doi 10.3732/ajb.95.4.441

Pankievicz V.C.S., Delaux P.M., Infante V., Hirsch H.H., Rajasekar S., Zamora P., Jayaraman D., Calderon C.I., Bennett A., Ané J.M. Nitrogen fixation and mucilage production on maize aerial roots is controlled by aerial root development and border cell functions. *Front. Plant Sci*. 2022;13:977056. doi 10.3389/fpls.2022.977056

Plancot B., Santaella C., Jaber R., Kiefer-Meyer M.C., Follet-Gueye M.L., Leprince J., Gattin I., Souc C., Driouich A., Vicré-Gibouin M. Deciphering the responses of root border-like cells of Arabidopsis and flax to pathogen-derived elicitors. *Plant Physiol*. 2013;163(4):1584-1597. doi 10.1104/pp.113.222356

Poulsen L.R., López-Marqués R.L., McDowell S.C., Okkeri J., Licht D., Schulz A., Pomorski T., Harper J.F., Palmgren M.G. The *Arabidopsis* P_4-ATPase ALA3 localizes to the Golgi and requires a β-subunit to function in lipid translocation and secretory vesicle formation. *Plant Cell*. 2008;20(3):658-676. doi 10.1105/tpc.107.054767

Ropitaux M., Bernard S., Follet-Gueye M.L., Vicré M., Boulogne I., Driouich A. Xyloglucan and cellulose form molecular cross-bridges connecting root border cells in pea (*Pisum sativum*). *Plant Physiol. Biochem*. 2019;139:191-196. doi 10.1016/j.plaphy.2019.03.023

Ropitaux M., Bernard S., Schapman D., Follet-Gueye M.L., Vicré M., Boulogne I., Driouich A. Root border cells and mucilage secretions of soybean, *Glycine max* (Merr) L.: characterization and role in interactions with the oomycete *Phytophthora parasitica*. *Cells*. 2020;9(10):2215. doi 10.3390/cells9102215

Shi C.-L., von Wangenheim D., Herrmann U., Wildhagen M., Kulik I., Kopf A., Ishida T., Olsson V., Anker M.K., Albert M., Butenko M.A., Felix G., Sawa S., Claassen M., Friml J., Aalen R.B. The dynamics of root cap sloughing in *Arabidopsis* is regulated by peptide signalling. *Nat. Plants*. 2018;4(8):596-604. doi 10.1038/s41477-018-0212-z

Shirakawa M., Matsushita N., Fukuda K. Visualization of root extracellular traps in an ectomycorrhizal woody plant (*Pinus densiflora*) and their interactions with root-associated bacteria. *Planta*. 2023; 258(6):112. doi 10.1007/s00425-023-04274-1

Tran T.M., MacIntyre A., Hawes M., Allen C. Escaping underground nets: extracellular DNases degrade plant extracellular traps and contribute to virulence of the plant pathogenic bacterium *Ralstonia solanacearum*. *PLoS Pathog*. 2016;12(6):e1005686. doi 10.1371/journal.ppat.1005686

Ulloa-Ogaz A.L., Muñoz-Castellanos L.N., Nevárez-Moorillón G.V. Biocontrol of phytopathogens: Antibiotic production as mechanism of control. In: Méndez-Vilas A. (Ed.). The Battle Against Microbial Pathogens: Basic Science, Technological Advances and Educational Programs. Formatex, 2015;305-309

Vermeer J., McCully M.E. The rhizosphere in *Zea*: New insight into its structure and development. *Planta.* 1982;156:45-61. doi 10.1007/BF00393442

Vicré M., Santaella C., Blanchet S., Gateau A., Driouich A. Root border-like cells of Arabidopsis. Microscopical characterization and role in the interaction with rhizobacteria. *Plant Physiol*. 2005;138: 998-1008. doi 10.1104/pp.104.051813

Wang P., Chen X., Goldbeck C., Chung E., Kang B.H. A distinct class of vesicles derived from the *trans*-Golgi mediates secretion of xylogalacturonan in the root border cell. *Plant J*. 2017;92(4):596-610. doi 10.1111/tpj.13704

Watson B.S., Bedair M.F., Urbanczyk-Wochniak E., Huhman D.V., Yang D.S., Allen S.N., Li W., Tang Y., Sumner L.W. Integrated metabolomics and transcriptomics reveal enhanced specialized metabolism in *Medicago truncatula* root border cells. *Plant Physiol*. 2015;167(4):1699-1716. doi 10.1104/pp.114.253054

Weiller F., Moore J.P., Young P., Driouich A., Vivier M.A. The Brassicaceae species *Heliophila coronopifolia* produces root border-like cells that protect the root tip and secrete defensin peptides. *Ann. Bot*. 2017;119(5):803-813. doi 10.1093/aob/mcw141

Wen F., Zhu Y., Hawes M.C. Effect of pectin methylesterase gene expression on pea root development. *Plant Cell*. 1999;11(6):1129-1140. doi 10.1105/tpc.11.6.1129

Wen F., VanEtten H.D., Tsaprailis G., Hawes M.C. Extracellular proteins in pea root tip and border cell exudates. *Plant Physiol*. 2007; 143(2):773-783. doi 10.1104/pp.106.091637

Wen F., White G.J., VanEtten H.D., Xiong Z., Hawes M.C. Extracellular DNA is required for root tip resistance to fungal infection. *Plant Physiol*. 2009;151(2):820-829. doi 10.1104/pp.109.142067

Wen F., Curlango-Rivera G., Huskey D.C., Xiong Z., Hawes M.C. Visualization of extracellular DNA released during border cell separation from the root cap. *Am. J. Bot*. 2017;104(7):970-978. doi 10.3732/ajb.1700142

Wuyts N., Maung Z.T.Z., Swennen R., De Waele D. Banana rhizodeposition: characterization of root border cell production and effects on chemotaxis and motility of the parasitic nematode *Radopholus similis*. *Plant Soil*. 2006;283:217-228. doi 10.1007/s11104-006-0013-4

Zhao X., Misaghi I.J., Hawes M.C. Stimulation of border cell production in response to increased carbon dioxide levels. *Plant Physiol.* 2000;122:181-186. doi 10.1104/pp.122.1.181

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 08.04.2024. После доработки 06.11.2024. Принята к публикации 07.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):927-939

doi 10.18699/vjgb-24-100

Метод генных сетей и метаболомный анализ позволили выявить специфические пути изменения профиля аминокислот и ацилкарнитинов в плазме крови при болезни Паркинсона и сосудистом паркинсонизме

А.А. Макарова [1, 2], П.М. Мельникова[2], А.Д. Рогачев [2, 3], П.С. Деменков [1, 2, 4], Т.В. Иванисенко [1, 2, 4], Е.В. Предтеченская[2], С.Ю. Карманов[1, 2], В.В. Коваль [5], А.Г. Покровский [2], И.Н. Лаврик[1], Н.А. Колчанов [1, 2], В.А. Иванисенко [1, 2, 4]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Новосибирский институт органической химии им. Н.Н. Ворожцова Сибирского отделения Российской академии наук, Новосибирск, Россия
[4] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[5] Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
makarovaaa@bionet.nsc.ru

Аннотация. Болезнь Паркинсона (БП) и сосудистый паркинсонизм (СП) характеризуются схожими неврологическими синдромами, но различаются патогенезом, морфологией и терапевтическими подходами. Их молекулярно-генетические механизмы многофакторны и задействуют множество биологических процессов. Для комплексного анализа патофизиологии этих заболеваний необходимо применение методов системной биологии и реконструкции генных сетей. В данном исследовании проведен метаболомный скрининг аминокислот и ацилкарнитинов в плазме крови трех групп испытуемых: пациентов с БП, пациентов с СП и контрольной группы. Сравнительный статистический анализ метаболомных профилей групп пациентов по сравнению с контролем определил значимо измененные уровни метаболитов при болезни Паркинсона и при сосудистом паркинсонизме. Для выявления потенциальных механизмов нарушения метаболизма аминокислот и ацилкарнитинов при БП и СП были реконструированы регуляторные генные сети с помощью когнитивной системы ANDSystem. Пути регуляции ферментов метаболизма значимых метаболитов были найдены для трех групп генетических маркеров: специфических для БП, специфических для СП, а также группы общих маркеров двух заболеваний. Сравнительный анализ молекулярно-генетических путей в генных сетях позволил выявить как специфические, так и общие для БП и СП молекулярные механизмы, ассоциированные с изменением метаболомного профиля. Обнаружены регуляторные пути, функция которых потенциально нарушена при этих патологиях. Специфическими для генетических маркеров БП оказались пути регуляции ферментов ALDH2, BCAT1, AL1B1 и UD11, а для генетических маркеров СП – пути регуляции ферментов OCTC, FURIN и S22A6. Регуляторные пути к ферментам BCAT2, ODPB и P4HA1 были связаны с общими для обоих заболеваний генетическими маркерами. Полученные результаты углубляют понимание патологических процессов при БП и СП и могут быть использованы для применения диагностических систем на основе оценки метаболомного профиля аминокислот и ацилкарнитинов в плазме крови пациентов с болезнью Паркинсона и сосудистым паркинсонизмом.

Ключевые слова: метаболомика; аминокислоты; ацилкарнитины; генные сети; генетический маркер; болезнь Паркинсона; сосудистый паркинсонизм; нейродегенерация; сухие пятна плазмы крови; биомаркер.

Для цитирования: Макарова А.А., Мельникова П.М., Рогачев А.Д., Деменков П.С., Иванисенко Т.В., Предтеченская Е.В., Карманов С.Ю., Коваль В.В., Покровский А.Г., Лаврик И.Н., Колчанов Н.А., Иванисенко В.А. Метод генных сетей и метаболомный анализ позволили выявить специфические пути изменения профиля аминокислот и ацилкарнитинов в плазме крови при болезни Паркинсона и сосудистом паркинсонизме. *Вавиловский журнал генетики и селекции*. 2024;28(8):927-939. doi 10.18699/vjgb-24-100

Финансирование. Авторы выражают благодарность компании «Новартис Фарма» за финансовую поддержку, благодаря которой были приобретены наборы реактивов для анализа аминокислот и ацилкарнитинов в рамках данной работы. Экспериментальные исследования выполнены при поддержке государственного задания ИХБФМ СО РАН № 121031300045-2. Биоинформатический анализ проведен при поддержке бюджетного проекта FWNR-2022-0020.

© Макарова А.А., Мельникова П.М., Рогачев А.Д., Деменков П.С., Иванисенко Т.В., Предтеченская Е.В., Карманов С.Ю., Коваль В.В., Покровский А.Г., Лаврик И.Н., Колчанов Н.А., Иванисенко В.А., 2024

Контент доступен под лицензией Creative Commons Attribution 4.0

Gene networks and metabolomic screening analysis revealed specific pathways of amino acid and acylcarnitine profile alterations in blood plasma of patients with Parkinson's disease and vascular parkinsonism

A.A. Makarova [1, 2], P.M. Melnikova[2], A.D. Rogachev [2, 3], P.S. Demenkov [1, 2, 4], T.V. Ivanisenko [1, 2, 4], E.V. Predtechenskaya[2], S.Y. Karmanov[1, 2], V.V. Koval [5], A.G. Pokrovsky [2], I.N. Lavrik[1], N.A. Kolchanov [1, 2], V.A. Ivanisenko [1, 2, 4]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[4] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
[5] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
makarovaaa@bionet.nsc.ru

Abstract. Parkinson's disease (PD) and vascular parkinsonism (VP) are characterized by similar neurological syndromes but differ in pathogenesis, morphology, and therapeutic approaches. The molecular genetic mechanisms of these pathologies are multifactorial and involve multiple biological processes. To comprehensively analyze the pathophysiology of PD and VP, the methods of systems biology and gene network reconstruction are essential. In the current study, we performed metabolomic screening of amino acids and acylcarnitines in blood plasma of three groups of subjects: PD patients, VP patients and the control group. Comparative statistical analysis of the metabolic profiles identified significantly altered metabolites in the PD and the VP group. To identify potential mechanisms of amino acid and acylcarnitine metabolism disorders in PD and VP, regulatory gene networks were reconstructed using ANDSystem, a cognitive system . Regulatory pathways to the enzymes converting significant metabolites were found from PD-specific genetic markers, VP-specific genetic markers, and the group of genetic markers common to the two diseases. Comparative analysis of molecular genetic pathways in gene networks allowed us to identify both specific and non-specific molecular mechanisms associated with changes in the metabolomic profile in PD and VP. Regulatory pathways with potentially impaired function in these pathologies were discovered. The regulatory pathways to the enzymes ALDH2, BCAT1, AL1B1, and UD11 were found to be specific for PD, while the pathways regulating OCTC, FURIN, and S22A6 were specific for VP. The pathways regulating BCAT2, ODPB and P4HA1 were associated with genetic markers common to both diseases. The results obtained deepen the understanding of pathological processes in PD and VP and can be used for application of diagnostic systems based on the evaluation of the amino acids and acylcarnitines profile in blood plasma of patients with PD and VP.

Key words: metabolomics; amino acids; acylcarnitines; gene networks; genetic markers; Parkinson's disease; vascular parkinsonism; neurodegeneration; dry plasma stains; biomarker.

For citation: Makarova A.A., Melnikova P.M., Rogachev A.D., Demenkov P.S., Ivanisenko T.V., Predtechenskaya E.V., Karmanov S.Y., Koval V.V., Pokrovsky A.G., Lavrik I.N., Kolchanov N.A., Ivanisenko V.A. Gene networks and metabolomic screening analysis revealed specific pathways of amino acid and acylcarnitine profile alterations in blood plasma of patients with Parkinson's disease and vascular parkinsonism *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):927-939. doi 10.18699/vjgb-24-100

Введение

Болезнь Паркинсона (БП) и сосудистый паркинсонизм (СП) являются комплексными заболеваниями, при которых у пациентов наблюдаются брадикинезия, ригидность мышц, нарушения походки и равновесия. Болезнь Паркинсона относится к нейродегенеративным заболеваниям, а сосудистый паркинсонизм, также известный как «болезнь малых сосудов», возникает на фоне цереброваскулярных заболеваний.

В патогенезе болезни Паркинсона ключевую роль играют нарушения в нигростриатальном дофаминергическом пути с существенной потерей нейронов компактной части черной субстанции и истощением запасов дофамина (Alexander, 2004). Нейродегенеративные процессы характеризуются четкой морфологической ступенчатостью развития: от поражения ядер блуждающего нерва и обонятельной луковицы до критической гибели нейронов компактной зоны черной субстанции (Braak et al., 2003). Эта ступенчатость согласуется с постепенным развитием клинических симптомов БП, начиная от вегетативных расстройств и заканчивая ядром моторных нарушений (брадикинезия, тремор, ригидность) и когнитивными нарушениями. Молекулярные механизмы развития БП активно изучаются в научном сообществе. Основное значение придается протеолитическому стрессу, нарушению энергетического метаболизма нейронов черной субстанции и митохондриальной дисфункции (Левин и др., 2022), а также накоплению альфа-синуклеина (Rocha et al., 2018).

В то же время механизмы, лежащие в основе сосудистого паркинсонизма, связанного с цереброваскулярными

заболеваниями (ЦВЗ), остаются недостаточно изученными. СП часто возникает на фоне ЦВЗ и хронического нарушения мозгового кровообращения, хронической ишемии мозга, вызванной болезнью малых сосудов, которая приводит к дисфункции нейроглиоваскулярной единицы (Che Mohd Nassir et al., 2021). При СП симптомы развиваются быстрее, чем при БП, и включают двусторонний паркинсонизм нижней части тела, отсутствие тремора, пирамидные знаки и когнитивные нарушения (Vale et al., 2012). В отличие от БП, где основное значение придается протеолитическому стрессу, нарушению энергетического метаболизма нейронов черной субстанции и митохондриальной дисфункции (Левин и др., 2022), механизмы патогенеза которой включают гибель дофаминергических нейронов и накопление α-синуклеина (Rocha et al., 2018), при СП на первый план выходят нарушения микроциркуляции и гемодинамики. Ключевым фактором развития СП является поражение малых сосудов головного мозга, часто ассоциированное с длительным анамнезом артериальной гипертензии (Che Mohd Nassir et al., 2021) и сахарного диабета (Thanvi et al., 2005). Хроническая ишемия, возникающая в результате цереброваскулярных нарушений, вызывает окислительный стресс, воспаление и митохондриальную дисфункцию. Эти патологические процессы приводят к существенным изменениям структуры и функции нейроглиоваскулярной единицы, включая эндотелиальную дисфункцию, нарушение проницаемости гематоэнцефалического барьера, а также изменения в астроцитах и перицитах (Narasimhan et al., 2022). В итоге происходит повреждение белого вещества головного мозга (лейкоареоз) и формирование множественных лакунарных инфарктов в стратегически важных зонах базальных ганглиев (Zijlmans et al., 1995; Chen Y.-F. et al., 2014; Korczyn, 2015).

Среди общих характеристик перечисленных патологических процессов можно выделить нарушения метаболизма липидов, аминокислот и энергетических молекул, что подчеркивает их важность в молекулярных механизмах патологий. Аминокислоты и ацилкарнитины вовлечены во многие процессы, включая биосинтез нейромедиаторов и энергетический обмен (Jones et al., 2010; Dalangin et al., 2020). В литературе представлены исследования, посвященные анализу метаболомных профилей при БП (Wuolikainen et al., 2016; Zhao et al., 2018; Ostrakhovitch et al., 2022), однако роль аминокислот и ацилкарнитинов требует дальнейшего изучения. Также надо отметить, что к настоящему времени мы не нашли метаболомных исследований сосудистого паркинсонизма.

Для изучения таких комплексных заболеваний, как БП и СП, широкое применение получили генные сети, позволяющие интегрировать накопленные данные и идентифицировать регуляторные механизмы патологий на молекулярно-генетическом уровне (Mercatelli et al., 2020). Известны исследования генных сетей болезни Паркинсона, включая сети белок-белковых взаимодействий маркеров БП (George et al., 2019a; Tomkins, Manzoni, 2021), сети коэкспрессии генов (George et al., 2019b), регуляторные пути (https://www.kegg.jp/entry/hsa05012) и др. В отличие от БП, исследования молекулярно-генетических механизмов СП на основе генных сетей слабо представлены в научной литературе (Chen Y. et al., 2022).

В Институте цитологии и генетики СО РАН была разработана когнитивная система ANDSystem для реконструкции и анализа генных сетей с использованием методов искусственного интеллекта (Demenkov et al., 2012; Ivanisenko V.A. et al., 2015, 2019; Ivanisenko T.V. et al., 2020, 2022). Эта программа применялась для интерпретации метаболомных (Rogachev et al., 2021; Ivanisenko V.A. et al., 2022, 2023) и протеомных (Pastushkova et al., 2013, 2019; Binder et al., 2014; Larina et al., 2015) данных. С помощью ANDSystem были проведены биоинформатические исследования, позволившие расширить представление о молекулярно-генетических процессах, ассоциированных с развитием различных заболеваний и формированием коморбидных состояний (Bragina et al., 2014, 2016, 2023; Saik et al., 2016, 2018, 2019; Zolotareva et al., 2019).

Целью настоящей работы был сравнительный анализ молекулярно-генетических механизмов заболеваний БП и СП с применением методов реконструкции генных сетей на основе данных метаболомного скрининга аминокислот и ацилкарнитинов.

Материалы и методы

Характеристика групп пациентов. Для исследования были отобраны две группы пациентов с установленными диагнозами БП и СП, а также группа контроля. Дифференциальный диагноз проводился по клиническим и томографическим признакам. На момент взятия образцов крови прием препарата L-ДОФА (L-диоксифенилаланин) был прекращен.

Группа БП включала 9 пациентов (5 женщин, 4 мужчин) со средним возрастом 72.2 года (от 64 до 88 лет). Критерии включения в группу: клинически достоверная БП, стадия IV по Хен и Яру, длительность болезни >5 лет, возраст начала 55–75 лет; симптомы – брадикинезия, тремор покоя или мышечная ригидность, ответ на L-ДОФА терапию. Группа СП включала 9 пациентов (7 женщин, 2 мужчин), средний возраст которых составлял 74.6 года (от 60 до 89 лет). Критерии включения: длительность болезни >3 лет, мультилакунарный статус и лейкоареоз по МРТ; симптомы – паркинсонизм нижней половины тела, двусторонний дебют, постуральная неустойчивость. Контрольная группа состояла из 17 условно здоровых человек (11 женщин, 6 мужчин). Средний возраст испытуемых 68 лет (от 51 до 82). Фоновыми заболеваниями данной группы были: хроническая артериальная гипертензия без транзиторных ишемических атак, инсульта, отсутствие неврологической симптоматики.

Забор биологического материала и анализ методом ВЭЖХ-МС/МС. Забор крови осуществлялся из периферической вены в дневное время, через 3 ч после приема пищи, в пробирки для получения плазмы с гепаринатом лития 68 МЕ, объемом 6 мл (Vacutainer, BD). Пробоподготовку плазмы крови и анализ аминокислот и ацилкарнитинов (список анализируемых метаболитов приведен в Приложении 1)[1] методом ВЭЖХ-МС/МС выполняли

[1] Приложения 1–9 см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx33.xlsx

Шаблоны молекулярно-генетических путей регуляции ферментов биосинтеза и деградации значимых метаболитов генетическими маркерами БП, СП или общими для обоих заболеваний

Название	Шаблон регуляторного пути
Белок-белковые взаимодействия	Генетические маркеры – белок-белковые взаимодействия → Ферменты
Регуляция функции белка	Генетические маркеры – регуляция активности/деградации/посттрансляционных модификаций/каталитических реакций/транспорта белка → Ферменты
Регуляция экспрессии	Генетические маркеры – регуляция экспрессии → Гены ферментов – экспрессия → Ферменты
Двойная регуляция экспрессии	Генетические маркеры – регуляция экспрессии → Гены человека – экспрессия → Белки человека – регуляция экспрессии → Гены ферментов – экспрессия → Ферменты

Примечание. Генетические маркеры – белки, кодируемые генетическими маркерами (БП, СП или общие маркеры заболеваний); Ферменты – ферменты превращения значимых метаболитов; Гены ферментов – гены, кодирующие ферменты превращения метаболитов.

по методике, описанной ранее (Kasakin et al., 2019), с использованием масс-спектрометра API 6500 QTRAP (AB SCIEX, США) в сочетании с хроматографом ВЭЖХ LC-20AD Prominence (Shimadzu, Япония), оснащенным автосамплером SIL-20AC (Shimadzu).

Статистический анализ экспериментальных данных. Для статистического анализа различий содержания метаболитов в метаболомных профилях исследуемых групп применяли критерии Манна–Уитни и Колмогорова–Смирнова с поправками на множественное сравнение FDR. Расчеты проведены на языке программирования Python 3.11 с помощью функций модуля scipy.stats.

Формирование списков ферментов и генетических маркеров БП и СП. Для каждого из метаболитов, концентрации которых были статистически значимо изменены в группах пациентов относительно контрольной группы, были сформированы списки ферментов биосинтеза и деградации. Ферменты, осуществляющие реакции превращения значимых метаболитов, извлечены из баз данных KEGG (Kanehisa, 2000) и HMDB (Wishart et al., 2022).

Списки генетических маркеров болезни Паркинсона и сосудистого паркинсонизма были взяты из базы данных MalaCards (https://www.malacards.org/, дата обращения 25.01.2024) (Rappaport et al., 2014). В качестве генетических маркеров СП рассматривались белок-кодирующие гены, ассоциированные с сосудистым паркинсонизмом, а также с васкулярной деменцией: "Vascular Parkinsonism", "Vascular Dementia". В качестве генетических маркеров БП рассматривались белок-кодирующие гены, ассоциированные с болезнью Паркинсона: "Parkinson's Disease".

Построение генных сетей. Реконструкция генных сетей осуществлялась с помощью программы ANDVisio – графического пользовательского интерфейса программно-информационной системы ANDSystem (Ivanisenko V.A. et al., 2015). Регуляторные пути четырех типов были построены в соответствии с шаблонами, описанными в таблице. Данные шаблоны позволяют найти такие молекулярно-генетические взаимодействия, как белок-белковые, регуляция активности, протеолиза, транспорта белков, а также регуляция экспрессии генов. Построение молекулярно-генетических путей регуляции ферментов превращения метаболитов проводилось по одинаковым шаблонам для трех множеств генетических маркеров: БП, СП и общих для обоих заболеваний маркеров.

Результаты

Статистический анализ метаболомных данных

Статистический анализ метаболомных данных (см. Приложение 1), направленный на выявление различий в уровнях метаболитов между группами БП и СП в сравнении с контрольной группой, показал, что из 44 метаболитов, у которых измерялись концентрации, статистически значимые различия (FDR < 0.05) в концентрациях наблюдались для 18 метаболитов при БП и для 21 метаболита при СП (Приложение 2).

Обе патологии отличались от группы контроля уровнем одних и тех же четырех из 14 анализируемых аминокислот (аланин, пролин, изолейцин и валин). Следует отметить, что уровень метионина оказался статистически значимо измененным в группе БП, хотя не отличал группу СП от контроля. Среди ацилкарнитинов выявлены значимые различия для концентраций одних и тех же 13 метаболитов для БП и СП (см. Приложение 2). Специфическими метаболитами ряда ацилкарнитинов, уровни которых значимо отличались от контроля только в группе СП, были ацилкарнитины C6, C10 и C10:1 и карнитин (Carnitine).

Реконструкция и анализ генных сетей

Для выявления молекулярно-генетических механизмов, которые могли способствовать изменению метаболомного профиля при БП и СП, мы использовали подход генных сетей. Генные сети позволили интегрировать знания о молекулярных взаимосвязях метаболитов и известных генетических маркеров БП и СП. В качестве генетических маркеров мы рассматривали гены, ассоциированные с БП и СП, согласно информации из базы данных MalaCards (Rappaport et al., 2014). Списки из 84 генетических маркеров болезни Паркинсона и 60 генетических маркеров сосудистого паркинсонизма приведены в Приложении 3. Пересечение списков генетических маркеров БП и СП показало, что 22 генетических маркера были общими для этих заболеваний.

Далее мы применили методы реконструкции генных сетей с помощью программы ANDVisio (Ivanisenko V.A. et al., 2019), чтобы исследовать роль генетических маркеров в регуляции ферментов превращения значимых метаболитов. Этот подход основан на автоматической реконструкции регуляторных молекулярно-генетических пу-

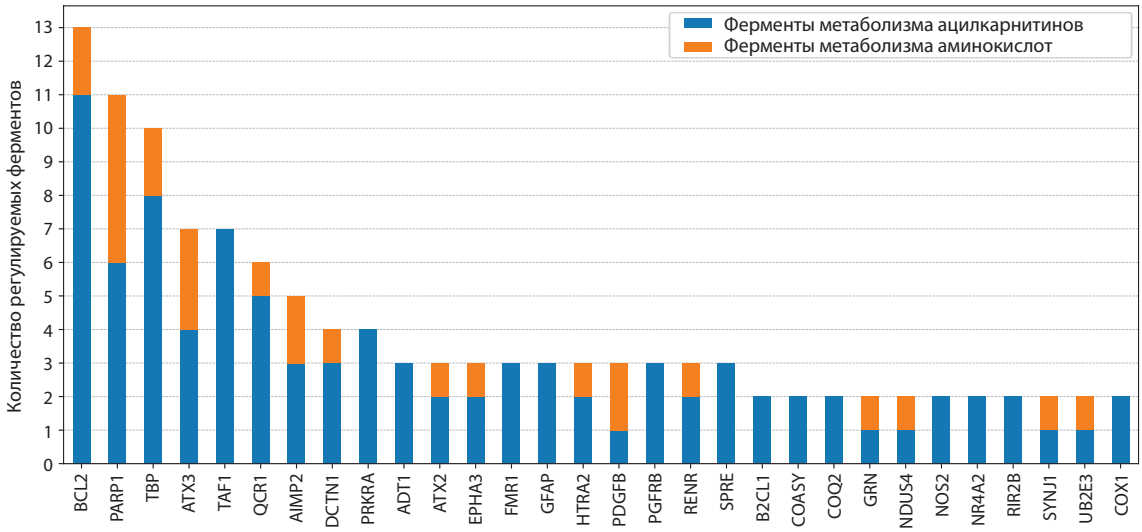


Рис. 1. Гистограмма распределения регуляторных связей от генетических маркеров БП к ферментам метаболизма аминокислот и ацилкарнитинов в реконструированных генных сетях.

тей с использованием задаваемых программе ANDVisio шаблонов путей (см. таблицу).

Были рассмотрены шаблоны, в которых все реконструируемые регуляторные пути стартуют с белков, кодируемых генетическими маркерами БП или СП или общими маркерами для обоих заболеваний. Списки этих белков были заданы в качестве входных данных при запуске модуля Pathway Wizard программы ANDVisio. Регуляторные пути заканчиваются ферментами биосинтеза и деградации соответствующих метаболитов, идентифицированных как значимые при статистическом анализе. Списки использованных для анализа ферментов приведены в Приложении 4. Также пути включают промежуточных участников, связывающих генетические маркеры с ферментами, в роли которых выступают белки человека. Все промежуточные участники регуляторных путей в явном виде не задавались во входных данных, программа проводила поиск таких посредников автоматически. В регуляторных путях были рассмотрены основные типы молекулярно-генетических взаимодействий, включая регуляцию экспрессии генов, белок-белковые взаимодействия, регуляцию активности, деградации, транспорта белков и др. Иллюстрации генных сетей приведены в Приложениях 5 и 6. Количество регуляторных связей к каждому ферменту, направленных от генетических маркеров БП, СП и группы общих маркеров, приведено в Приложениях 7–9.

Распределение числа регуляторных связей между участниками реконструированных генных сетей представлено на рис. 1 и 2. По оси абсцисс указаны имена генетических маркеров заболеваний, по оси ординат отображено количество регуляторных воздействий, реализуемых через молекулярно-генетические пути от генетических маркеров (БП, СП и общих маркеров обоих заболеваний) к ферментам превращения значимых метаболитов.

Генетические маркеры болезни Паркинсона BCL2, TBP, TAF1 оказывают большее регуляторное воздействие на ферменты метаболизма ацилкарнитинов, в то время как PARP1 в равной степени воздействует на ферменты

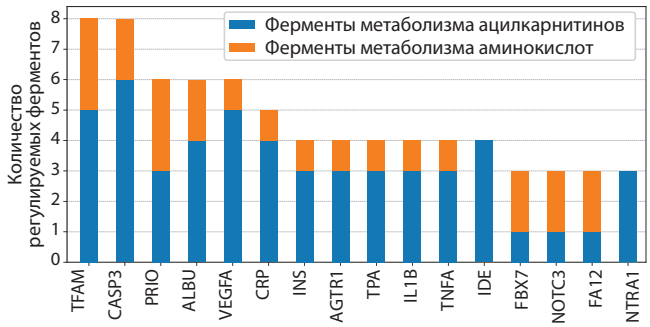


Рис. 2. Гистограмма распределения регуляторных связей от генетических маркеров СП к ферментам метаболизма аминокислот и ацилкарнитинов в реконструированных генных сетях.

метаболизма аминокислот и ацилкарнитинов (см. рис. 1). Такие генетические маркеры сосудистого паркинсонизма, как TFAM, CASP3, ALBU, VEGFA, имеют большее регуляторное воздействие на ферменты метаболизма ацилкарнитинов, тогда как FBX7, NOTC3 и FA12 – на ферменты метаболизма аминокислот (см. рис. 2).

Следует отметить, что общие для БП и СП генетические маркеры примерно равно участвуют в регуляции ферментов метаболизма аминокислот и ацилкарнитинов (рис. 3). Генетический маркер LRRK2, согласно генным сетям, оказывает большее воздействие на регуляцию метаболизма аминокислот.

К некоторым ферментам превращения метаболитов, уровни которых значимо различались у пациентов с БП и СП относительно контроля, регуляторные пути от генетических маркеров БП, СП и общих маркеров реализовывались в разном количественном соотношении. Для генных сетей (см. Приложения 5 и 6) были построены гистограммы, отражающие количество регуляторных воздействий от групп генетических маркеров к ферментам метаболизма аминокислот и ацилкарнитинов (рис. 4 и 5).

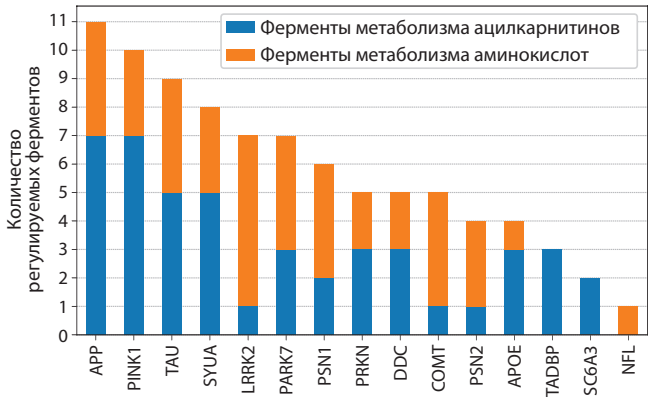


Рис. 3. Гистограмма распределения регуляторных связей от общих для БП и СП генетических маркеров к ферментам метаболизма аминокислот и ацилкарнитинов в реконструированных генных сетях.

К ферментам ALDH2, BCAT1, AL1B1, P5CR1 в большей степени реализуются связи от генетических маркеров БП, а к ферментам BCAT2, P4HA1 направлено больше связей от общих для БП и СП генетических маркеров (см. рис. 4). Среди ферментов метаболизма ацилкарнитинов наибольшему регуляторному влиянию оказалась подвержена синтаза жирных кислот FAS (см. рис. 5). Для ферментов FAS, ODPA (PDHA1), ACACA (ACC1) пути регуляции реализуются со стороны генетических маркеров, специфичных как для БП, так и для СП, в то время как в механизме регуляции фермента ODPB больше участвуют общие для двух заболеваний генетические маркеры.

Таким образом, анализ метаболомных данных позволил выявить 5 аминокислот и 17 ацилкарнитинов, концентрации которых были значимо изменены в группах пациентов с БП, СП относительно контрольной группы (см. При-

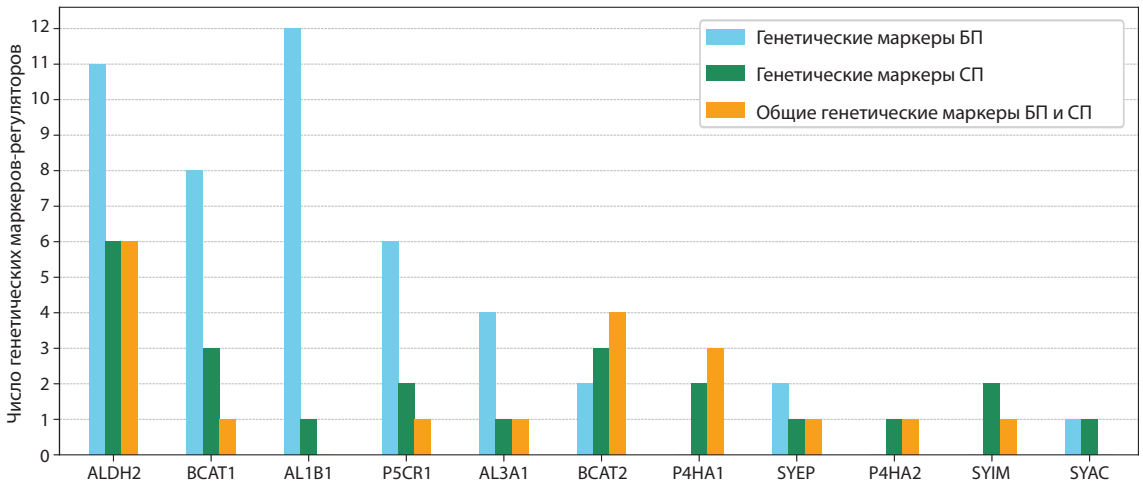


Рис. 4. Гистограмма распределения регуляторных связей от трех групп генетических маркеров (БП, СП и общих) к ферментам метаболизма аминокислот в реконструированных генных сетях.

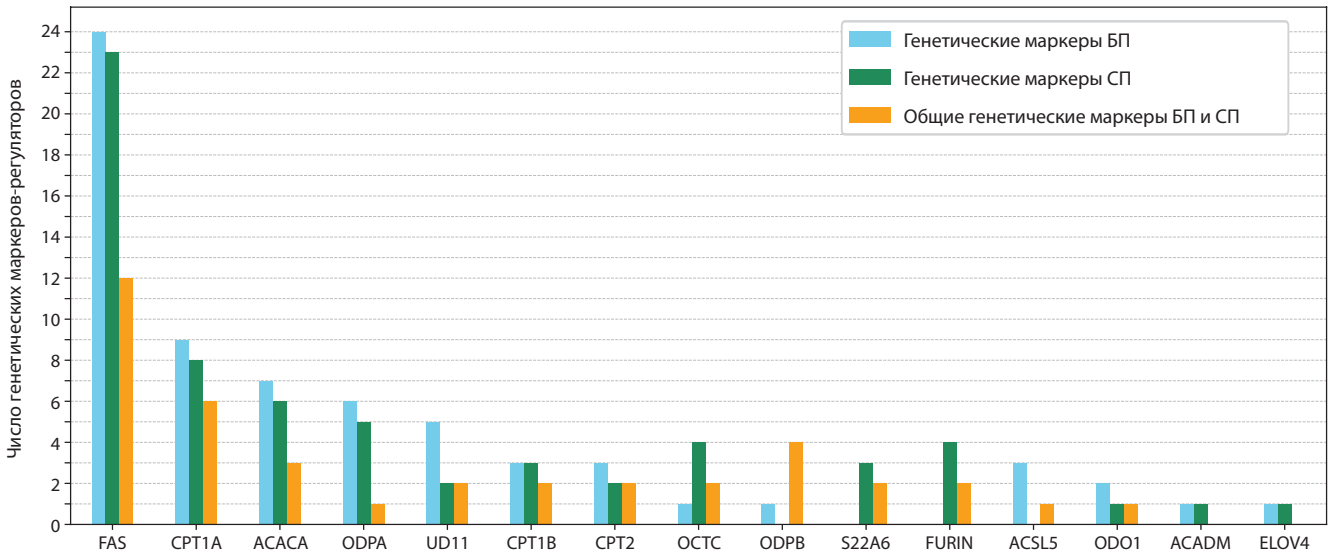


Рис. 5. Гистограмма распределения регуляторных связей от трех групп генетических маркеров (БП, СП и общих) к ферментам метаболизма ацилкарнитинов в реконструированных генных сетях.

ложение 2). Подход генных сетей позволил реконструировать и проанализировать регуляторные молекулярно-генетические пути от генетических маркеров БП и СП к ферментам, участвующим в метаболизме аминокислот и ацилкарнитинов.

Обсуждение

Специфические и неспецифические маркеры БП и СП

Согласно нашим результатам, списки из 18 и 21 маркеров БП и СП в пересечении имели 17 одинаковых метаболитов, которые можно назвать неспецифическими маркерами для дифференциальной диагностики БП и СП. В связи с этим анализ только метаболомных профилей был малоинформативным для дифференцирования особенностей болезни Паркинсона и сосудистого паркинсонизма. Мы предположили, что, хотя потенциальные метаболомные маркеры пересекаются для БП и СП, молекулярные механизмы нарушения их метаболизма могут иметь разную природу у двух заболеваний. Известно, что генетические маркеры заболеваний играют ключевую роль в патологических процессах. В связи с этим генетические маркеры также могут участвовать в нарушении метаболизма обнаруженных нами аминокислот и ацилкарнитинов.

Для проверки данной гипотезы мы реконструировали молекулярно-генетические пути, описывающие регуляторные связи к ферментам биосинтеза и деградации значимых метаболитов от генетических маркеров исследуемых заболеваний. Оказалось, что генетические маркеры заболеваний активно вовлечены в регуляцию функций ферментов и экспрессии кодирующих их генов (см. Приложения 5 и 6). Были составлены три группы генетических маркеров: специфические для БП, специфические для СП и общие для обоих заболеваний. Для выявления специфических молекулярных механизмов нарушенной регуляции метаболомных профилей при БП и СП мы рассмотрели регуляторные пути, в которых первым звеном выступают специфические генетические маркеры заболеваний. В свою очередь регуляторные пути, в которых принимают участие неспецифические генетические маркеры (общие для БП и СП), могут определять общие для БП и СП механизмы нарушения метаболомных профилей. В реконструированных генных сетях (см. Приложения 5 и 6) мы выделили сети регуляции конкретных ферментов метаболизма аминокислот и ацилкарнитинов, которые в научной литературе были исследованы в контексте болезни Паркинсона и сосудистого паркинсонизма.

Генные сети регуляции ферментов превращения аминокислот

Альдегиддегидрогеназа 2 (ALDH2) оказалась среди ферментов с наибольшим числом регуляторных связей от соответствующих генетических маркеров БП и СП (рис. 6, *а*). ALDH2 участвует в путях метаболизма пролина, аланина, жирных кислот. Это основной фермент, участвующий в метаболизме альдегидов и цитотоксических метаболитов. Роль ALDH2 в головном мозге заключается в предотвращении «альдегидной нагрузки» – накоплении альдегидов, которые при окислительном стрессе могут образовывать связи с липидами, нуклеиновыми кислотами и белками, провоцируя нейротоксический эффект (Chen C.-H. et al., 2016). Из исследований связи альдегиддегидрогеназ с БП известно, что активность митохондриальной ALDH2 повышена в скорлупе головного мозга пациентов с БП (Michel et al., 2014). ALDH2 может защищать нейроны от токсического действия метаболитов дофамина (Chiu et al., 2015), а усиление активности ALDH2 способствует восстановлению функции нейронов, нарушенной под действием гипоксии (Lin et al., 2022).

К ферменту AL1B1 (митохондриальная альдегиддегидрогеназа X) было направлено наибольшее количество регуляторных связей от генетических маркеров БП (рис. 6, *б*). Этот фермент участвует в путях метаболизма пролина, аланина, жирных кислот, а также играет важную роль в детоксикации ацетальдегидов и метаболизме нейромедиаторов (Shortall et al., 2021). Согласно имеющимся данным, дефицит AL1B1 отмечен в черной субстанции пациентов с болезнью Паркинсона и ассоциирован с развитием патологии (Grünblatt, Riederer, 2016; Odongo et al., 2023). Это может привести к накоплению альдегидов, таких как 4-гидрокси-2-ноненаль (4-HNE), которые могут нарушить функцию митохондрий, индуцировать агрегацию альфа-синуклеина и вызывать нейровоспаление и апоптоз (Wey et al., 2012; Grünblatt, Riederer, 2016).

В генных сетях ферменты P4HA1 и P4HA2 оказались в большей степени под регуляторным воздействием общих для БП и СП генетических маркеров (см. рис. 6, *в*). Группа ферментов P4HA (пролил 4-гидроксилаза альфа) катализирует образование 4-гидроксипролина, который необходим для правильного трехмерного складывания вновь синтезированных цепей проколлагена (Song et al., 2023). Помимо этой функции, известно об участии P4HA1 в процессе постишемического ангиогенеза (Xu et al., 2024).

Ферменты BCAT1 и BCAT2 катализируют обратимое трансаминирование аминокислот с разветвленной боковой цепью (BCAA) и α-кетоглутарата с образованием соответствующих α-кетокислот с разветвленной цепью и глутамата. По результатам метаболомного анализа мы выявили у пациентов с БП и СП повышенное содержание таких BCAA, как валин и изолейцин. По данным реконструированных нами генных сетей, BCAT1 являлся одним из ферментов, к которым направлено наибольшее число регуляторных связей от генетических маркеров БП, а регуляторные связи к BCAT2 были направлены в большей степени от общих маркеров БП и СП (см. рис. 6, *г*). Дефектный метаболизм BCAA, который включает нарушения BCAT1, связан с несколькими основными особенностями БП, включая двигательную дисфункцию и нейродегенерацию (Yao et al., 2018; Sohrabi et al., 2021). На моделях болезни Паркинсона *C. elegans* нокдаун *bcat1* приводил к истощению метаболитов цикла трикарбоновых кислот и к «митохондриальной гиперактивности», что в результате приводит к окислительному повреждению в нейронах (Mor et al., 2020). Кроме того, метаанализ полногеномных ассоциаций выявил связи между болезнью Паркинсона и генами ферментов метаболизма BCAA (Nalls et al., 2014). Нарушения ферментов метаболизма аминокислот с разветвленной боковой цепью были най-

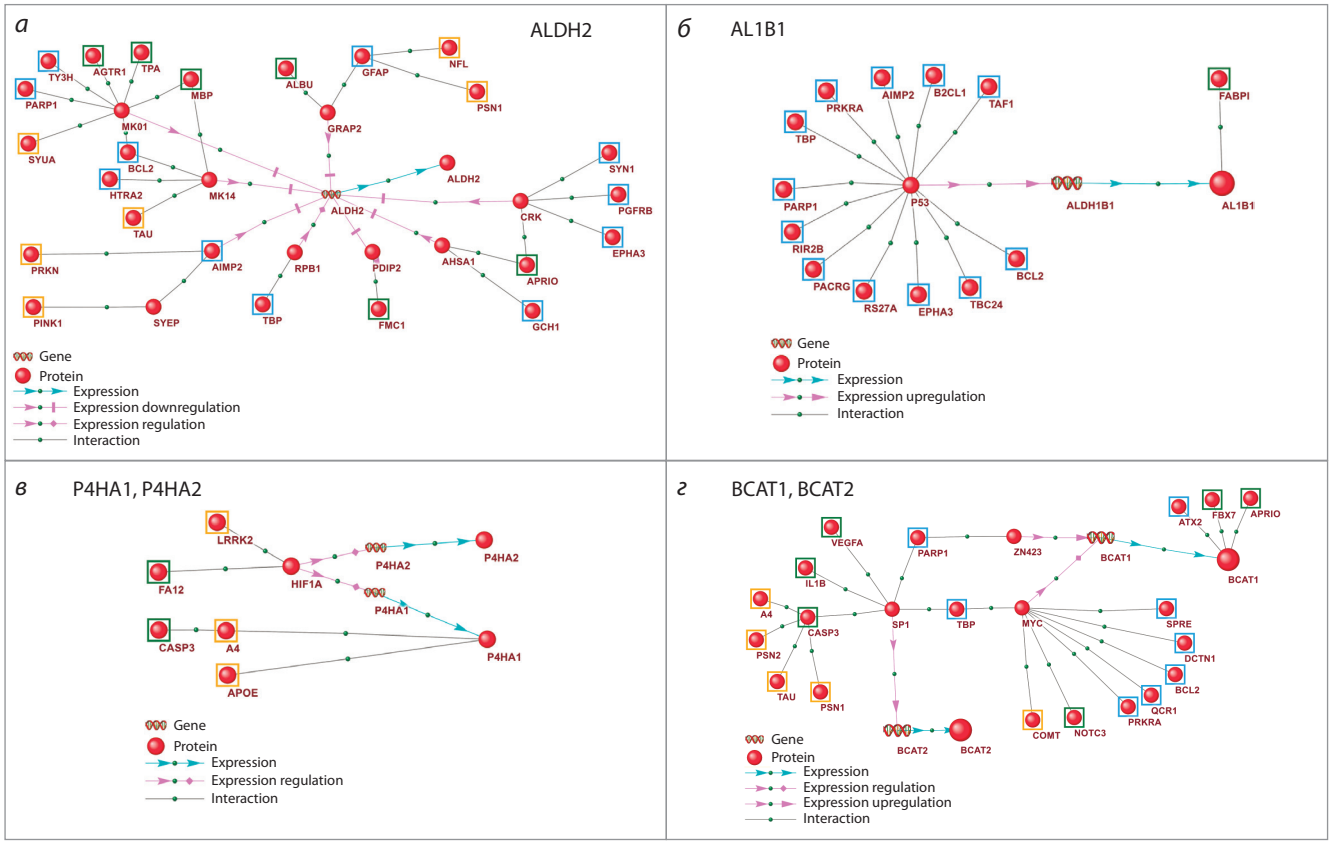


Рис. 6. Генные сети регуляции ферментов метаболизма аминокислот генетическими маркерами БП, СП и общими маркерами двух заболеваний.

Рамками показаны генетические маркеры: БП (голубые рамки), СП (зеленые), общие маркеры двух заболеваний (оранжевые).

дены и при васкулярной деменции. Так, повышенная экспрессия мРНК цитозольных и митохондриальных BCAT обнаружена в образцах коры мозга пациентов с васкулярной деменцией, что предположительно может защищать клетки от нейротоксического эффекта избытка глутамата (Ashby et al., 2017).

Генные сети регуляции ферментов превращения ацилкарнитинов

В обеих группах пациентов с патологиями мы идентифицировали изменение профиля ацилкарнитинов, являющихся важным звеном энергетических процессов в клетке. Поскольку ацилкарнитины – основные переносчики жирных кислот на внутреннюю мембрану митохондрий, метаболизм ацилкарнитинов и жирных кислот тесно взаимосвязан. Синтаза жирных кислот FAS катализирует элонгацию жирных кислот начиная от ацетил-КоА и малонил-КоА. На генной сети регуляции ферментов метаболизма ацилкарнитинов (см. Приложение 6) к гену, кодирующему фермент FAS, направлено наибольшее количество связей «регуляция экспрессии» от генетических маркеров БП, СП и общих для обоих заболеваний маркеров (рис. 7, *а*). Следует отметить генетический маркер PINK1, который ассоциирован с митохондриальной дисфункцией при БП (Narendra et al., 2010). Мутации в PINK1 приводят к дефициту этого белка при БП (Valente et al., 2004). Показано, что ингибирование FAS в PINK1-мутантных моделях восстанавливает метаболические процессы в митохондриях и снижает уровень пальмитата (Vos et al., 2017). Кроме того, известно о роли FAS в процессах миелинизации и ремиелинизации центральной нервной системы (Dimas et al., 2019).

К ферменту CPT1 (карнитин-пальмитоилтрансфераза 1), согласно реконструкции генных сетей, было направлено множество регуляторных связей от генетических маркеров, специфических как для БП, так и для СП (см. рис. 7, *б*). CPT1 – белок-транспортер, расположенный на внешней мембране митохондрий и в клетках млекопитающих, существующий в трех изоформах: CPT1A, CPT1B и CPT1C. При этом CPT1A в большей степени специфичен для липогенных тканей, например для печени, в то время как CPT1B преобладает в тканях с высокой окислительной способностью жирных кислот (сердце и скелетные мышцы), а CPT1C – в нейрональной ткани (Wang Muyun et al., 2021). Ферменты семейства CPT1 катализируют процесс переноса ацил-КоА групп с длиной цепи от C12 до C18 на L-карнитин с формированием ацилкарнитинов (Schlaepfer, Joshi, 2020). Ингибирование липидного обмена, регулируемого CPT1A, на мышиных моделях болезни Паркинсона показало многообещающие результаты в виде улучшения двигательных и сенсорно-моторных функций (Trabjerg et al., 2023). Описана роль CPT1 в развитии инсулинорезистентности – состояния, ассоциированного с нарушением функции черной субстанции (Virmani et al.,

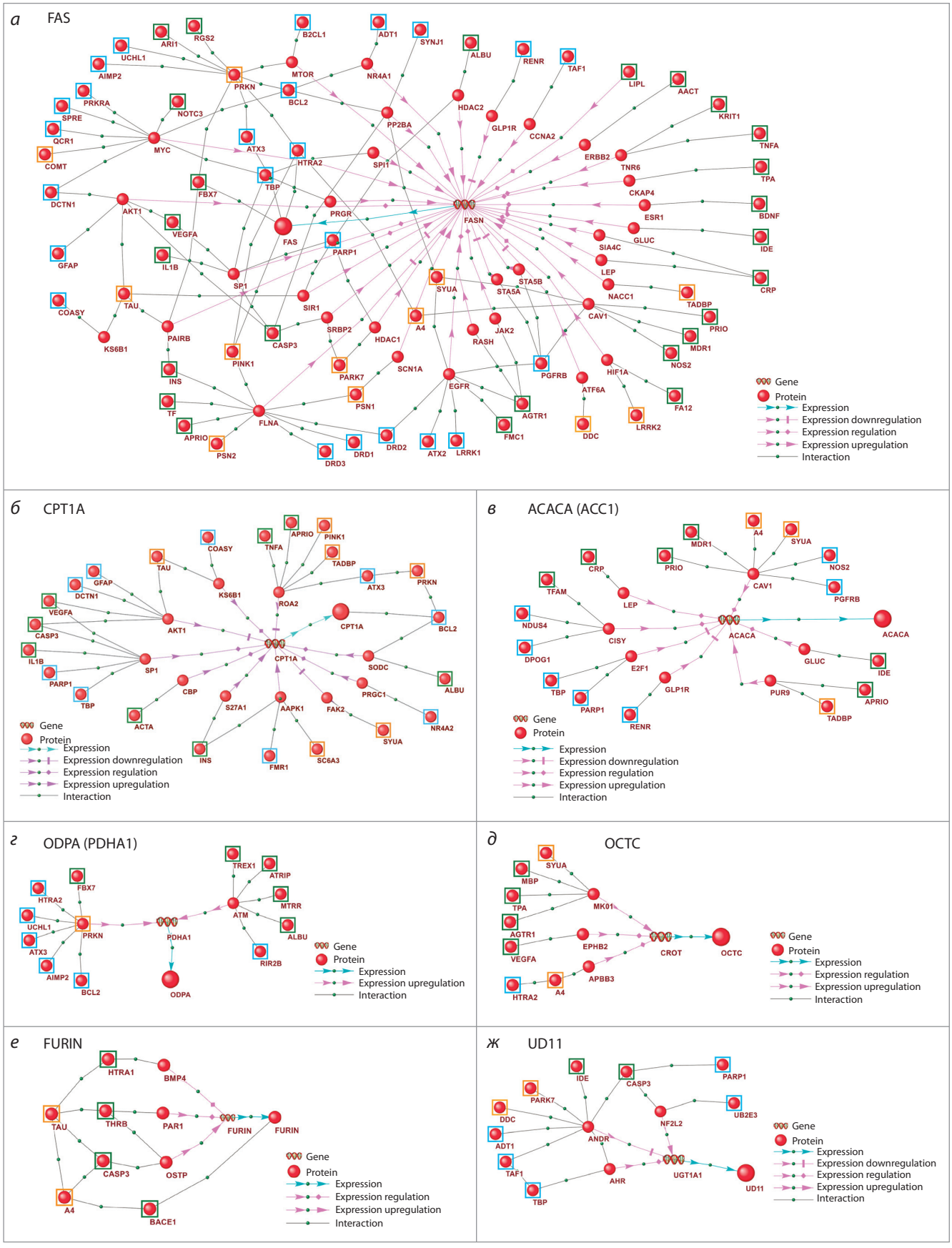


Рис. 7. Генные сети регуляции ферментов метаболизма ацилкарнитинов генетическими маркерами БП, СП и общими маркерами двух заболеваний.

Рамками показаны генетические маркеры: БП (голубые рамки), СП (зеленые), общие маркеры двух заболеваний (оранжевые).

2015). У пациентов с ранней стадией БП было идентифицировано снижение уровня длинноцепочечных ацилкарнитинов (C14–C18), предположительно связанное с дефицитом фермента CPT1 (Saiki et al., 2017).

По результатам, полученным в ходе реконструкции генных сетей, регуляторные связи к ферментам ACC1 и ODPA (PDHA1) были характерны как для БП, так и для СП (см. рис. 7, *в*, *г*). ACC1 (ацетил-коА карбоксилаза 1, ACACA) является лимитирующим ферментом в *de novo* синтезе жирных кислот, который обеспечивает превращение ацетил-КоА в малонил-КоА (Wang Y. et al., 2022). На экспериментальной модели БП было показано, что взаимодействие фосфорилированного альфа-синуклеина с ACC1 связано с низким уровнем АТФ, окислительным стрессом и митохондриальной дисфункцией (Grassi et al., 2018). Фермент PDHA1 (пируват дегидрогеназа E1 альфа, ODPA) – важный компонент комплекса, катализирующего декарбоксилирование пирувата до ацетил-КоА (Børglum et al., 1996). При стрессовых условиях ингибирование PDHA1 позволяет астроцитам использовать анаэробный гликолиз, вследствие чего увеличивается потребление нейронами лактата, осуществляется экономия глюкозы и защита от окислительного стресса (de Holanda Paranhos et al., 2024). Таким образом, PDHA1 выступает в роли посредника между цитозольным гликолизом и митохондриальным окислительным фосфорилированием (Pavlú-Pereira et al., 2024). В исследовании (Miki et al., 2017) продемонстрировано участие PDHA1 в качестве компонента телец Леви при идиопатической БП и PARK14-связанном паркинсонизме (семейная форма БП). Кроме того, показано снижение уровня белка PDHA1 в таких структурах головного мозга, как скорлупа и черная субстанция, у пациентов с идиопатической БП.

Согласно генным сетям, более специфичными для генетических маркеров СП оказались регуляторные связи к ферментам OCTC и FURIN (см. рис. 7, *д*, *е*). Кодируемый геном *CROT* фермент OCTC (пероксисомальная карнитин О-октаноилтрансфераза) участвует в транспорте из пероксисом средне- и длинноцепочечных ацил-КоА, которые задействованы в бета-окислении жирных кислот. Показано, что OCTC ассоциирован с кальцификацией гладкомышечных клеток артерий – высокое содержание OCTC было обнаружено вблизи кальцифицированных участков бляшек (Okui et al., 2021).

Фурин (Furin, PACE) – сериновая конвертаза, принимающая участие в процессах атерогенеза. Повышение активности фурина ассоциировано с развитием сердечно-сосудистых заболеваний (Wichaiyo et al., 2024). Ингибирование фурина замедляет прогрессирование атеросклеротического поражения у мышей (Yakala et al., 2019). Кроме того, известны эффекты фурина на нейрональную ткань. В частности, фурин стимулирует превращение нейротрофического фактора головного мозга (BDNF) из pro-BDNF в зрелую форму, что может играть роль в развитии нейродегенеративных заболеваний (Wang Mingyue et al., 2021). Ингибиторы фурина могут предотвращать повреждение нейронов, вызванное обработкой NMDA (Yamada et al., 2018). Было продемонстрировано, что фурин действует на фактор роста нейронов (NGF), способствуя переходу из pro-NGF в β-NGF, который, в свою очередь, воздействует на гладкомышечные клетки сосудов (Urban et al., 2013). Описана также роль фурина при БП на моделях болезни Паркинсона. Так, у дрозофил нокдаун *Furin1* уменьшал потерю дофаминергических нейронов, вызванную мутациями LRRK2 (Maksoud et al., 2019). Также в экспериментах на дрозофилах идентифировано, что фурин необходим при кэп-зависимой трансляции LRRK2, что влияет на постсинаптическую передачу (Penney et al., 2016).

Регуляторные связи к ферменту UD11 в большей степени реализовались от генетических маркеров БП (см. рис. 7, *ж*). UDP-глюкуронозилтрансферазы – ферменты, участвующие в детоксикации различных соединений путем глюкуронирования субстратов, что облегчает их выведение из организма (Tukey, Strassburg, 2000). В контексте исследований БП и СП UDP-глюкуронозилтрансферазы мало изучены. Отметим работу, в которой продемонстрирована ассоциация генотипа UDP-глюкуронозилтрансферазы 1A9 с побочными реакциями на ингибиторы катехол-О-метилтрансферазы у пациентов с БП (Ferrari et al., 2012).

Таким образом, причиной изменения метаболомного профиля аминокислот и ацилкарнитинов для болезни Паркинсона и сосудистого паркинсонизма могут быть различные молекулярно-генетические механизмы. Специфическими для БП в нашей работе были отмечены пути регуляции ферментов ALDH2, BCAT1, AL1B1, UD11, а для СП – пути регуляции ферментов OCTC, FURIN, S22A6. Со стороны общих для БП и СП генетических маркеров, согласно генным сетям, наибольшее регуляторное воздействие было направлено к ферментам BCAT2, ODPB, P4HA1. При обоих заболеваниях нарушалась регуляция липидного обмена, метаболизма валина и изолейцина, а также обнаружено вовлечение множества механизмов, приводящих к процессам окислительного стресса и митохондриальной дисфункции.

Заключение

Для выявления специфических молекулярно-генетических механизмов при БП и СП мы реконструировали генные сети, описывающие регуляцию ферментов метаболизма потенциальных маркеров БП, СП, включающих ряд аминокислот (аланин, пролин, валин, изолейцин, метионин) и 17 ацилкарнитинов, идентифицированных методом ВЭЖХ-МС/МС. Сравнительный анализ регуляторных путей в генных сетях позволил выявить специфические и неспецифические для БП и СП молекулярные механизмы, ассоциированные с изменением метаболомного профиля при этих патологиях. Полученные результаты описывают особенности молекулярно-генетических механизмов БП и СП, а также могут быть использованы для разработки и применения диагностических систем на основе оценки метаболомного профиля аминокислот и ацилкарнитинов в плазме крови пациентов. Отметим, что в нашей работе впервые проанализирован с применением генных сетей метаболомный профиль аминокислот и ацилкарнитинов у пациентов с сосудистым паркинсонизмом и болезнью Паркинсона, что представляет важность в сравнительном изучении этих заболеваний.

Список литературы / References

Левин О.С., Боголепова А.Н., Лобзин В.Ю. Общие механизмы патогенеза нейроденеративных и цереброваскулярных заболеваний и возможности их коррекции. *Журн. неврологии и психиатрии им. С.С. Корсакова*. 2022;122(5):11-16. doi 10.17116/jnevro202212205111
[Levin O.S., Bogolepova A.N., Lobzin V.Yu. General mechanisms of the pathogenesis of neurodenerative and cerebrovascular diseases and the possibilities of their correction. *Zhurnal Nevrologii i Psikhiatrii Imeni S.S. Korsakova = S.S. Korsakov Journal of Neurology and Psychiatry*. 2022;122(5):11-16. doi 10.17116/jnevro202212205111 (in Russian)]

Alexander G.E. Biology of Parkinson's disease: pathogenesis and pathophysiology of a multisystem neurodegenerative disorder. *Dialogues Clin. Neurosci*. 2004;6(3):259-280. doi 10.31887/DCNS.2004.6.3/galexander

Ashby E.L., Kierzkowska M., Hull J., Kehoe P.G., Hutson S.M., Conway M.E. Altered expression of human mitochondrial branched chain aminotransferase in dementia with Lewy bodies and vascular dementia. *Neurochem. Res*. 2017;42(1):306-319. doi 10.1007/s11064-016-1855-7

Binder H., Wirth H., Arakelyan A., Lembcke K., Tiys E.S., Ivanisenko V.A., Kolchanov N.A., Kononikhin A., Popov I., Nikolaev E.N., Pastushkova L.K., Larina I.M. Time-course human urine proteomics in space-flight simulation experiments. *BMC Genomics*. 2014;15(S12):S2. doi 10.1186/1471-2164-15-S12-S2

Børglum A.D., Flint T., Hansen L.L., Kruse T.A. Refined localization of the pyruvate dehydrogenase E1α gene (PDHA1) by linkage analysis. *Hum. Genet*. 1996;99(1):80-82. doi 10.1007/s004390050315

Braak H., Tredici K.D., Rüb U., De Vos R.A.I., Jansen Steur E.N.H., Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging*. 2003;24(2):197-211. doi 10.1016/S0197-4580(02)00065-9

Bragina E.Yu., Tiys E.S., Freidin M.B., Koneva L.A., Demenkov P.S., Ivanisenko V.A., Kolchanov N.A., Puzyrev V.P. Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis. *Immunogenetics*. 2014;66(7-8):457-465. doi 10.1007/s00251-014-0786-1

Bragina E.Yu., Tiys E.S., Rudko A.A., Ivanisenko V.A., Freidin M.B. Novel tuberculosis susceptibility candidate genes revealed by the reconstruction and analysis of associative networks. *Infect. Genet. Evol*. 2016;46:118-123. doi 10.1016/j.meegid.2016.10.030

Bragina E.Yu., Gomboeva D.E., Saik O.V., Ivanisenko V.A., Freidin M.B., Nazarenko M.S., Puzyrev V.P. Apoptosis genes as a key to identification of inverse comorbidity of Huntington's disease and cancer. *Int. J. Mol. Sci*. 2023;24(11):9385. doi 10.3390/ijms24119385

Che Mohd Nassir C.M.N., Damodaran T., Yusof S.R., Norazit A., Chilla G., Huen I., Kn B.P., Mohamed Ibrahim N., Mustapha M. Aberrant neurogliovascular unit dynamics in cerebral small vessel disease: a rheological clue to vascular Parkinsonism. *Pharmaceutics*. 2021;13(8):1207. doi 10.3390/pharmaceutics13081207

Chen C.-H., Joshi A.U., Mochly-Rosen D. The role of mitochondrial aldehyde dehydrogenase 2 (ALDH2) in neuropathology and neurodegeneration. *Acta Neurol. Taiwan*. 2016;25(4)(4):111-123

Chen Y., Liu Q., Liu J., Wei P., Li B., Wang N., Liu Z., Wang Z. Revealing the modular similarities and differences among Alzheimer's disease, vascular dementia, and Parkinson's disease in genomic networks. *Neuromol. Med*. 2022;24(2):125-138. doi 10.1007/s12017-021-08670-2

Chen Y.-F., Tseng Y.-L., Lan M.-Y., Lai S.-L., Su C.-S., Liu J.-S., Chang Y.-Y. The relationship of leukoaraiosis and the clinical severity of vascular Parkinsonism. *J. Neurol. Sci*. 2014;346(1-2):255-259. doi 10.1016/j.jns.2014.09.002

Chiu C.-C., Yeh T.-H., Lai S.-C., Wu-Chou Y.-H., Chen C.-H., Mochly-Rosen D., Huang Y.-C., Chen Y.-J., Chen C.-L., Chang Y.-M., Wang H.-L., Lu C.-S. Neuroprotective effects of aldehyde dehydrogenase 2 activation in rotenone-induced cellular and animal models of parkinsonism. *Exp. Neurol*. 2015;263:244-253. doi 10.1016/j.expneurol.2014.09.016

Dalangin R., Kim A., Campbell R.E. The role of amino acids in neurotransmission and fluorescent tools for their detection. *Int. J. Mol. Sci*. 2020;21(17):6197. doi 10.3390/ijms21176197

De Holanda Paranhos L., Magalhães R.S.S., De Araújo Brasil A., Neto J.R.M., Ribeiro G.D., Queiroz D.D., Dos Santos V.M., Eleutherio E.C.A. The familial amyotrophic lateral sclerosis-associated A4V SOD1 mutant is not able to regulate aerobic glycolysis. *Biochim. Biophys. Acta Gen. Subjt.* 2024;1868(8):130634. doi 10.1016/j.bbagen.2024.130634

Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol*. 2012;11(3-4):149-161. doi 10.3233/ISB-2012-0449

Dimas P., Montani L., Pereira J.A., Moreno D., Trötzmüller M., Gerber J., Semenkovich C.F., Köfeler H.C., Suter U. CNS myelination and remyelination depend on fatty acid synthesis by oligodendrocytes. *eLife*. 2019;8:e44702. doi 10.7554/eLife.44702

Ferrari M., Martignoni E., Blandini F., Riboldazzi G., Bono G., Marino F., Cosentino M. Association of UDP-glucuronosyltransferase 1A9 polymorphisms with adverse reactions to catechol-O-methyltransferase inhibitors in Parkinson's disease patients. *Eur. J. Clin. Pharmacol*. 2012;68(11):1493-1499. doi 10.1007/s00228-012-1281-y

George G., Singh S., Lokappa S.B., Varkey J. Gene co-expression network analysis for identifying genetic markers in Parkinson's disease – a three-way comparative approach. *Genomics*. 2019a;111(4):819-830. doi 10.1016/j.ygeno.2018.05.005

George G., Valiya Parambath S., Lokappa S.B., Varkey J. Construction of Parkinson's disease marker-based weighted protein-protein interaction network for prioritization of co-expressed genes. *Gene*. 2019b;697:67-77. doi 10.1016/j.gene.2019.02.026

Grassi D., Howard S., Zhou M., Diaz-Perez N., Urban N.T., Guerrero-Given D., Kamasawa N., Volpicelli-Daley L.A., LoGrasso P., Lasmézas C.I. Identification of a highly neurotoxic α-synuclein species inducing mitochondrial damage and mitophagy in Parkinson's disease. *Proc. Natl. Acad. Sci. USA.* 2018;115(11):E2634-E2643. doi 10.1073/pnas.1713849115

Grünblatt E., Riederer P. Aldehyde dehydrogenase (ALDH) in Alzheimer's and Parkinson's disease. *J. Neural. Transm*. 2016;123(2):83-90. doi 10.1007/s00702-014-1320-1

Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*. 2020;21(S11):228. doi 10.1186/s12859-020-03557-8

Ivanisenko T.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. The new version of the ANDDigest tool with improved ai-based short names recognition. *Int. J. Mol. Sci*. 2022;23(23):14934. doi 10.3390/ijms232314934

Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol*. 2015;9(Suppl.2):S2. doi 10.1186/1752-0509-9-S2-S2

Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(S1):34. doi 10.1186/s12859-018-2567-6

Ivanisenko V.A., Gaisler E.V., Basov N.V., Rogachev A.D., Cheresiz S.V., Ivanisenko T.V., Demenkov P.S., Mishchenko E.L., Khripko O.P., Khripko Yu.I., Voevoda S.M., Karpenko T.N., Velichko A.J., Voevoda M.I., Kolchanov N.A., Pokrovsky A.G. Plasma metabolomics and gene regulatory networks analysis reveal the role of nonstructural SARS-CoV-2 viral proteins in metabolic dysregulation

in COVID-19 patients. *Sci. Rep*. 2022;12(1):19977. doi 10.1038/s41598-022-24170-0

Ivanisenko V.A., Basov N.V., Makarova A.A., Venzel A.S., Rogachev A.D., Demenkov P.S., Ivanisenko T.V., Kleshchev M.A., Gaisler E.V., Moroz G.B., Plesko V.V., Sotnikova Y.S., Patrushev Y.V., Lomivorotov V.V., Kolchanov N.A., Pokrovsky A.G. Gene networks for use in metabolomic data analysis of blood plasma from patients with postoperative delirium. *Vavilov J. Genet. Breed*. 2023;27(7):768-775. doi 10.18699/VJGB-23-89

Jones L.L., McDonald D.A., Borum P.R. Acylcarnitines: role in brain. *Prog. Lipid Res*. 2010;49(1):61-75. doi 10.1016/j.plipres.2009.08.004

Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi 10.1093/nar/28.1.27

Kasakin M.F., Rogachev A.D., Predtechenskaya E.V., Zaigraev V.J., Koval V.V., Pokrovsky A.G. Targeted metabolomics approach for identification of relapsing-remitting multiple sclerosis markers and evaluation of diagnostic models. *Med. Chem. Commun.* 2019;10(10):1803-1809. doi 10.1039/c9md00253g

Korczyn A.D. Vascular Parkinsonism – characteristics, pathogenesis and treatment. *Nat. Rev. Neurol*. 2015;11(6):319-326. doi 10.1038/nrneurol.2015.61

Larina I.M., Pastushkova L.Kh., Tiys E.S., Kireev K.S., Kononikhin A.S., Starodubtseva N.L., Popov I.A., Custaud M.-A., Dobrokhotov I.V., Nikolaev E.N., Kolchanov N.A., Ivanisenko V.A. Permanent proteins in the urine of healthy humans during the Mars-500 experiment. *J. Bioinform. Comput. Biol.* 2015;13(01):1540001. doi 10.1142/S0219720015400016

Lin L., Tao J.-P., Li M., Peng J., Zhou C., Ouyang J., Si Y.-Y. Mechanism of ALDH2 improves the neuronal damage caused by hypoxia/reoxygenation. *Eur. Rev. Med. Pharmacol. Sci.* 2022;26(8):2712-2720. doi 10.26355/eurrev_202204_28601

Maksoud E., Liao E.H., Haghighi A.P. A neuron-glial trans-signaling cascade mediates LRRK2-induced neurodegeneration. *Cell Rep*. 2019;26(7):1774-1786.e4. doi 10.1016/j.celrep.2019.01.077

Mercatelli D., Scalambra L., Triboli L., Ray F., Giorgi F.M. Gene regulatory network inference resources: a practical overview. *Biochim. Biophys. Acta Gene Regul. Mech.* 2020;1863(6):194430. doi 10.1016/j.bbagrm.2019.194430

Michel T.M., Käsbauer L., Gsell W., Jecel J., Sheldrick A.J., Cortese M., Nickl-Jockschat T., Grünblatt E., Riederer P. Aldehyde dehydrogenase 2 in sporadic Parkinson's disease. *Parkinsonism Relat. Disord*. 2014;20:S68-S72. doi 10.1016/S1353-8020(13)70018-X

Miki Y., Tanji K., Mori F., Kakita A., Takahashi H., Wakabayashi K. Alteration of mitochondrial protein PDHA1 in Lewy body disease and PARK14. *Biochem. Biophys. Res. Commun*. 2017;489(4):439-444. doi 10.1016/j.bbrc.2017.05.162

Mor D.E., Sohrabi S., Kaletsky R., Keyes W., Tartici A., Kalia V., Miller G.W., Murphy C.T. Metformin rescues Parkinson's disease phenotypes caused by hyperactive mitochondria. *Proc. Natl. Acad. Sci. USA.* 2020;117(42):26438-26447. doi 10.1073/pnas.2009838117

Nalls M.A., Pankratz N., Lill C.M., Do C.B., Hernandez D.G., Saad M., DeStefano A.L., Kara E., Bras J., Sharma M., … Brice A., Scott W.K., Gasser T., Bertram L., Eriksson N., Foroud T., Singleton A.B. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet*. 2014;46(9):989-993. doi 10.1038/ng.3043

Narasimhan M., Schwartz R., Halliday G. Parkinsonism and cerebrovascular disease. *J. Neurol. Sci*. 2022;433:120011. doi 10.1016/j.jns.2021.120011

Narendra D.P., Jin S.M., Tanaka A., Suen D.-F., Gautier C.A., Shen J., Cookson M.R., Youle R.J. PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS Biol*. 2010;8(1):e1000298. doi 10.1371/journal.pbio.1000298

Odongo R., Bellur O., Abdik E., Çakır T. Brain-wide transcriptome-based metabolic alterations in Parkinson's disease: human inter-region and human-experimental model correlations. *Mol. Omics*. 2023;19(7):522-537. doi 10.1039/D2MO00343K

Okui T., Iwashita M., Rogers M.A., Halu A., Atkins S.K., Kuraoka S., Abdelhamid I., Higashi H., Ramsaroop A., Aikawa M., Singh S.A., Aikawa E. CROT (Carnitine O-Octanoyltransferase) is a novel contributing factor in vascular calcification via promoting fatty acid metabolism and mitochondrial dysfunction. *Arterioscler. Thromb. Vasc. Biol*. 2021;41(2):755-768. doi 10.1161/ATVBAHA.120.315007

Ostrakhovitch E.A., Song E.-S., Macedo J.K.A., Gentry M.S., Quintero J.E., Van Horne C., Yamasaki T.R. Analysis of circulating metabolites to differentiate Parkinson's disease and essential tremor. *Neurosci. Lett*. 2022;769:136428. doi10.1016/j.neulet.2021.136428

Pastushkova L.Kh., Kireev K.S., Kononikhin A.S., Tiys E.S., Popov I.A., Starodubtseva N.L., Dobrokhotov I.V., Ivanisenko V.A., Larina I.M., Kolchanov N.A., Nikolaev E.N. Detection of renal tissue and urinary tract proteins in the human urine after space flight. *PLoS One*. 2013;8(8):e71652. doi 10.1371/journal.pone.0071652

Pastushkova L.Kh., Kashirina D.N., Brzhozovskiy A.G., Kononikhin A.S., Tiys E.S., Ivanisenko V.A., Koloteva M.I., Nikolaev E.N., Larina I.M. Evaluation of cardiovascular system state by urine proteome after manned space flight. *Acta Astronaut*. 2019;160:594-600. doi 10.1016/j.actaastro.2019.02.015

Pavlú-Pereira H., Florindo C., Carvalho F., Tavares De Almeida I., Vicente J., Morais V., Rivera I. Evaluation of mitochondrial function on pyruvate dehydrogenase complex deficient patient-derived cell lines. *Endocr. Metab. Immune Disord. Drug Targets*. 2024;24(16):20. doi 10.2174/0118715303280072231004082458

Penney J., Tsurudome K., Liao E.H., Kauwe G., Gray L., Yanagiya A., Calderon M.R., Sonenberg N., Haghighi A.P. LRRK2 regulates retrograde synaptic compensation at the *Drosophila* neuromuscular junction. *Nat. Commun*. 2016;7(1):12188. doi 10.1038/ncomms12188

Rappaport N., Twik M., Nativ N., Stelzer G., Bahir I., Stein T.I., Safran M., Lancet D. MalaCards: a comprehensive automatically-mined database of human diseases. *Curr. Protoc. Bioinformatics*. 2014;47(1):1.24.1-19. doi 10.1002/0471250953.bi0124s47

Rocha E.M., De Miranda B., Sanders L.H. Alpha-synuclein: pathology, mitochondrial dysfunction and neuroinflammation in Parkinson's disease. *Neurobiol. Dis*. 2018;109:249-257. doi 10.1016/j.nbd.2017.04.004

Rogachev A.D., Alemasov N.A., Ivanisenko V.A., Ivanisenko N.V., Gaisler E.V., Oleshko O.S., Cheresiz S.V., Mishinov S.V., Stupak V.V., Pokrovsky A.G. Correlation of metabolic profiles of plasma and cerebrospinal fluid of high-grade glioma patients. *Metabolites*. 2021;11(3):133. doi 10.3390/metabo11030133

Saik O.V., Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Interactome of the hepatitis C virus: literature mining with ANDSystem. *Virus Res*. 2016;218:40-48. doi 10.1016/j.virusres.2015.12.003

Saik O.V., Demenkov P.S., Ivanisenko T.V., Bragina E.Yu., Freidin M.B., Dosenko V.E., Zolotareva O.I., Choynzonov E.L., Hofestaedt R., Ivanisenko V.A. Search for new candidate genes involved in the comorbidity of asthma and hypertension based on automatic analysis of scientific literature. *J. Integr. Bioinform*. 2018;15(4):20180054. doi 10.1515/jib-2018-0054

Saik O.V., Nimaev V.V., Usmonov D.B., Demenkov P.S., Ivanisenko T.V., Lavrik I.N., Ivanisenko V.A. Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema using an analysis of associative gene networks with ANDSystem. *BMC Med. Genomics*. 2019;12(S2):47. doi 10.1186/s12920-019-0492-9

Saiki S., Hatano T., Fujimaki M., Ishikawa K.-I., Mori A., Oji Y., Okuzumi A., Fukuhara T., Koinuma T., Imamichi Y., Nagumo M., Furuya N., Nojiri S., Amo T., Yamashiro K., Hattori N. Decreased long-chain acylcarnitines from insufficient β-oxidation as potential early diagnostic markers for Parkinson's disease. *Sci. Rep*. 2017;7(1):7328. doi 10.1038/s41598-017-06767-y

Schlaepfer I.R., Joshi M. CPT1A-mediated fat oxidation, mechanisms, and therapeutic potential. *Endocrinology*. 2020;161(2):bqz046. doi 10.1210/endocr/bqz046

Shortall K., Djeghader A., Magner E., Soulimane T. Insights into aldehyde dehydrogenase enzymes: a structural perspective. *Front. Mol. Biosci.* 2021;8:659550. doi 10.3389/fmolb.2021.659550

Sohrabi S., Mor D.E., Kaletsky R., Keyes W., Murphy C.T. High-throughput behavioral screen in *C. elegans* reveals Parkinson's disease drug candidates. *Commun. Biol*. 2021;4(1):203. doi 10.1038/s42003-021-01731-z

Song M., Schnettler E., Venkatachalam A., Wang Y., Feldman L., Argenta P., Rodriguez-Rodriguez L., Ramakrishnan S. Increased expression of collagen prolyl hydroxylases in ovarian cancer is associated with cancer growth and metastasis. *Am. J. Cancer Res*. 2023;13(12):6051-6062

Thanvi B., Lo N., Robinson T. Vascular parkinsonism – an important cause of parkinsonism in older people. *Age Ageing*. 2005;34(2):114-119. doi 10.1093/ageing/afi025

Tomkins J.E., Manzoni C. Advances in protein-protein interaction network analysis for Parkinson's disease. *Neurobiol. Dis*. 2021;155:105395. doi 10.1016/j.nbd.2021.105395

Trabjerg M.S., Andersen D.C., Huntjens P., Mørk K., Warming N., Kullab U.B., Skjønnemand M.-L.N., Oklinski M.K., Oklinski K.E., Bolther L., Kroese L.J., Pritchard C.E.J., Huijbers I.J., Corthals A., Søndergaard M.T., Kjeldal H.B., Pedersen C.F.M., Nieland J.D.V. Inhibition of carnitine palmitoyl-transferase 1 is a potential target in a mouse model of Parkinson's disease. *NPJ Parkinsons Dis.* 2023;9(1):6. doi 10.1038/s41531-023-00450-y

Tukey R.H., Strassburg C.P. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol.* 2000;40(1):581-616. doi 10.1146/annurev.pharmtox.40.1.581

Urban D., Lorenz J., Meyborg H., Ghosh S., Kintscher U., Kaufmann J., Fleck E., Kappert K., Stawowy P. Proprotein convertase furin enhances survival and migration of vascular smooth muscle cells via processing of pro-nerve growth factor. *J. Biochem*. 2013;153(2):197-207. doi 10.1093/jb/mvs137

Vale T.C., Barbosa M.T., Caramelli P., Cardoso F. Vascular Parkinsonism and cognitive impairment: literature review, Brazilian studies and case vignettes. *Dement. Neuropsychol.* 2012;6(3):137-144. doi 10.1590/S1980-57642012DN06030005

Valente E.M., Abou-Sleiman P.M., Caputo V., Muqit M.M.K., Harvey K., Gispert S., Ali Z., Del Turco D., Bentivoglio A.R., Healy D.G., Albanese A., Nussbaum R., González-Maldonado R., Deller T., Salvi S., Cortelli P., Gilks W.P., Latchman D.S., Harvey R.J., Dallapiccola B., Auburger G., Wood N.W. Hereditary early-onset Parkinson's disease caused by mutations in *PINK1*. *Science*. 2004;304(5674):1158-1160. doi 10.1126/science.1096284

Virmani A., Pinto L., Bauermann O., Zerelli S., Diedenhofen A., Binienda Z.K., Ali S.F., Van Der Leij F.R. The carnitine palmitoyl transferase (CPT) system and possible relevance for neuropsychiatric and neurological conditions. *Mol. Neurobiol*. 2015;52(2):826-836. doi 10.1007/s12035-015-9238-7

Vos M., Geens A., Böhm C., Deaulmerie L., Swerts J., Rossi M., Craessaerts K., Leites E.P., Seibler P., Rakovic A., Lohnau T., De Strooper B., Fendt S.-M., Morais V.A., Klein C., Verstreken P. Cardiolipin promotes electron transport between ubiquinone and complex I to rescue *PINK1* deficiency. *J. Cell Biol*. 2017;216(3):695-708. doi 10.1083/jcb.201511044

Wang Mingyue, Xie Y., Qin D. Proteolytic cleavage of proBDNF to mBDNF in neuropsychiatric and neurodegenerative diseases. *Brain Res. Bull*. 2021;166:172-184. doi 10.1016/j.brainresbull.2020.11.005

Wang Muyun, Wang K., Liao X., Hu H., Chen L., Meng L., Gao W., Li Q. Carnitine palmitoyltransferase system: a new target for anti-inflammatory and anticancer therapy? *Front. Pharmacol.* 2021;12:760581. doi 10.3389/fphar.2021.760581

Wang Yu, Yu W., Li S., Guo D., He J., Wang Yugang. Acetyl-CoA carboxylases and diseases. *Front. Oncol.* 2022;12:836058. doi 10.3389/fonc.2022.836058

Wey M.C.-Y., Fernandez E., Martinez P.A., Sullivan P., Goldstein D.S., Strong R. Neurodegeneration and motor dysfunction in mice lacking cytosolic and mitochondrial aldehyde dehydrogenases: implications for Parkinson's disease. *PLoS One*. 2012;7(2):e31522. doi 10.1371/journal.pone.0031522

Wichaiyo S., Koonyosying P., Morales N.P. Functional roles of furin in cardio-cerebrovascular diseases. *ACS Pharmacol. Transl. Sci.* 2024;7(3):570-585. doi 10.1021/acsptsci.3c00325

Wishart D.S., Guo A., Oler E., Wang F., Anjum A., Peters H., Dizon R., Sayeeda Z., Tian S., Lee B.L., Berjanskii M., Mah R., Yamamoto M., Jovel J., Torres-Calzada C., Hiebert-Giesbrecht M., Lui V.W., Varshavi Dorna, Varshavi Dorsa, Allen D., Arndt D., Khetarpal N., Sivakumaran A., Harford K., Sanford S., Yee K., Cao X., Budinski Z., Liigand J., Zhang L., Zheng J., Mandal R., Karu N., Dambrova M., Schiöth H.B., Greiner R., Gautam V. HMDB 5.0: the human metabolome database for 2022. *Nucleic Acids Res*. 2022;50(D1):D622-D631. doi 10.1093/nar/gkab1062

Wuolikainen A., Jonsson P., Ahnlund M., Antti H., Marklund S.L., Moritz T., Forsgren L., Andersen P.M., Trupp M. Multi-platform mass spectrometry analysis of the CSF and plasma metabolomes of rigorously matched amyotrophic lateral sclerosis, Parkinson's disease and control subjects. *Mol. BioSyst.* 2016;12(4):1287-1298. doi 10.1039/C5MB00711A

Xu Y., Xia D., Huang K., Liang M. Hypoxia-induced P4HA1 overexpression promotes post-ischemic angiogenesis by enhancing endothelial glycolysis through downregulating FBP1. *J. Transl. Med*. 2024;22(1):74. doi 10.1186/s12967-024-04872-x

Yakala G.K., Cabrera-Fuentes H.A., Crespo-Avilan G.E., Rattanasopa C., Burlacu A., George B.L., Anand K., Mayan D.C., Corlianò M., Hernández-Reséndiz S., Wu Z., Schwerk A.M.K., Tan A.L.J., Trigueros-Motos L., Chèvre R., Chua T., Kleemann R., Liehn E.A., Hausenloy D.J., Ghosh S., Singaraja R.R. FURIN inhibition reduces vascular remodeling and atherosclerotic lesion progression in mice. *Arterioscler. Thromb. Vasc. Biol.* 2019;39(3):387-401. doi 10.1161/ATVBAHA.118.311903

Yamada M., Hayashi H., Yuuki M., Matsushima N., Yuan B., Takagi N. Furin inhibitor protects against neuronal cell death induced by activated NMDA receptors. *Sci. Rep*. 2018;8(1):5212. doi 10.1038/s41598-018-23567-0

Yao V., Kaletsky R., Keyes W., Mor D.E., Wong A.K., Sohrabi S., Murphy C.T., Troyanskaya O.G. An integrative tissue-network approach to identify and test human disease genes. *Nat. Biotechnol*. 2018;36(11):1091-1099. doi 10.1038/nbt.4246

Zhao H., Wang C., Zhao N., Li W., Yang Z., Liu X., Le W., Zhang X. Potential biomarkers of Parkinson's disease revealed by plasma metabolic profiling. *J. Chromatogr. B*. 2018;1081-1082:101-108. doi 10.1016/j.jchromb.2018.01.025

Zijlmans J.C.M., Thijssen H.O.M., Vogels O.J.M., Kremer H.P.H.M.P., Poels P.J.E., Schoonderwaldt H.C., Merx J.L., van't Hof M.A., Thien Th., Horstink M.W.I.M. MRI in patients with suspected vascular parkinsonism. *Neurology*. 1995;45(12):2183-2188. doi 10.1212/WNL.45.12.2183

Zolotareva O., Saik O.V., Könige C., Bragina E.Yu., Goncharova I.A., Freidin M.B., Dosenko V.E., Ivanisenko V.A., Hofestädt R. Comorbidity of asthma and hypertension may be mediated by shared genetic dysregulation and drug side effects. *Sci. Rep*. 2019;9(1):16302. doi 10.1038/s41598-019-52762-w

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 16.09.2024. После доработки 07.11.2024. Принята к публикации 08.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):940-949

doi 10.18699/vjgb-24-101

Онтологии в моделировании и анализе больших генетических данных

Н.Л. Подколодный [1, 2, 3, 4], О.А. Подколодная [1], В.А. Иванисенко [1, 4], М.А. Марченко [2, 3]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[4] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

pnl@bionet.nsc.ru

Аннотация. Для систематизации и эффективного использования огромного объема экспериментальных данных, накопленных в области биоинформатики и биомедицины, необходимы новые подходы, основанные на онтологиях, включая автоматизированные методы семантической интеграции гетерогенных экспериментальных данных, методы создания больших баз знаний и самоинтерпретируемые методы анализа больших разнородных данных на основе глубокого обучения. В статье кратко представлены особенности предметной области (биоинформатика, системная биология, биомедицина), формальные определения понятия онтологии и графов знаний, приведены примеры применения онтологий для семантической интеграции гетерогенных данных и создания больших баз знаний, а также интерпретации результатов глубокого обучения на больших данных. В качестве примера успешного проекта описана база знаний Gene Ontology, которая помимо терминологических знаний и аннотаций генов (GOA) включает модели причинных влияний (GO-CAM). Это делает ее полезной не только для геномной биологии, но и для системной биологии, а также для интерпретации крупномасштабных экспериментальных данных. Обсуждается подход к созданию больших онтологий с использованием шаблонов проектирования на примере онтологии биологических атрибутов (OBA). Здесь большая часть классификации автоматически вычисляется на основе ранее созданных эталонных онтологий с помощью автоматизированного логического вывода, за исключением небольшого числа высокоуровневых понятий. Одной из основных проблем глубокого обучения является отсутствие интерпретируемости, поскольку нейронные сети часто функционируют как «черные ящики», не способные объяснить свои решения. В нашей статье описаны подходы к созданию методов интерпретации моделей глубокого обучения и представлены два примера самообъясняемых моделей глубокого обучения на основе онтологий. Модель Deep GONet, которая интегрирует Gene Ontology в иерархическую архитектуру нейронной сети, где каждый нейрон представляет биологическую функцию. Эксперименты с наборами данных диагностики рака показывают, что Deep GONet легко интерпретируется и обладает высокой производительностью для различения раковых и нераковых образцов. Модель ONN4MST, использующая онтологии биома для отслеживания микробных источников образцов, ниши которых ранее были мало изучены или неизвестны, и обнаружения микробных загрязнителей. ONN4MST может отличать образцы от онтологически близких биомов и, таким образом, предлагает количественный способ охарактеризовать развитие микробного сообщества кишечника человека. Оба примера демонстрируют высокую производительность и интерпретируемость, что делает их ценными инструментами для анализа и интерпретации больших данных в биологии.

Ключевые слова: онтологии; биоинформатика; системная биология; анализ больших данных; глубокое обучение; интерпретируемость.

Для цитирования: Подколодный Н.Л., Подколодная О.А., Иванисенко В.А., Марченко М.А. Онтологии в моделировании и анализе больших генетических данных. *Вавиловский журнал генетики и селекции*. 2024;28(8):940-949. doi 10.18699/vjgb-24-101

Финансирование. Работа выполнена при поддержке бюджетных проектов FWNR-2022-0020 и FWNM-2022-0005.

Ontologies in modelling and analysing of big genetic data

N.L. Podkolodnyy [1, 2, 3, 4], O.A. Podkolodnaya [1], V.A. Ivanisenko [1, 4], M.A. Marchenko [2, 3]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Institute of Computational Mathematics and Mathematical Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
[4] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

pnl@bionet.nsc.ru

Abstract. To systematize and effectively use the huge volume of experimental data accumulated in the field of bioinformatics and biomedicine, new approaches based on ontologies are needed, including automated methods for semantic integration of heterogeneous experimental data, methods for creating large knowledge bases and self-interpreting methods for analyzing large heterogeneous data based on deep learning. The article briefly presents the

© Подколодный Н.Л., Подколодная О.А., Иванисенко В.А., Марченко М.А., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

features of the subject area (bioinformatics, systems biology, biomedicine), formal definitions of the concept of ontology and knowledge graphs, as well as examples of using ontologies for semantic integration of heterogeneous data and creating large knowledge bases, as well as interpreting the results of deep learning on big data. As an example of a successful project, the Gene Ontology knowledge base is described, which not only includes terminological knowledge and gene ontology annotations (GOA), but also causal influence models (GO-CAM). This makes it useful not only for genomic biology, but also for systems biology, as well as for interpreting large-scale experimental data. An approach to building large ontologies using design patterns is discussed, using the ontology of biological attributes (OBA) as an example. Here, most of the classification is automatically computed based on previously created reference ontologies using automated inference, except for a small number of high-level concepts. One of the main problems of deep learning is the lack of interpretability, since neural networks often function as "black boxes" unable to explain their decisions. This paper describes approaches to creating methods for interpreting deep learning models and presents two examples of self-explanatory ontology-based deep learning models: (1) Deep GONet, which integrates Gene Ontology into a hierarchical neural network architecture, where each neuron represents a biological function. Experiments on cancer diagnostic datasets show that Deep GONet is easily interpretable and has high performance in distinguishing cancerous and non-cancerous samples. (2) ONN4MST, which uses biome ontologies to trace microbial sources of samples whose niches were previously poorly studied or unknown, detecting microbial contaminants. ONN4MST can distinguish samples from ontologically similar biomes, thus offering a quantitative way to characterize the evolution of the human gut microbial community. Both examples demonstrate high performance and interpretability, making them valuable tools for analyzing and interpreting big data in biology.

Key words: ontologies; big data analysis; bioinformatics; systems biology; deep learning; interpretability.

For citation: Podkolodnyy N.L., Podkolodnaya O.A., Ivanisenko V.A., Marchenko M.A. Ontologies in modelling and analysing of big genetic data. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8): 940-949. doi 10.18699/vjgb-24-101

Введение

Термин «большие данные» (Big Data) обозначает объемные наборы данных, которые отличаются значительными размерами, разнообразием и сложностью, что делает их трудными для обработки и анализа с помощью традиционных методов. Кроме того, такие данные часто имеют неполноту и неопределенность, что усложняет задачу контроля их качества и точности (Qaiser, Ghulam, 2023).

Появление качественно новых возможностей для проведения исследований, основанных на высокопроизводительных экспериментальных технологиях, таких как массовое параллельное секвенирование ДНК, многолокусное генотипирование, многопараметрическое профилирование экспрессии генов с использованием ДНК-чипов, технологии ChIP-on-chip, а также протеомные и метаболомные технологии, привело к накоплению беспрецедентно больших объемов экспериментальных данных и знаний (Stephens et al., 2015). Гетерогенность молекулярно-биологической информации и ее сложность затрудняют анализ, систематизацию и применение этих данных для решения конкретных задач в биоинформатике, биотехнологии, фармакологии и персонализированной медицине.

Для освоения, систематизации и эффективного использования огромного объема данных необходимы новые подходы к их обработке. В частности, это автоматизированные методы семантической интеграции гетерогенных данных, одним из ключевых этапов которой является согласование понятий предметной области, а также способов их описания и использования. Согласованное описание конкретной предметной области называется онтологией.

Онтологии позволяют представлять понятия в формате, пригодном для машинной обработки, и выступают в роли посредника между пользователем и информационной системой, а также между членами научного сообщества при обмене данными. Таким образом, онтологии становятся важным инструментом в биоинформатике и системной биологии, способствуя семантической интеграции экспериментальных данных и знаний с целью создания «единой картины мира». Кроме того, они помогают решать проблемы, возникающие при анализе больших данных, преодолевая разнородность и недостатки качества данных, а также улучшая интерпретацию результатов глубокого обучения. Онтологии повышают масштабируемость и эффективность обработки больших объемов информации, что делает их незаменимыми в современных научных исследованиях.

Ранее в обзоре (Podkolodnyy et al., 2016) были представлены примеры онтологий, описывающих биологические системы на различных уровнях организации живых систем. В настоящей статье будут рассмотрены примеры применения онтологий для интеграции гетерогенных данных и создания больших баз знаний, а также интерпретации результатов анализа данных.

Формальное представление онтологий

В информатике термин «онтология» обозначает концептуальную модель, которая описывает объекты, их свойства и взаимосвязи между ними (Chandrasekaran et al., 1999). Онтология включает в себя набор понятий (терминов) конкретной предметной области и их определения, а также всю информацию, связанную с этими понятиями, такую как свойства, отношения, ограничения, аксиомы и утверждения. Эта информация необходима для описания и решения задач в выбранной предметной области (Podkolodnyy et al., 2016).

Таким образом, формальная модель онтологии может быть представлена в виде упорядоченной тройки конечных множеств O = <T, R, F>, где T – это конечное и непустое множество классов и концептов (понятий, терминов) предметной области, которая рассматривается в определенном контексте (в нашем случае: биоинформатика, системная биология, биотехнология и биомедицина);

R – конечное множество отношений между концептами данной предметной области; F – конечное множество функций интерпретации, заданных на понятиях и/или отношениях онтологии O, а также аксиом, используемых для моделирования утверждений, которые всегда являются истинными. Это ограничивает интерпретацию и обеспечивает корректное использование понятий.

Одним из наиболее эффективных подходов к описанию и использованию знаний о предметной области являются дескриптивные логики, которые определяют формальный язык для описания понятий (концептов, классов, категорий или сущностей) и отношений между ними (называемых ролями), а также для формулирования утверждений о фактах и запросов к ним, включая проверку выполнимости и включения. Кроме того, дескриптивная логика включает конструкторы (операции) для создания понятийных выражений, такие как конъюнкция, дизъюнкция и определение отношений.

В базе знаний предметной области с точки зрения дескриптивной логики можно выделить две основные категории знаний. Первая категория включает общие знания о множестве классов понятий, их свойствах и отношениях между ними, что обозначается как терминологические знания (terminological knowledge), или T-Box. Вторая категория охватывает знания об индивидуальных объектах (экземплярах классов), их свойствах и связях с другими объектами, известные как утверждающие знания (assertional knowledge), или A-Box. Таким образом, T-Box описывает предметную область на уровне абстрактных понятий, в то время как A-Box является базой данных. Важно отметить, что обе компоненты базы знаний взаимосвязаны и дополняют друг друга.

Для систематического моделирования сложных систем, организмов и заболеваний, а также для представления знаний в области биоинформатики и системной биологии часто применяются графы знаний (ГЗ). В соответствии с определением, приведенным в работе (Callahan et al., 2024), граф знаний является структурой данных, которая отображает множество разнородных сущностей и различные типы отношений между ними. Эта структура служит абстрактной основой, способной генерировать новые знания, а также выявлять и разрешать расхождения или противоречия, что делает ее полезной для решения различных задач и применения в различных сценариях.

Можно выделить три типа графов знаний в зависимости от сложности представления и функциональности использования.

Простые графы – это наиболее распространенный и базовый тип графов. В таких графах сущности представлены в виде узлов, а ребра описывают отношения между ними. Обычно в простых графах отсутствует формальная семантика для ребер и узлов, что делает их простыми в применении, но ограничивает возможности для более глубокого анализа и интерпретации данных.

Гибридный граф, или граф свойств. Гибридные графы предназначены для моделирования сущностей и их отношений с использованием сочетания стандартных представлений сети и формальной семантики, таких как Resource Description Framework (RDF: https://www.w3.org/RDF) и RDF Schema (RDFS: https://www.w3.org/TR/rdf11-mt). В отличие от простых графов, гибридные графы, основанные на этих стандартах, облегчают интеграцию с другими ресурсами и обеспечивают большую возможность для автоматизированного вывода знаний. Это делает их более мощным инструментом для представления и обработки сложной информации.

Сложный граф, такой как в системе KaBOB (Livingston et al., 2015; Podkolodnyy et al., 2016), часто строится на основе Web Ontology Language (OWL). Сложные графы обладают высокой выразительностью, что позволяет эффективно генерировать новые знания с помощью дедуктивного вывода (Подколодный и др., 2012). Благодаря явной семантике, OWL предоставляет значительные преимущества в интеграции больших объемов биомедицинских данных по сравнению с RDF/RDFS, что делает его особенно полезным для сложных задач в области биоинформатики и системной биологии.

В качестве примера рассмотрим сеть основных взаимосвязанных биомедицинских концепций, необходимых для моделирования знаний о путях, генетических вариантах, заболеваниях и фармацевтических методах лечения (рис. 1). На верхнем уровне представлены анатомические сущности, такие как ткани, клетки и биологические жидкости (компартменты), содержащие геномные сущности (ДНК, РНК, мРНК и белки). ДНК кодирует гены, транскрибирующиеся в мРНК и транслирующиеся в белки, которые имеют молекулярные функции, могут взаимодействовать друг с другом и участвовать в путях и биологических процессах.

В последнее время было разработано несколько программных систем, таких как KG-HUB (Caufield et al., 2023), Clinical KG (CKG) (Santos et al., 2022), RTX-KG2 (Wood et al., 2022), BioCypher (Lobentanzer et al., 2023) и Knowledge Base Of Biomedicine (KaBOB) (Livingston et al., 2015; Podkolodnyy et al., 2016), которые обеспечивают широкие функциональные возможности создания и использования графов знаний в биоинформатике и биомедицине, включая интеграцию больших гетерогенных данных.

В работе (Callahan et al., 2024) описана семантическая экосистема PheKnowLator (Phenotype Knowledge Translator) для автоматизации построения онтологических ГЗ с полностью настраиваемым представлением знаний. Экосистема включает в себя различные компоненты для создания и использования ГЗ для решения различных прикладных задач, а также предварительно построенные графы знаний.

Интеграция больших данных и создание баз знаний на основе онтологий

В настоящее время в области биоинформатики, системной биологии, агробиологии, биомедицины разработано более тысячи онтологий, которые можно использовать для описания и интеграции знаний, анализа данных, а также вывода новых знаний (https://bioportal.bioontology.org/ontologies).

В качестве примера одного из самых успешных проектов создания онтологий можно привести проект Gene Ontology (GO) (http://www.geneontology.org/). В GO описаны

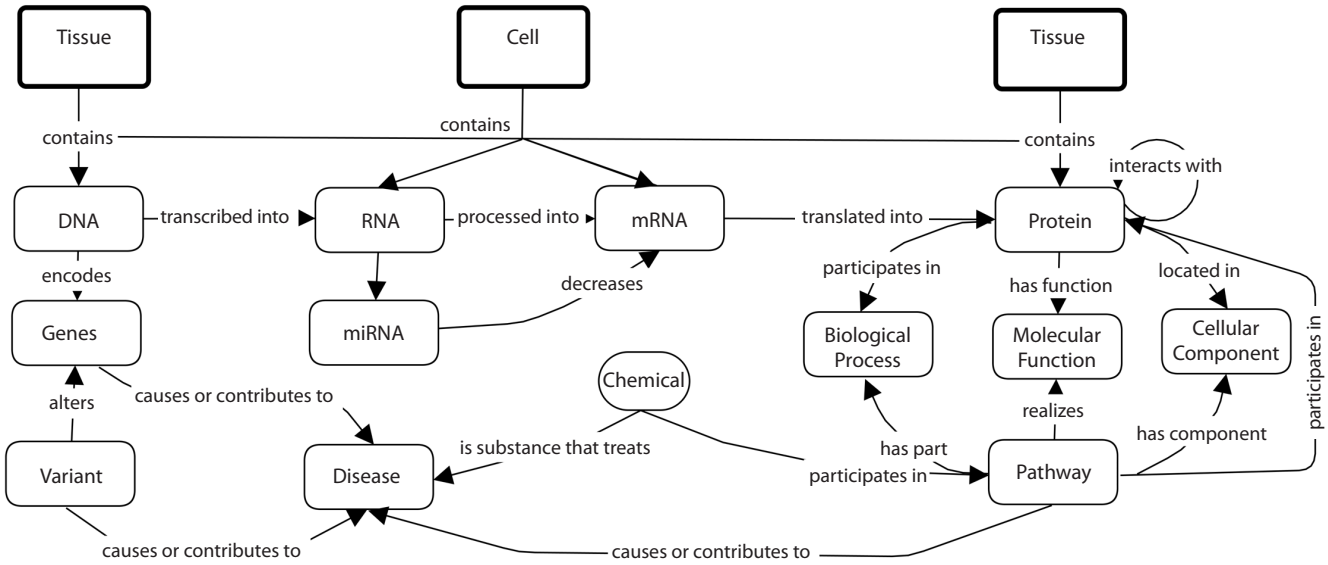


Рис. 1. Представление знаний об уровнях биологической организации, лежащих в основе описания заболеваний человека (Callahan et al., 2024).

текущие знания о типах функциональных характеристик (всего более 40 тыс. понятий), которыми может обладать генный продукт. В состав GO входят три раздела:

- Молекулярная функция (Molecular function) – элементарная молекулярная активность или роль, которую может выполнять ген, продукт гена в каких-либо биологических процессах. Всего описано 10365 терминов (https://geneontology.org/stats.html, дата обращения 08.09.2024).
- Биологические процессы (Biological process) – «биологическая программа», включающая набор молекулярных событий или активностей, действующих согласованно для достижения определенного результата и относящихся к функционированию интегрированных живых единиц: клеток, ткани, органов и организмов. В отличие от функции, процесс должен иметь несколько различающихся этапов с определенным началом и концом. Всего описано 26552 термина (https://geneontology.org/stats.html, дата обращения 08.09.2024).
- Клеточные компоненты (Cellular component) – часть анатомической структуры, в которой описывается локализация гена или его продукта в организме на уровнях клеточных структур и макромолекулярных комплексов или групп продуктов генов. Всего описано 4022 термина (https://geneontology.org/stats.html, дата обращения 08.09.2024).

Основные отношения между понятиями, которые используются в GO, включают простое отношение класс-подкласс (is_a), отношение часть-целое (part_of), отношения regulates, positively_regulates и negatively_regulates, которые описывают отношения между биологическими процессами, молекулярными функциями или биологическими свойствами. Свойство транзитивности отношений, используемых в GO, позволяет строить решетку отношений между понятиями и выполнять логический вывод о свойствах понятий и их отношений (Podkolodnyy et al., 2016).

На основе GO создана база знаний, которая кроме терминологических знаний (онтологии генов GO) включает результаты аннотации генов GOA (Gene Ontology Annotation, http://www.ebi.ac.uk/GOA), т. е. знания об индивидуальных объектах – генах и их продуктах (Huntley et al., 2015). В настоящее время GOA включает более 7.6 млн GO аннотаций для почти 1.54 млн белков и более 4.4 тыс. видов организмов.

Исходно на ранней стадии развития GO аннотация гена или его продукта (белка или РНК) осуществлялась независимо по молекулярным функциям, биологическим процессам или клеточным компонентам. Для того чтобы получить информацию о том, какую функцию выполняет ген или его продукт (РНК, белок) в том или ином биологическом процессе и той или иной клеточной структуре, понадобилась разработка еще одной компоненты базы знаний GO – модели причинных влияний между генными продуктами GO-CAM (Thomas et al., 2019).

GO-CAM связывает несколько аннотаций GO вместе, чтобы создать модели биологических процессов, соединяющие активности более чем одного генного продукта вместе в причинно-следственные сети и позволяющие специфицировать биологический контекст (например, тип клетки/ткани), в котором происходят активности. В качестве примера можно привести представление одной и той же биологической модели, описывающей, как убиквитин-протеинлигаза E3 NEDD4 подавляет транскрипцию РНК в ответ на повреждение ДНК, вызванное УФ-излучением, двумя способами: как набор разрозненных аннотаций GO, каждая из которых охватывает частичное описание общей функции (рис. 2, *а*), и как схема GO-CAM, связывающая аннотации GO в структурированную модель функций NEDD4, включая влияние активности NEDD4 на активность макромолекулярного комплекса РНК-полимеразы II (см. рис. 2, *б*).

Базовой единицей GO-CAM является единица активности продукта гена, которая объединяет аннотацию GO MF

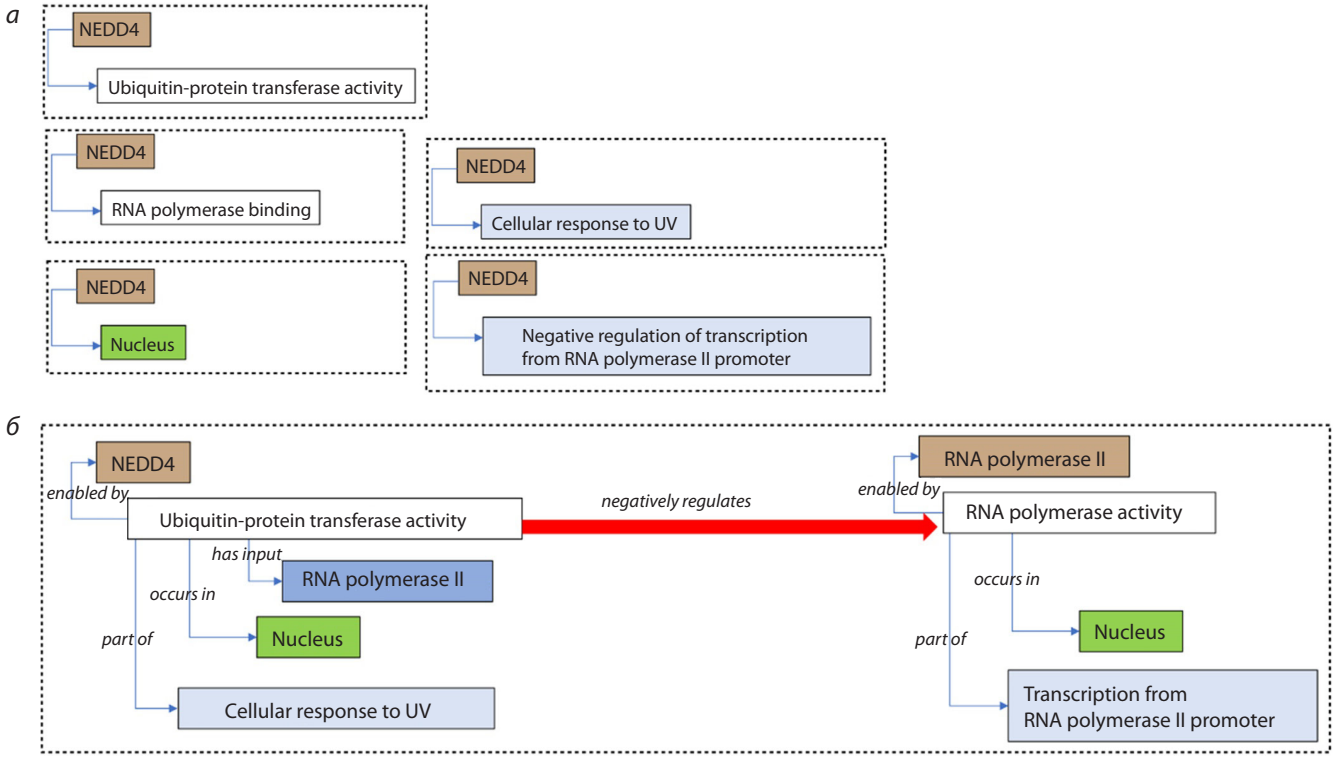


Рис. 2. Одна и та же биологическая модель, описывающая двумя способами, как NEDD4 подавляет транскрипцию РНК в ответ на повреждение ДНК, вызванное УФ-излучением: *а* – как набор разрозненных аннотаций GO, каждая из которых охватывает частичное описание общей функции; *б* – как схема GO-CAM, связывающая аннотации GO в структурированную модель функций NEDD4, включая влияние активности NEDD4 на активность макромолекулярного комплекса РНК-полимеразы II (Thomas et al., 2019).

(молекулярная активность) вместе с аннотациями GO CC (клеточные компоненты) и GO BP (биологический процесс), обеспечивающими биологический контекст активности. Контекст может быть дополнительно задан другими онтологиями, включая онтологии типа клетки (Cell Type Ontology) (Diehl et al., 2016), тканевого/анатомического расположения (с использованием нескольких различных онтологий в зависимости от вида организма, например, интегрированная онтология межвидовой анатомии, охватывающая животных и объединяющая несколько видоспецифических онтологий – Uberon (https://obophenotype.github.io/uberon/) (Mungall et al., 2012)), или неживотные онтологии, такие как Plant ontology (https://planteome.org/) (Cooper, Jaiswal, 2016), или описание временного периода (например, биологическая фаза GO). Единицы активности связаны между собой причинно-следственными связями из Онтологии Отношений (Smith et al., 2005).

Причинно-следственные сети в моделях GO-CAM позволяют использовать совершенно новые приложения, такие как сетевой анализ геномных данных и логическое моделирование биологических систем. Кроме того, модели могут оказаться полезными для визуализации пути. Например, основанное на активности представление GO-CAM совместимо с «диаграммами потока активности» стандарта Systems Biology Graphical Notation (SBGN) (Bergmann et al., 2020).

Таким образом, GO-CAM дает возможность использовать массивную базу знаний GO и GOA, собранную за последние двадцать лет, в качестве основы не только для геномной биологии представления функции гена, но и для более обширного представления системной биологии и его новых приложений к интерпретации крупномасштабных экспериментальных данных.

Пример GO анализа генов ассоциативной генной сети ревматоидного артрита

Ранее в Институте цитологии и генетики СО РАН была разработана программно-информационная система ANDSystem для автоматизированного извлечения медико-биологических знаний из научных публикаций и большого числа биологических и биомедицинских фактографических баз данных (Ivanisenko et al., 2015, 2019). База знаний ANDSystem представляет собой уникальный ресурс, содержащий формализованную информацию в виде ассоциативных генных сетей (графов знаний) о почти 44 млн взаимодействий различных типов между молекулярно-генетическими объектами.

Оригинальная онтология, лежащая в основе ANDSystem, дает очень подробное описание предметной области. В базе знаний ANDSystem описаны молекулярно-генетические объекты (белки, гены, метаболиты, микроРНК), биологические процессы, фенотипические признаки, лекарственные средства и их побочные эффекты, заболевания и т. д., а также более 25 типов взаимодействий между этими объектами, включая: физические взаимодействия с образованием молекулярных комплексов (белок/белок, белок/ДНК, метаболит/белок); каталитические реакции и протеолитические события с участием субстрата/фер-

мента/продукта; регуляторные взаимодействия; функции/активности, транспорта и стабильности белков; регуляцию трансляции белка с участием миРНК; регуляцию биологических процессов и фенотипических признаков с участием белков, метаболитов и лекарств; ассоциативные взаимодействия генов, белков, метаболитов, биологических процессов, фенотипических признаков с заболеваниями и т. д.

Примером типичной задачи использования ANDsystem является реконструкция ассоциативной генной сети (графа знаний) ревматоидного артрита (РА), содержащей 1025 генов/белков и более 20 тыс. взаимодействий между ними. Анализ перепредставленности терминов биологических процессов в Gene Ontology для множества генов ревматоидного артрита, выполненный с помощью системы DAVID (https://david.ncifcrf.gov/tools.jsp), выявил 376 биологических процессов, статистически значимо связанных с ревматоидным артритом (см. таблицу). Значения *p*-value вычислялись на основе гипергеометрического распределения. Для учета множественного тестирования использовалась поправка Бонферрони.

Рассмотрим более детально процесс иммунного ответа (GO:0006955~immune response). В Gene Ontology описано 420 генов, связанных с термином "GO:0006955~immune response". Из них 158 присутствуют в ассоциативной сети ревматоидного артрита (рис. 3). Вероятность выявления столь большого числа генов по случайным причинам очень низка (*p*-value с поправкой Бонферрони $<4.69 \cdot 10^{-79}$),

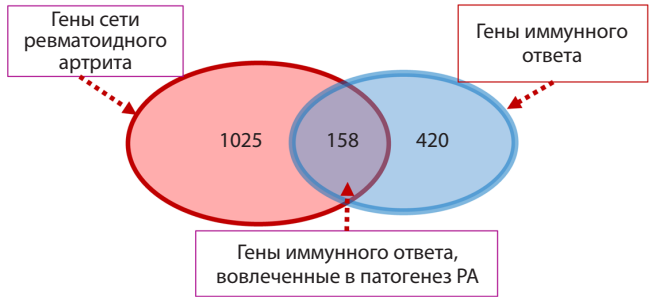


Рис. 3. Диаграмма Венна, описывающая пересечение генов сети ревматоидного артрита и генов иммунного ответа, связанных с термином "GO:0006955~immune response".

что указывает на высокую значимость взаимосвязи ревматоидного артрита с процессом иммунного ответа и на важнейшую роль иммунной системы в патогенезе этого заболевания.

Список первых из упорядоченных по статистической значимости (*p*-value с поправкой Бонферрони) биологических процессов, взаимосвязанных с ревматоидным артритом, представлен в таблице. Большая часть этих терминов так или иначе связана с процессами иммунного ответа и воспаления, играющими важную роль в патогенезе ревматоидного артрита. Эти процессы не являются независимыми.

Список первого 21 биологического процесса, статистически наиболее значимо связанного с ревматоидным артритом

Биологический процесс (Gene Ontology)	*p*-value с поправкой Бонферрони
GO:0006955~immune response	$4.69 \cdot 10^{-79}$
GO:0006954~ inflammatory response	$2.13 \cdot 10^{-70}$
GO:0060326~chemotaxis	$2.49 \cdot 10^{-30}$
GO:0007267~cell-cell signaling	$8.59 \cdot 10^{-28}$
GO:0032496~response to lipopolysaccharide	$7.41 \cdot 10^{-27}$
GO:0070098~chemokine-mediated signaling pathway	$3.91 \cdot 10^{-25}$
GO:1990256~signal transduction	$2.91 \cdot 10^{-24}$
GO:0071222~cellular response to lipopolysaccharide	$5.45 \cdot 10^{-24}$
GO:0050729~positive regulation of inflammatory response	$6.31 \cdot 10^{-24}$
GO:2001023~regulation of response to drug	$1.70 \cdot 10^{-23}$
GO:0070374~positive regulation of ERK1 and ERK2 cascade	$8.26 \cdot 10^{-23}$
GO:0001666~response to hypoxia	$9.11 \cdot 10^{-23}$
GO:0071864~positive regulation of cell proliferation	$2.52 \cdot 10^{-22}$
GO:0042102~positive regulation of T cell proliferation	$6.90 \cdot 10^{-22}$
GO:0045087~innate immune response	$2.09 \cdot 10^{-18}$
GO:0032729~positive regulation of interferon-gamma production	$2.38 \cdot 10^{-18}$
GO:0045766~positive regulation of angiogenesis	$2.90 \cdot 10^{-18}$
GO:0043066~negative regulation of apoptotic process	$5.72 \cdot 10^{-18}$
GO:0050731~positive regulation of peptidyl-tyrosine phosphorylation	$8.13 \cdot 10^{-18}$
GO:0007166~cell surface receptor signaling pathway	$8.40 \cdot 10^{-18}$
GO:0007568~aging	$1.28 \cdot 10^{-17}$

В частности, процесс иммунного ответа (термин “GO:0006955~immune response”) является обобщением таких терминов (см. таблицу), как “GO:0045087~innate immune response”, “GO:0032729~positive regulation of interferon-gamma production”, “GO:0060326~chemotaxis”, “GO:0042102~positive regulation of T cell proliferation”, “GO:1990256~signal transduction” и др.

Процесс воспалительного ответа (GO:0006954~inflammatory response) включает более частные процессы, описываемые терминами “GO:0032496~response to lipopolysaccharide”, “GO:0050729~positive regulation of inflammatory response”, “GO:1990256~signal transduction”, “GO:0001666~response to hypoxia”, “GO:0045766~positive regulation of angiogenesis”. И даже термин “GO:0007568~aging” связан с термином “GO:0006954~inflammatory response”, так как одним из механизмов старения являются хронические неинфекционные воспаления.

Приведенные на примере ревматоидного артрита результаты свидетельствуют, что подход к поиску генов, ассоциированных с конкретным заболеванием, с помощью ANDsystem и дальнейший GO анализ этой группы генов позволяет выявлять ключевые биологические процессы, участвующие в патогенезе данного заболевания.

Использование шаблонов проектирования онтологий для интеграции онтологий фенотипов и биологических атрибутов

Онтологии, имеющие логически богатую аксиоматизацию, обеспечивают такие мощные возможности, как автоматизированные рассуждения, классификацию и логические запросы. Однако ручное создание такого рода онтологий крайне затратно и требует от аннотаторов не только быть специалистами в предметной области, но и обладать знаниями в области логического моделирования (Slater et al., 2020).

Популярным подходом к решению данной проблемы является применение шаблонов проектирования и систем шаблонов для логических аксиом (Osumi-Sutherland et al., 2017). Это позволяет отделить курирование ссылочных терминов и их логических определений от их точной аксиоматической картины. Центральная идея состоит в том, чтобы задействовать небольшое количество шаблонов аксиом, реализующих шаблоны проектирования, которые могут быть созданы и поддерживаться экспертами по логике, и чтобы кураторы контента сосредоточились на выборе соответствующих терминов из различных ссылочных онтологий (например, терминов из онтологии Uberon для определения анатомических атрибутов).

Онтология биологических атрибутов (OBA) – это стандартизированная структура для наблюдаемых атрибутов, которые являются характеристиками организмов или частей организмов (Stefancsik et al., 2023). В отличие от большинства фенотипических онтологий, в OBA логические аксиомы определяют общие атрибуты без ссылки на какие-либо конкретные фенотипические изменения или состояния.

При создании OBA использовался шаблон проектирования типа Сущность-Качество (EQ), в котором фенотипическое качество (Q), такое как “height”, “mass” или “amount” из онтологии фенотипа и признаков (PATO) (Gkoutos et al., 2018), объединяется с сущностью (E), такой как анатомическая или химическая сущность, чтобы сформировать концепцию «биологического атрибута», называемого “trait”. Например, понятие “blood glucose amount” (OBA:VT0000188) включает класс “amount” (PATO:000070), который определяет характеристику глюкозы – “glucose” (CHEBI:17234) в крови – “blood” (UBERON:0000178).

В настоящее время в OBA применяется десять шаблонов признаков из Dead Simple Ontology Design Patterns (DOS-DP) (Osumi-Sutherland et al., 2017). Они были выбраны, потому что охватывают большинство анатомических, химических и клеточных атрибутов, являющихся центральными для интеграции данных геномики.

Богатая логическая аксиоматизация, основанная на шаблонах проектирования, необходима для обеспечения совместимости с существующими онтологиями фенотипов и другими типами данных, такими как анатомические, химические и биологические данные о метаболических путях и генных сетях.

Большинство атрибутов в OBA логически можно определить с помощью OWL, используя термины из соответствующих ссылочных онтологий, например Uberon (Mungall et al., 2012) или Chebi (Hastings et al., 2016). За исключением незначительного числа высокоуровневых понятий, большая часть классификации в OBA автоматически вычисляется на основе классификаций различных эталонных онтологий, с помощью автоматизированного логического вывода. Можно выделить два преимущества этого подхода: во-первых, не нужно вручную классифицировать какие-либо понятия, что значительно снижает стоимость кураторства классификации при одновременном повышении ее полноты. Во-вторых, многочисленные ссылки на ссылочные онтологии могут использоваться для самых разных целей, включая запрос (например, выбор всех данных, где затрагивается морфология части почечной системы), интеграцию графов знаний (например, автоматическая привязка к фенотипическим аномалиям из широко используемых онтологий, таких как онтология фенотипа человека (HPO) или фенотипа млекопитающих (MP)) и вывод знаний (например, логический вывод недостающих данных) (Dececchi et al., 2015).

Применение онтологий для интерпретации глубокого обучения

Глубокое обучение (DL) наглядно продемонстрировало свою эффективность при решении задач в области геномики, протеомики, биомедицины, включая анализ и автоматическую функциональную аннотацию последовательностей ДНК, РНК и белков, поиск ДНК/РНК-мишеней регуляторных РНК и белков, прогнозирование свойств и функций биомолекул, поиск 3D структуры белка, реконструкцию структур биомолекул с заданными свойствами, прогнозирование взаимодействия биомолекул и выявление на этой основе потенциальных кандидатов на лекарственные препараты, обработку и анализ изображений, интеграцию омиксных данных, анализ сложных, гетерогенных и взаимосвязанных биологических сетей (в том числе сетей белок-белковых взаимодействий, генных ре-

гуляторных сетей, семантических сетей и метаболических путей), моделирование биологических систем и процессов и т. д. (Li et al., 2019; Sapoval et al., 2022).

Одной из ключевых проблем глубокого обучения в биоинформатике, системной биологии и современной биомедицине является недостаточная интерпретируемость моделей нейронных сетей, которые часто функционируют как модели «черного ящика».

Интерпретируемость алгоритмов машинного обучения в биоинформатике и биомедицине важна по трем основным причинам. Во-первых, при анализе сложных систем, когда нет теории и четкого алгоритма принятия решений, требуется понять, почему модель делает такое предсказание. Во-вторых, важно гарантировать, что модель основывает свои прогнозы на надежном представлении данных и не фокусируется на нерелевантных артефактах. И наконец, модель с высокоточными предсказаниями, возможно, выявила интересные закономерности, которые биологи хотели бы изучить.

В формально-логическом смысле интерпретация – это отображение формальной конструкции на сущности и их отношения, которые она представляет. В этом смысле можно сказать, что кто-то понимает формальную конструкцию, если может соотнести ее с соответствующими сущностями и предложениями реального мира и рассуждать о последствиях. При этом важно отличать понятность модели от понимания того, почему модель истинна или как модель была получена из данных.

Можно выделить два основных подхода к интерпретации черных ящиков: апостериорные методы и самообъясняемые модели (Adadi, Berrada, 2018). В апостериорном методе сначала изучается модель «черного ящика», а затем используется метод интерпретации для объяснения того, как делались прогнозы. Часто полученные таким образом объяснения не совпадают с тем, как алгоритм глубокого обучения на самом деле получает решение. Кроме того, процедура объяснения является отдельным методом со своими ошибками, которые влияют на качество принимаемых решений. Поэтому такое объяснение не всегда подходит для биомедицины.

Необходимо отметить, что интерпретируемость – это понятие, специфичное для конкретной области, поэтому не может быть универсального определения. Очень часто в интерпретируемой модели машинного обучения добавляются ограничения в модельной форме, так чтобы она либо была полезна кому-то, либо подчинялась структурным знаниям области, таким как монотонность (Gupta et al., 2016), причинность, структурные (генеративные) ограничения, аддитивность (Lou et al., 2013) или физические ограничения, которые исходят из знания предметной области (онтологий).

К настоящему времени уже опубликовано несколько работ по построению самообъясняемых нейронных сетей на основе данных об экспрессии генов с использованием знания Gene Ontology. Так, например, в работе (Bourgeais et al., 2021) предложена самообъясняемая модель глубокого обучения Deep GONet, интегрирующая Gene Ontology в иерархическую архитектуру нейронной сети. Эта модель основана на полностью связанной архитектуре, ограниченной аннотациями GO, так что каждый нейрон представляет биологическую функцию. Эксперименты с наборами данных диагностики рака показывают, что Deep GONet легко интерпретируется и обладает высокой производительностью для различения раковых и нераковых образцов.

Другим примером самообъясняемой нейронной сети является ONN4MST – обобщение вычислительной модели Neural Network, основанной на онтологии (ONN), для отслеживания микробных источников (Zha, Ning, 2022). Модель ONN использует новый онтологический подход, который поощряет предсказание, удовлетворяющее онтологии «биомов». Другими словами, модель ONN может использовать информацию онтологии биома для моделирования зависимостей между биомами и оценки доли различных биомов в выборке сообщества.

Способность ONN4MST к открытию новых знаний продемонстрирована в различных приложениях отслеживания источников. Она позволяет отслеживать источники образцов, ниши которых ранее были менее изучены или неизвестны, обнаруживать микробные загрязнители, а также идентифицировать аналогичные образцы из онтологически удаленных биомов, показывающих уникальную важность ONN4MST в открытии знаний из огромного количества микробных образцов сообщества гетерогенных биомов.

ONN4MST может отличать образцы от онтологически близких биомов и, таким образом, предлагает количественный способ охарактеризовать развитие микробного сообщества кишечника человека. В частности, показано, что микробиом кишечника долгожителей отличается от микробиома нормальных пожилых людей и более похож на микробиом молодых людей (Zha, Ning, 2022).

Заключение

Бурное развитие экспериментальных технологий в области молекулярной биологии привело к тому, что онтологическое моделирование становится базовым методом в биоинформатике и системной биологии. Создание нескольких сотен базовых ссылочных онтологий и их верификация позволяют применять эти онтологии в качестве источников знаний для интеграции и построения сложных моделей предметной области и баз знаний, ориентированных на решение конкретных задач биомедицины.

Особое значение онтологии имеют для интерпретации результатов компьютерных предсказаний, полученных с помощью методов глубокого обучения. Чтобы ученые доверяли глубокому обучению, которое часто представлено как модели «черного ящика», должны использоваться специальные методы интерпретации на основе дополнительных знаний о предметной области или онтологий. Онтологии, паттерны их конструирования, интеграция больших данных и создание графов знаний играют ключевую роль в повышении интерпретируемости моделей глубокого обучения. Эти инструменты позволяют не только улучшить понимание результатов, но и обеспечить более качественный анализ данных. В условиях стремительного роста объемов информации и сложности моделей глубокого обучения использование онтологий становится необходимым шагом к созданию более прозрачных и объяснимых систем.

Можно ожидать, что новое поколение систем интерпретации будет не только в состоянии объяснить полученные решения понятным для человека способом, с указанием количественного уровня неопределенности, но и предложит дополнительные шаги (например, дополнительные эксперименты, клинические исследования и т. д.), необходимые для уточнения или надежного подтверждения своих решений.

Список литературы / References

Подколодный Н.Л., Игнатьева Е.В., Подколодная О.А., Колчанов Н.А. Информационная поддержка исследования механизмов регуляции транскрипции: онтологический подход. *Вавиловский журнал генетики и селекции*. 2012;16(4/1):742-755 [Podkolodnyy N.L., Ignatyeva E.V., Podkolodnaya O.A., Kolchanov N.A. Information support of research on transcriptional regulatory mechanisms: an ontological approach. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2012;16(4/1):742-755 (in Russian)]

Adadi A., Berrada M. Peeking inside the Black-Box: a survey on explainable artificial intelligence (XAI). *IEEE Access*. 2018;6:52138-52160. doi 10.1109/ACCESS.2018.2870052

Bergmann F.T., Czauderna T., Dogrusoz U., Rougny A., Drager A., Toure V., Mazein A., Blinov M.L., Luna A. Systems biology graphical notation markup language (SBGNML) version 0.3. *J. Integr. Bioinform*. 2020;17(2-3):20200016. doi 10.1515/jib-2020-0016

Bourgeais V., Zehraoui F., Ben Hamdoune M., Hanczar B. Deep GONet: self-explainable deep neural network based on Gene Ontology for phenotype prediction from gene expression data. *BMC Bioinformatics*. 2021;22(S10):455. doi 10.1186/s12859-021-04370-7

Callahan T.J., Tripodi I.J., Stefanski A.L., Cappelletti L., Taneja S.B., Wyrwa J.M., Casiraghi E., Matentzoglu N.A., Reese J., Silverstein J.C., Hoyt C.T., Boyce R.D., Malec S.A., Unni D.R., Joachimiak M.P., Robinson P.N., Mungall C.J., Cavalleri E., Fontana T., Valentini G., Mesiti M., Gillenwater L.A., Santangelo B., Vasilevsky N.A., Hoehndorf R., Bennett T.D., Ryan P.B., Hripcsak G., Kahn M.G., Bada M., Baumgartner W.A., Hunter L.E. An open source knowledge graph ecosystem for the life sciences. *Sci. Data*. 2024;11(1):363. doi 10.1038/s41597-024-03171-w

Caufield J.H., Putman T., Schaper K., Unni D.R., Hegde H., Callahan T.J., Cappelletti L., Moxon S.A.T., Ravanmehr V., Carbon S., Chan L.E., Cortes K., Shefchek K.A., Elsarboukh G., Balhoff J., Fontana T., Matentzoglu N., Bruskiewich R.M., Thessen A.E., Harris N.L., Munoz-Torres M.C., Haendel M.A., Robinson P.N., Joachimiak M.P., Mungall C.J., Reese J.T. KG-Hub – building and exchanging biological knowledge graphs. *Bioinformatics*. 2023;39(7):btad418. doi 10.1093/bioinformatics/btad418

Chandrasekaran B., Josephson J., Benjamins V. What are ontologies, and why do we need them? *IEEE Intell. Syst. Appl*. 1999;14(1):20-26. doi 10.1109/5254.747902

Cooper L., Jaiswal P. The plant ontology: a tool for plant genomics. In Edwards D. (Ed.) Plant Bioinformatics. Methods in Molecular Biology. Vol. 1374. New York: Humana Press, 2016;89-114. doi 10.1007/978-1-4939-3167-5_5

Dececchi T.A., Balhoff J.P., Lapp H., Mabee P.M. Toward synthesizing our knowledge of morphology: using ontologies and machine reasoning to extract presence/absence evolutionary phenotypes across studies. *Syst. Biol*. 2015;64(6):936-952. doi 10.1093/sysbio/syv031

Diehl A.D., Meehan T.F., Bradford Y.M., Brush M.H., Dahdul W.M., Dougall D.S., He Y., Osumi-Sutherland D., Ruttenberg A., Sarntivijai S., Van Slyke C.E., Vasilevsky N.A., Haendel M.A., Blake J.A., Mungall C.J. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semantics*. 2016;7(1):44. doi 10.1186/s13326-016-0088-7

Gkoutos G.V., Schofield P.N., Hoehndorf R. The anatomy of phenotype ontologies: principles, properties and applications. *Brief Bioinform*. 2018;19(5):1008-1021. doi 10.1093/bib/bbx035

Gupta M., Cotter A., Pfeifer J., Voevodski K., Canini K., Mangylov A., Moczydlowski W., van Esbroeck A. Monotonic calibrated interpolated look-up tables. *J. Mach. Learn. Res*. 2016;17:1-47

Hastings J., Owen G., Dekker A., Ennis M., Kale N., Muthukrishnan V., Turner S., Swainston N., Mendes P., Steinbeck C. ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res*. 2016;44(D1):D1214-D1219. doi 10.1093/nar/gkv1031

Huntley R.P., Sawford T., Mutowo-Meullenet P., Shypitsyna A., Bonilla C., Martin M.J., O'Donovan C. The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res*. 2015;43(D1):D1057-D1063. doi 10.1093/nar/gku1113

Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol*. 2015;9(Suppl.2):S2. doi 10.1186/1752-0509-9-S2-S2

Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl.1):34. doi 10.1186/s12859-018-2567-6

Li Y., Huang C., Ding L., Li Z., Pan Y., Gao X. Deep learning in bioinformatics: introduction, application, and perspective in big data era. *Methods*. 2019;166:4-21. doi 10.1016/j.ymeth.2019.04.008

Livingston K.M., Bada M., Baumgartner W.A., Hunter L.E. KaBOB: ontology-based semantic integration of biomedical databases. *BMC Bioinformatics*. 2015;16(1):126. doi 10.1186/s12859-015-0559-3

Lobentanzer S., Aloy P., Baumbach J., Bohar B., Carey V.J., Charoentong P., Danhauser K., Doğan T., Dreo J., Dunham I., Farr E., Fernandez-Torras A., Gyori B.M., Hartung M., Hoyt C.T., Klein C., Korcsmaros T., Maier A., Mann M., Ochoa D., Pareja-Lorente E., Popp F., Preusse M., Probul N., Schwikowski B., Sen B., Strauss M.T., Turei D., Ulusoy E., Waltemath D., Wodke J.A.H., Saez-Rodriguez J. Democratizing knowledge representation with BioCypher. *Nat. Biotechnol*. 2023;41(8):1056-1059. doi 10.1038/s41587-023-01848-y

Lou Y., Caruana R., Gehrke J., Hooker G. Accurate intelligible models with pairwise interactions. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: Assoc. for Computing Machinery, 2013;623-631. doi 10.1145/2487575.2487579

Mungall C.J., Torniai C., Gkoutos G.V., Lewis S.E., Haendel M.A. Uberon, an integrative multi-species anatomy ontology. *Genome Biol*. 2012;13(1):R5. doi 10.1186/gb-2012-13-1-r5

Osumi-Sutherland D., Courtot M., Balhoff J., Mungall C. Dead simple OWL design patterns. *J. Biomed. Semant*. 2017;8:18. doi 10.1186/s13326-017-0126-0

Podkolodnyy N.L., Podkolodnaya O.A. Ontologies in bioinformatics and systems biology. *Russ. J. Genet. Appl. Res.* 2016;6(7):749-758. doi 10.1134/S2079059716070091

Qaiser A., Ghulam S. Bioinformatics and big data analytics in genomic research. *Med. Pap.* 2023;3(1):165-179. doi 10.31219/osf.io/5grpc

Santos A., Colaço A.R., Nielsen A.B., Niu L., Strauss M., Geyer P.E., Coscia F., Albrechtsen N.J.W., Mundt F., Jensen L.J., Mann M. A knowledge graph to interpret clinical proteomics data. *Nat. Biotechnol*. 2022;40(5):692-702. doi 10.1038/s41587-021-01145-6

Sapoval N., Aghazadeh A., Nute M.G., Antunes D.A., Balaji A., Baraniuk R., Barberan C.J., Dannenfelser R., Dun C., Edrisi M., Elworth R.A.L., Kille B., Kyrillidis A., Nakhleh L., Wolfe C.R., Yan Z., Yao V., Treangen T.J. Current progress and open challenges for applying deep learning across the biosciences. *Nat. Commun*. 2022;13(1):1728. doi 10.1038/s41467-022-29268-7

Slater L.T., Gkoutos G.V., Hoehndorf R. Towards semantic interoperability: finding and repairing hidden contradictions in biomedical ontologies. *BMC Med. Inform. Decis. Mak*. 2020;20(Suppl.10):311. doi 10.1186/s12911-020-01336-2

Smith B., Ceusters W., Klagges B., Kohler J., Kumar A., Lomax J., Mungall C., Neuhaus F., Rector A.L., Rosse C. Relations in biomedical ontologies. *Genome Biol*. 2005;6(5):R46. doi 10.1186/gb-2005-6-5-r46

Stefancsik R., Balhoff J.P., Balk M.A., Ball R.L., Bello S.M., Caron A.R., Chesler E.J., de Souza V., Gehrke S., Haendel M., Harris L.W., Harris N.L., Ibrahim A., Koehler S., Matentzoglu N., McMurry J.A., Mungall C.J., Munoz-Torres M.C., Putman T., Robinson P., Smedley D., Sollis E., Thessen A.E., Vasilevsky N., Walton D.O., Osumi-Sutherland D. The Ontology of Biological Attributes (OBA)-computational traits for the life sciences. *Mamm. Genome*. 2023;34(3):364-378. doi 10.1007/s00335-023-09992-1

Stephens Z.D., Lee S.Y., Faghri F., Campbell R.H., Zhai C., Efron M.J., Iyer R., Schatz M.C., Sinha S., Robinson G.E. Big Data: astronomical or genomical? *PLoS Biol*. 2015;13(7):e1002195. doi 10.1371/journal.pbio.1002195

Thomas P.D., Hill D.P., Mi H., Osumi-Sutherland D., Van Auken K., Carbon S., Balhoff J.P., Albou L.-P., Good B., Gaudet P., Lewis S.E., Mungall C.J. Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. *Nat. Genet*. 2019;51(10):1429-1433. doi 10.1038/s41588-019-0500-1

Wood E.C., Glen A.K., Kvarfordt L.G., Womack F., Acevedo L., Yoon T.S., Ma C., Flores V., Sinha M., Chodpathumwan Y., Termehchy A., Roach J.C., Mendoza L., Hoffman A.S., Deutsch E.W., Koslicki D., Ramsey S.A. RTX-KG2: a system for building a semantically standardized knowledge graph for translational biomedicine. *BMC Bioinformatics*. 2022;23(1):400. doi 10.1186/s12859-022-04932-3

Zha Y., Ning K. Ontology-aware neural network: a general framework for pattern mining from microbiome data. *Brief. Bioinform*. 2022;23(2):bbac005. doi 10.1093/bib/bbac005

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 28.10.2024. После доработки 08.11.2024. Принята к публикации 11.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):950-959

doi 10.18699/vjgb-24-102

PlantReg: реконструкция связей между регуляторными сетями транскрипционных факторов и контролируемыми ими признаками

В.В. Лавреха [1, 2#], Н.А. Омельянчук [1#], А.Г. Богомолов [1], Е.В. Землянская [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

ezemlyanskaya@bionet.nsc.ru

Аннотация. Описание пути от гена к признаку как основная задача многих отраслей биологии в настоящее время оснащается новыми методами не только в технике экспериментов, но и в системном анализе их результатов. Плейотропный эффект гена осуществляется за счет его участия во многих биологических процессах, вовлеченных в разные признаки. Широкое распространение полногеномного секвенирования транскриптов и районов связывания транскрипционных факторов (ТФ) сделало актуальными задачи установления плейотропных эффектов ТФ за счет знаний о функциях их мишеней, составление списков ТФ, регулирующих интересующие исследователя биологические процессы, описание путей их действия (первичные и вторичные мишени, мишени следующих порядков, взаимодействие между ТФ в исследуемом процессе). Ранее мы разработали метод реконструкции регуляторных сетей ТФ и предложили подход, позволяющий выявлять, какие биологические процессы и каким образом эти сети регулируют. В данной работе мы реализовали этот подход в виде программы PlantReg, доступной пользователям через веб-интерфейс. В статье описан принцип работы программы. На вход подаются список генов и список ТФ – известных или предполагаемых регуляторов транскрипции этих генов. На выходе программа выдает список биологических процессов, которые обогащены в этих генах, а также информацию о том, какими ТФ и через какие гены эти процессы регулируются. Работа PlantReg проиллюстрирована на примере исследования регуляции процессов, инициируемых на начальных этапах ответа на солевой стресс у *Arabidopsis thaliana* L. С помощью программы PlantReg нами выявлены биологические процессы, стимулируемые в раннем ответе на солевой стресс, и специфический набор ТФ, активирующих каждый из этих процессов. На примере одного из таких процессов – ответа на фитогормон абсцизовую кислоту – мы показали, что солевой стресс активирует в основном сигнальный путь этого гормона, и выделили ключевые ТФ в этой регуляции. Таким образом, программа PlantReg – удобный инструмент для создания гипотез о молекулярных механизмах регуляции признаков растений.

Ключевые слова: генная онтология; биологические процессы; регуляторные генные сети; *Arabidopsis thaliana*.

Для цитирования: Лавреха В.В., Омельянчук Н.А., Богомолов А.Г., Землянская Е.В. PlantReg: реконструкция связей между регуляторными сетями транскрипционных факторов и контролируемыми ими признаками. *Вавиловский журнал генетики и селекции*. 2024;28(8):950-959. doi 10.18699/vjgb-24-102

Финансирование. Работа выполнена в рамках бюджетного проекта FWNR-2022-0020.

PlantReg: the reconstruction of links between transcription factor regulatory networks and biological processes under their control

V.V. Lavrekha [1, 2#], N.A. Omelyanchuk [1#], A.G. Bogomolov [1], E.V. Zemlyanskaya [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia

ezemlyanskaya@bionet.nsc.ru

Abstract. The description of the path from a gene to a trait, as the main task of many areas in biology, is currently being equipped with new methods affecting not only experimental techniques, but also analysis of the results. The pleiotropic effect of a gene is due to its participation in numerous biological processes involved in different traits. A widespread use of genome-wide sequencing of transcripts and transcription factor (TF) binding regions has made the following tasks relevant: unveiling pleiotropic effects of TFs based on the functions of their target genes; compiling the lists of TFs that regulate biological processes of interest; and describing the ways of TF functioning (their primary and secondary targets, higher order targets, TF interactions in the process under study). We have previously developed a method for the reconstruction of TF regulatory networks and proposed an approach that allows identifying which biological processes are controlled by these networks and how this control is exerted. In this paper, we have

© Лавреха В.В., Омельянчук Н.А., Богомолов А.Г., Землянская Е.В., 2024

[#] Равный вклад авторов.

Контент доступен под лицензией Creative Commons Attribution 4.0

implemented the approach as PlantReg, a program available as a web service. The paper describes how the program works. The input consists of a list of genes and a list of TFs – known or putative transcriptional regulators of these genes. As an output, the program provides a list of biological processes enriched for these genes, as well as information about by which TFs and through which genes these processes are controlled. We illustrated the use of PlantReg deciphering transcriptional regulation of processes initiated at the early salt stress response in *Arabidopsis thaliana* L. With PlantReg, we identified biological processes stimulated by the stress, and specific sets of TFs that activate each process. With one of these processes (response to abscisic acid) as an example, we showed that salt stress mainly affects abscisic acid signaling and identified key TFs in this regulation. Thus, PlantReg is a convenient tool for generating hypotheses about the molecular mechanisms that control plant traits.

Key words: gene ontology; biological processes; gene regulatory networks; *Arabidopsis thaliana*.

For citation: Lavrekha V.V., Omelyanchuk N.A., Bogomolov A.G., Zemlyanskaya E.V. PlantReg: the reconstruction of links between transcription factor regulatory networks and biological processes under their control. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):950-959. doi 10.18699/vjgb-24-102

Введение

Эффективное развитие методов секвенирования транскриптомов открыло большие возможности не только для изучения изменения экспрессии генов на уровне транскрипции, но и для отслеживания регуляции этих изменений транскрипционными факторами (ТФ) и влияния этих изменений на биологические процессы (Chen J.W. et al., 2023). В связи с этим в настоящее время широко применяются методы извлечения списков ТФ на основании наличия их сайтов связывания в промоторах дифференциально экспрессирующихся генов (ДЭГ) и методы выявления обогащения определенными терминами генной онтологии (ГО) списков генов, т. е. их функциональная аннотация. Тем не менее нахождение связи между результатами применения этих методов (т. е. установление, какие ТФ изменяют конкретные биологические процессы, на каких их стадиях, есть ли ТФ, общие для всех процессов или специфичные для каких-либо из них) остается слабо проработанным звеном в анализе результатов транскриптомных экспериментов. Создание компьютерных программ для этого позволит сделать транскриптомные исследования более системными и выстроить связь от изменений в экспрессии генов до изменений в течении биологических процессов.

Транскрипционные факторы могут регулировать друг друга на уровне транскрипции, такую регуляторную конструкцию часто представляют в виде графов – регуляторных сетей транскрипционных факторов (РСТФ), которые могут быть реконструированы с помощью различных методов (Hecker et al., 2023). Реконструированные РСТФ позволяют установить иерархию в их структуре и определить хабы – ТФ, связанные с наибольшим числом других ТФ. Изменение в результате экспериментальных манипуляций экспрессией генов, кодирующих хабы, наиболее вероятно изменит функции всей РСТФ, а значит, и биологические процессы, регулируемые ею (He, Zhang, 2006).

Ранее мы разработали методику и программное средство для реконструкции РСТФ (Omelyanchuk et al., 2024). Мы также предложили биоинформатический подход, позволяющий выявлять, какие биологические процессы эти сети регулируют и каким образом. Он заключается в следующем: на основании списка ДЭГ автоматизированным способом находится список ТФ, обогащенных в промоторах этих ДЭГ, и реконструируется РСТФ. После функциональной аннотации списка ДЭГ для каждого ДЭГ в биологическом процессе ищется потенциальный регулятор среди обогащенных ТФ. Таким образом устанавливается, какие стадии биологического процесса какими факторами регулируются. Использование этого подхода проиллюстрировано в статье (Omelyanchuk et al., 2024) на примерах регуляции ауксином биосинтеза хлорофилла и лигнина, пути передачи сигнала абсцизовой кислоты и биогенеза рибосом.

В настоящей работе мы реализовали этот подход в виде компьютерной программы PlantReg, доступной пользователям через веб-интерфейс (https://plamorph.sysbio.ru/fannotf/). Мы использовали PlantReg для исследования регуляции процессов на начальных этапах ответа на солевой стресс у *Arabidopsis thaliana* L. с помощью транскриптомных данных из статьи (Wu et al., 2021a). С использованием программы PlantReg нами установлены процессы, координируемые всеми ТФ в составе РСТФ раннего ответа на солевой стресс, – это ответы на тепло, красный и дальний красный свет и салициловую кислоту. Наибольшее число процессов – запрограммированная гибель клеток, старение листьев и ответы на синий свет, гипоксию, активные формы кислорода, обезвоживание, абсцизовую и жасмоновую кислоты – регулируется не менее 70 % ТФ, входящих в состав РСТФ. В контроле ответа эндоплазматического ретикулума на потерю третичной структуры белков, биосинтеза индол-содержащих соединений и S-гликозидов, а также транспорта воды участвует менее 50 % регуляторных сетей транскрипционных факторов.

Далее мы более подробно рассмотрели результаты PlantReg по регуляции ответа на абсцизовую кислоту (АБК) при солевом стрессе и обнаружили, что эта регуляция осуществляется преимущественно за счет контроля сигнального пути АБК, и более всего модулируется последний этап этого пути – активация главных ТФ ответа на АБК. Оба хаба РСТФ ответа на солевой стресс, WRKY8 и DEAR2, задействованы на этом этапе, DEAR2 также контролирует работу рецепторов АБК. Таким образом, программа PlantReg – эффективный инструмент для анализа данных по дифференциальной экспрессии генов в транскриптомах и создания гипотез о молекулярных механизмах регуляции биологических процессов.

Материалы и методы

Программная реализация PlantReg. Схема работы программы PlantReg представлена на рис. 1. На вход программы подаются список генов (в данной работе мы

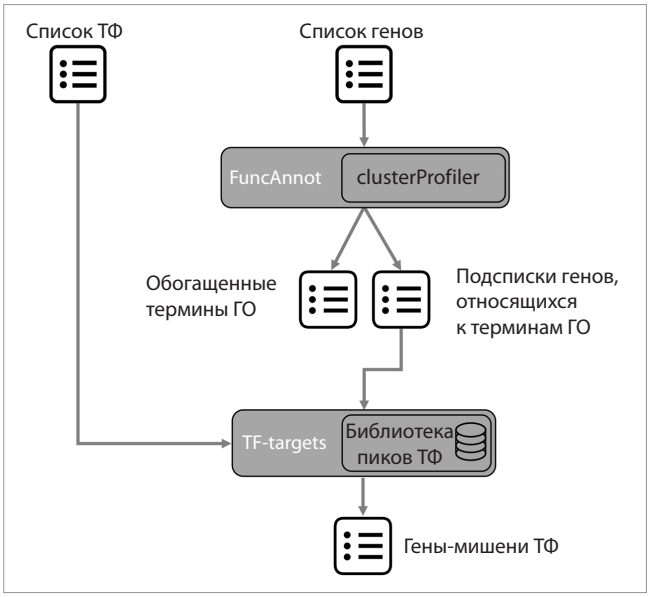


Рис. 1. Схема работы программы PlantReg.

фокусируемся на анализе списков ДЭГ), а также список ТФ – известных или предполагаемых регуляторов транскрипции этих генов. На этапе FuncAnnot осуществляется функциональная аннотация списка генов с использованием R-пакета clusterProfiler (Yu et al., 2012; Wu et al., 2021b). Результатом является файл, содержащий информацию об обогащенных в списке ДЭГ терминах ГО, а также подсписки генов из входного списка, которые относятся к обогащенным терминам ГО. На следующем шаге происходит поиск пересечения районов связывания ТФ из входного списка с промоторами генов из подсписков. Для этого используется функция TF-targets, разработанная нами ранее в рамках создания программы CisCross-FindTFnet (Omelyanchuk et al., 2024). В качестве выходных данных пользователь получает файл с информацией об обогащенных терминах ГО, ассоциированных с ними ДЭГ, кодах обоснования аннотаций, ТФ, районы связывания которых в геноме картируются в 5′-регуляторные районы ДЭГ, ассоциированных с обогащенными терминами генной онтологии.

Ядро программы PlantReg реализовано на языке Perl с привлечением R-пакета clusterProfiler. PlantReg доступна пользователям через веб-интерфейс (https://plamorph.sysbio.ru/fannotf/). В веб-версии программы PlantReg для поиска генов-мишеней ТФ предоставляются две коллекции полногеномных профилей связывания ТФ. Первая, GTRD-MACS2, включает 306 наборов пиков ChIP-seq для 131 ТФ *A. thaliana*, загруженных в формате BED из базы данных GTRD (https://gtrd.biouml.org/#!) (Kolmykov et al., 2021). Вторая, CisCross-MACS2, получена путем широкомасштабного профилирования сайтов связывания ТФ *A. thaliana* методом DAP-seq (O'Malley et al., 2016) и представляет собой результат повторной обработки сырых данных из оригинального исследования (Lavrekha et al., 2022). Эта коллекция содержит 608 наборов пиков для 404 ТФ *A. thaliana*. Для определения 5′-регуляторных районов генов (500, 1000, 1500, 2000 или 2500 п. н. выше старта транскрипции) в веб-версии PlantReg используется аннотация генома *A. thaliana* ARAPORT11 (https://bar.utoronto.ca/thalemine/begin.do).

Реконструкция РСТФ ответа на солевой стресс. Для реконструкции РСТФ мы использовали публично доступные данные секвенирования транскриптомов семидневных проростков *A. thaliana* (экотип Col-0), выращенных на свету, до и после солевого стресса (100 мМ NaCl) длительностью 1 ч (Wu et al., 2021a). В качестве ДЭГ отбирали гены, изменение экспрессии которых под действием солевого стресса было достоверным при FDR < 0.05, среди них были выделены активируемые и подавляемые ДЭГ: аДЭГ и пДЭГ соответственно. Реконструкцию РСТФ осуществляли с помощью программы CisCross-FindTFnet (Omelyanchuk et al., 2024) со следующими параметрами: для картирования районов связывания ТФ использовали библиотеку пиков CisCross-MACS2, длину 5′-регуляторных районов генов установили как 1000 п. н., позиции стартов транскрипции определяли в соответствии с аннотацией генома *A. thaliana* ARAPORT11. Анализ обогащения районов связывания ТФ в 5′-регуляторных областях аДЭГ и пДЭГ проводили, контролируя FDR на уровне 0.001 по методу Бенджамини–Хохберга. Для реконструкции пар «ТФ-регулятор–ТФ-мишень» в составе РСТФ использовали наборы пиков, полученные в результате связывания ТФ с нативной геномной ДНК листьев, содержащей метильные метки.

Реконструкцию связей между РСТФ раннего ответа на солевой стресс и контролируемыми ею биологическими процессами осуществляли с помощью PlantReg. На вход программе подавали список ТФ, входящих в состав РСТФ ответа на солевой стресс, а также список генов, дифференциально экспрессирующихся в ответ на обработку солью (аДЭГ и пДЭГ анализировали по отдельности). Для 5′-регуляторных районов ДЭГ была установлена длина 1000 п. н., для картирования районов связывания ТФ в 5′-регуляторных районах генов использовали библиотеку CisCross-MACS2. Далее для анализа и интерпретации брали только пары «ТФ-регулятор–ген-мишень», реконструированные на основании DAP-seq-профилей связывания ТФ с геномной ДНК листьев, содержащей метильные метки.

Результаты и обсуждение

Биологическая интерпретация выходных данных PlantReg

Программа PlantReg разработана для реконструкции молекулярных механизмов генетической регуляции признаков. Для начала работы пользователю необходимо иметь список известных или предполагаемых регуляторов дифференциальной экспрессии генов. PlantReg осуществляет функциональную аннотацию списка ДЭГ, далее среди ДЭГ, ассоциированных с обогащенными биологическими процессами, выполняет поиск потенциальных мишеней для ТФ. Картирование районов связывания ТФ в 5′-регуляторных районах генов осуществляется с использованием представительной коллекции полногеномных профилей связывания ТФ исследуемого организма. В веб-версии программы PlantReg доступны две коллекции полногеном-

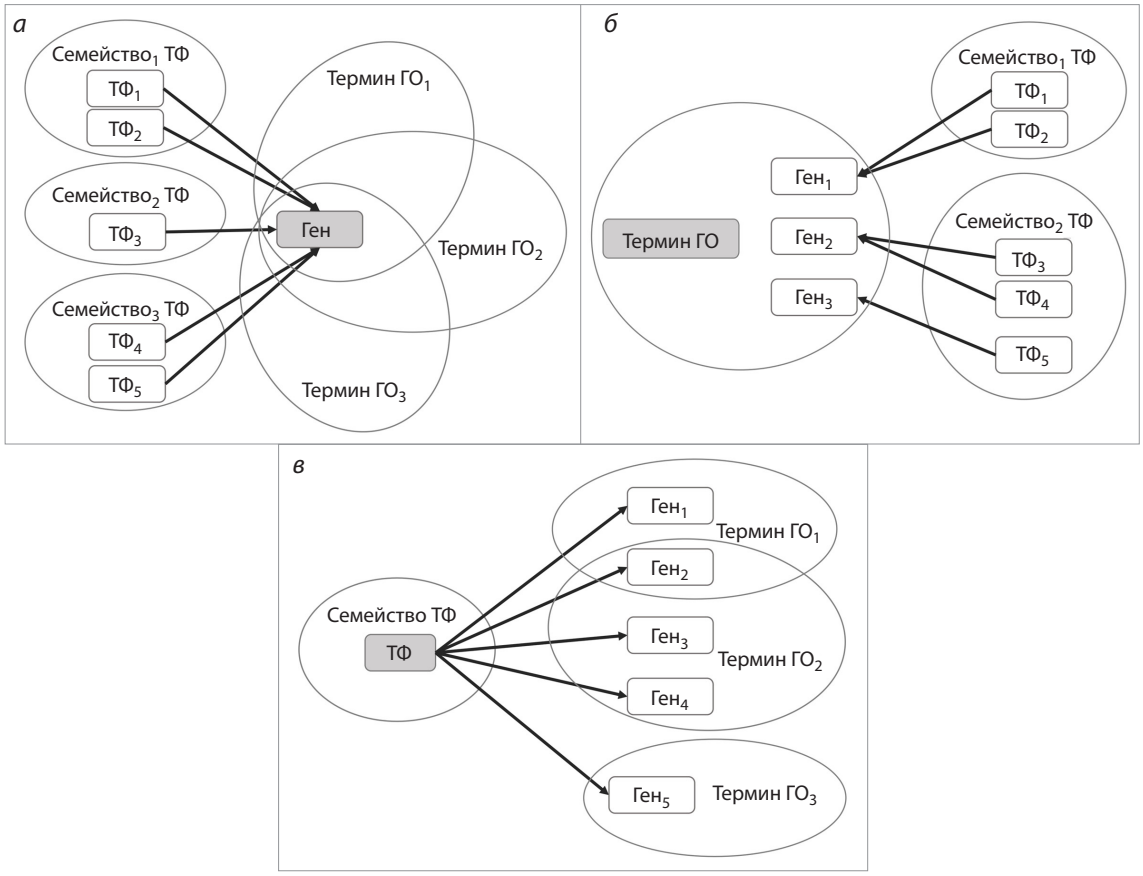


Рис. 2. Альтернативные представления выходных данных программы PlantReg.

Панели *а*, *б* и *в* соответствуют блокам выдачи 1, 2 и 3, описанным в тексте. Серым цветом выделен центральный элемент выдачи.

ных профилей связывания ТФ *A. thaliana*, полученных на основании анализа данных, ChIP-seq или DAP-seq. Пользователь может выбрать для работы одну из них. На выходе программа устанавливает взаимосоответствие между ДЭГ, регулирующими их экспрессию ТФ и обогащенными терминами генной онтологии.

Для удобства биологической интерпретации и последующего анализа полученных результатов выходные данные PlantReg представлены пятью блоками. Первые четыре блока предлагают четыре альтернативных представления одних и тех же результатов. Так, блок 1 характеризует гены. Он содержит подсписок ДЭГ, аннотированных обогащенными терминами ГО, со списками потенциальных транскрипционных регуляторов для каждого гена, их количеством и указанием для каждого ТФ принадлежности к определенному семейству (рис. 2, *а*). Для каждого гена также отмечаются общее количество и список обогащенных терминов ГО с кодом обоснования аннотации, что облегчает выявление среди ДЭГ генов, вовлеченных в широкий спектр биологических процессов, и генов, специфичных для конкретных процессов.

Центральный элемент блока 2 – биологические процессы. В этом блоке для каждого обогащенного термина ГО определен подсписок ассоциированных с ним ДЭГ с кодами обоснования аннотации, а также подсписок ТФ, потенциально регулирующих дифференциальную экспрессию этих ДЭГ, с указанием принадлежности к определенному семейству (см. рис. 2, *б*). Этот блок выдачи данных позволяет реконструировать механизм генетической регуляции для заданного биологического процесса.

Блок выдачи 3 характеризует транскрипционные регуляторы дифференциальной экспрессии генов. Он содержит список ТФ, для которых среди ДЭГ были найдены гены-мишени, ассоциированные с обогащенными терминами ГО (см. рис. 2, *в*). Этот блок выдачи данных полезен при планировании экспериментов для верификации предсказанных механизмов генетической регуляции биологических процессов.

Блок выдачи 4 содержит таблицу, где в каждой строчке приведен один ген, один из ТФ, потенциально регулирующих экспрессию этого гена, семейство ТФ и один из терминов ГО с кодом обоснования аннотации. Эта выдача может быть использована для дальнейшего анализа с применением программных средств.

Блок выдачи 5, вспомогательный, содержит результаты функциональной аннотации списка ДЭГ с помощью программы clusterProfiler с указанием значимости обогащения терминов генной онтологии.

Функциональная аннотация РСТФ ответа на солевой стресс у *A. thaliana*

Мы использовали программу PlantReg для исследования механизмов, регулирующих ответ на солевой стресс у модельного вида растений *A. thaliana*. Список ДЭГ,

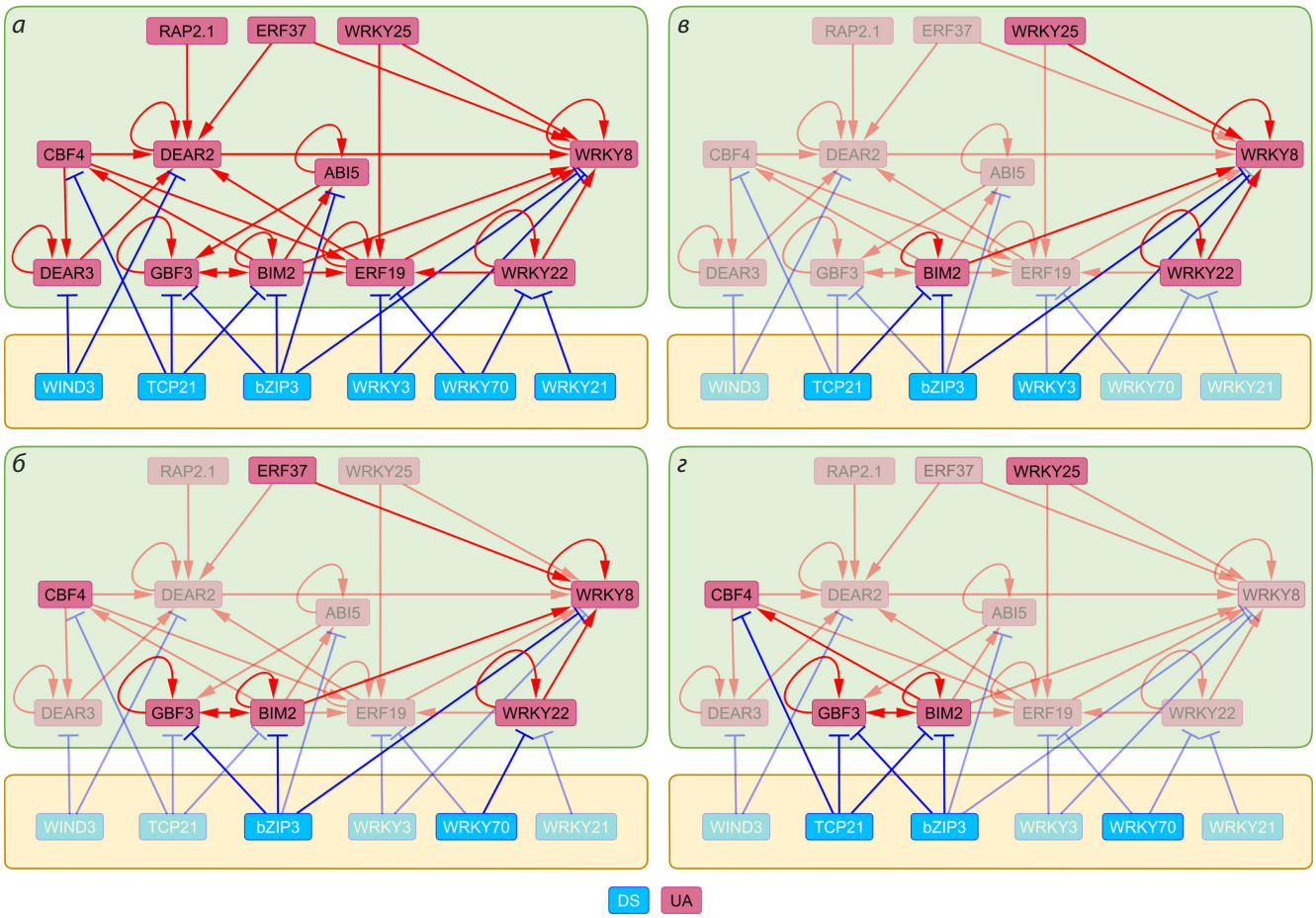


Рис. 3. Реконструированная РСТФ ответа на ранний солевой стресс у *A. thaliana* (*а*) и ее участие в регуляции процессов, сопровождающих солевой стресс: ответа эндоплазматического ретикулума на потерю третичной структуры белков (*б*), биосинтеза индол-содержащих соединений (*в*) и S-гликозидов (*г*).

Вершины графов соответствуют транскрипционным факторам. ТФ1 и ТФ2 соединены дугой, направленной от ТФ1 (регулятора) к ТФ2 (мишени), если в 5′-регуляторном районе гена, кодирующего ТФ2, картируется пик связывания ТФ1. Блок зеленого цвета выделяет группу ТФ (UA), которые активируются и активируют свои мишени в ответ на солевой стресс. Блок желтого цвета выделяет группу ТФ (DS), которые подавляли гены в норме и сами подавляются солевым стрессом, и, таким образом, происходит пассивная активация ранее подавляемых ими генов. На панелях *б–г* выделены вершины и ребра РСТФ, вовлеченные в регуляцию соответствующих процессов, сопровождающих солевой стресс. DS и UA – подавляемые супрессоры и стимулируемые активаторы, согласно (Omelyanchuk et al., 2024).

реагирующих на высокое содержание соли в среде, был взят из публично доступных данных секвенирования транскриптомов (Wu et al., 2021a). Для того чтобы сформировать список потенциальных ТФ-регуляторов этих генов, мы применяли разработанный нами ранее инструмент CisCross-FindTFnet (Omelyanchuk et al., 2024). На основании совместного анализа ДЭГ и полногеномных профилей связывания ТФ этот инструмент выявляет потенциальные ТФ-регуляторы ДЭГ, классифицирует их по типу регуляции, а также определяет связи между ними и реконструирует РСТФ.

Классы ТФ по типу регуляции выделяются на базе определенного набора правил и соответствуют следующим свойствам регуляторов: US (upregulated suppressor) – индуцируемый стимулом (в нашем случае высокой концентрацией соли) репрессор. Он подавляет экспрессию генов-мишеней, которые были активны до воздействия стимула. UA (upregulated activator) индуцируется стимулом и активирует экспрессию генов-мишеней. DA (downregulated activator) и DS (downregulated suppressor) активны в отсутствие стимула. Появление стимула снижает экспрессию DA, а следовательно, и экспрессию его генов-мишеней. DS подавляют работу генов-мишеней в отсутствие стимула; под действием стимула экспрессия DS снижается, и активность его мишеней деблокируется.

Структура РСТФ раннего ответа на солевой стресс, реконструированной с помощью программы CisCross-FindTFnet, представлена на рис. 3, *а* и состоит только из ТФ, сайты связывания которых были обогащены в аДЭГ, т. е. реакция на солевой стресс начинается с активации транскрипции, причем ТФ в составе РСТФ относятся только к DS и UA, – активация гена происходит пассивно из-за подавления стрессом супрессора (DS) или активно за счет стимулирования активатора (UA). Среди генов, кодирующих предсказанные UA, увеличение активности при солевом стрессе было ранее экспериментально показано для *CBF4/DREB1D* (Sakuma et al., 2002), *ERF37/DREB A-4* (Hossain et al., 2016), *RAP2.1/DEAR6* (Ghorbani et al., 2019),

WRKY25 (Jiang, Deyholos, 2009), *ABI5* (Yuan et al., 2011), *GBF3* (Zhang L. et al., 2012, 2017) и *WRKY8* (Hu et al., 2013). При этом *ABI5* (Yuan et al., 2011), *GBF3* (Zhang L. et al., 2012, 2017) и *WRKY8* (Hu et al., 2013) играют ключевую роль в ответе на солевой стресс.

Для обнаруженных нами DS ранее было продемонстрировано, что инактивация *WRKY70* увеличивала устойчивость растений к солевому стрессу (Li J. et al., 2013), а экспрессия *bZIP3* ингибировалась солевым стрессом (Liu Y. et al., 2013). Примечательно, что более продолжительный солевой стресс (4 ч) активировал *WRKY3* (Li P. et al., 2021). Таким образом, состав реконструированной нами РСТФ достаточно хорошо согласуется с литературными данными. При этом только для 4 из 18 ТФ (22 %) ранее была выявлена ключевая роль в солевом стрессе, и всего для 10 (56 %) была описана реакция на солевой стресс, т. е. РСТФ содержит ряд новых потенциальных регуляторов этого процесса.

Функциональная аннотация списков ДЭГ показала, что ранняя реакция на солевой стресс сопровождается ответом эндоплазматического ретикулума на потерю третичной структуры белков, а также активацией следующих процессов: запрограммированной гибели клеток, старения листьев, транспорта воды, биосинтеза индол-содержащих соединений и S-гликозидов, ответа на тепло, красный и дальний красный свет, абсцизовую, салициловую и жасмоновую кислоты, синий свет, гипоксию, активные формы кислорода, обезвоживание. Связь ответа на солевой стресс и ответа на тепло была обнаружена ранее, так как белки теплового шока усиливают устойчивость к солевому стрессу, и наоборот: сверхэкспрессия белков солевого стресса обеспечивает устойчивость к тепловому стрессу (Azameti et al., 2024; Chaffai et al., 2024; Chang et al., 2024). Также была продемонстрирована в экспериментах связь солевого стресса со старением листьев, гипоксией, транспортом воды, ответом на голубой, красный и дальний красный свет, активные формы кислорода, обезвоживание, абсцизовую, салициловую и жасмоновую кислоты (Serraj et al., 1994; Szepesi et al., 2009; Joseph, Jini, 2010; Khan et al., 2012; Kumar et al., 2014; Sharma et al., 2022; Kesawat et al., 2023; Lu, Fricke, 2023; Tan et al., 2023; Peng et al., 2024).

Солевой стресс приводит к нарушению укладки третичной структуры белков в эндоплазматическом ретикулуме (так называемому стрессу эндоплазматического ретикулума), и ответом на это является некоторая оптимизация формирования третичной структуры белков, выражающаяся в снижении накопления белков с нарушениями третичной структуры (Liu et al., 2007; Wang et al., 2011). Имеются данные об участии биосинтеза такого индол-содержащего соединения, как мелатонин, в ответе на солевой стресс (Qi et al., 2020; Shamloo-Dashtpagerdi et al., 2022). Обогащение генов ответа на солевой стресс термином генной онтологии «метаболизм S-гликозидов» было замечено ранее (Rodriquez et al., 2021).

В регуляцию ответа на тепло, красный и дальний красный свет и салициловую кислоту оказались вовлечены все ТФ сети. Остальные биологические процессы распались на две группы: контролируемые не менее 70 % и контролируемые менее 50 % ТФ сети. В первую группу попали запрограммированная гибель клеток, старение листьев и ответы на синий свет, гипоксию, активные формы кислорода, обезвоживание, абсцизовую и жасмоновую кислоты. Ко второй группе были отнесены ответ эндоплазматического ретикулума на потерю третичной структуры белков (см. рис. 3, *б*), биосинтез индол-содержащих соединений (см. рис. 3, *в*) и S-гликозидов (см. рис. 3, *г*), а также транспорт воды (последний регулируют лишь три ТФ сети, BIM2, bZIP3 и WIND3). Таким образом, с помощью PlantReg мы показали, что реакция на солевой стресс сопровождается как процессами, регулируемыми всей сетью ТФ, так и процессами, контролируемыми отдельными частями этой сети.

До настоящего времени среди ТФ, выявленных нами как контролирующих ответ эндоплазматического ретикулума на потерю третичной структуры белков, только для WRKY70 была продемонстрирована роль в этом процессе (Wang L.Y. et al., 2023), а bZIP3 был указан как один из возможных кандидатов на эту роль (Ko et al., 2023).

Глюкозинолаты – наиболее разнообразная и изученная группа среди S-гликозидов. Они представляют собой вторичные метаболиты растений семейства Крестоцветных, участвующие в защите растений от патогенов (Halkier, Gershenzon, 2006). В настоящее время они интенсивно изучаются благодаря их лечебным и профилактическим свойствам в отношении онкологических, сердечно-сосудистых или неврологических заболеваний. Глюкозинолаты подразделяются на три группы в зависимости от аминокислот, из которых они производятся. Алифатические глюкозинолаты образуются из метионина, аланина, лейцина, изолейцина и валина, ароматические – из фенилаланина и тирозина, а индольные – из триптофана. По крайней мере для трех из семи ТФ, которые, согласно нашим данным, контролируют биосинтез глюкозинолатов, эта функция была известна ранее. CBF4 запускает синтез алифатических глюкозинолатов, которые к тому же увеличивают устойчивость к солевому стрессу (Defoort et al., 2018), в то время как WRKY70 подавляет биосинтез индол-3-улметил глюкозинолата (Li J. et al., 2006). Экспрессия *GBF3* значительно уменьшена в мутантах по гену *SUR2*/*CYP83B1*, который контролирует точку метаболического разветвления между биосинтезом ауксина и индол-глюкозинолатов (Morant et al., 2010).

Регуляция сигнального пути абсцизовой кислоты при солевом стрессе у *A. thaliana*

Наряду с определением состава ТФ, контролирующих определенные процессы, PlantReg позволяет вычленять ТФ, регулирующие активность отдельных генов в этих процессах. Последнее предоставляет возможность выделить модуляторы экспрессии генов по стадиям изучаемых процессов. В данной статье мы демонстрируем это на примере реконструкции механизма транскрипционной регуляции ответа на АБК при солевом стрессе. Согласно результатам PlantReg, все ТФ в составе РСТФ ответа на солевой стресс, за исключением WRKY21, контролируют ответ на АБК. Эта регуляция начинается с контроля уровня АБК в клетке.

На этой стадии (стадия 1 на рис. 4) мишенями РСТФ являются гены *ABCG25* и *ATAF1*, соответственно коди-

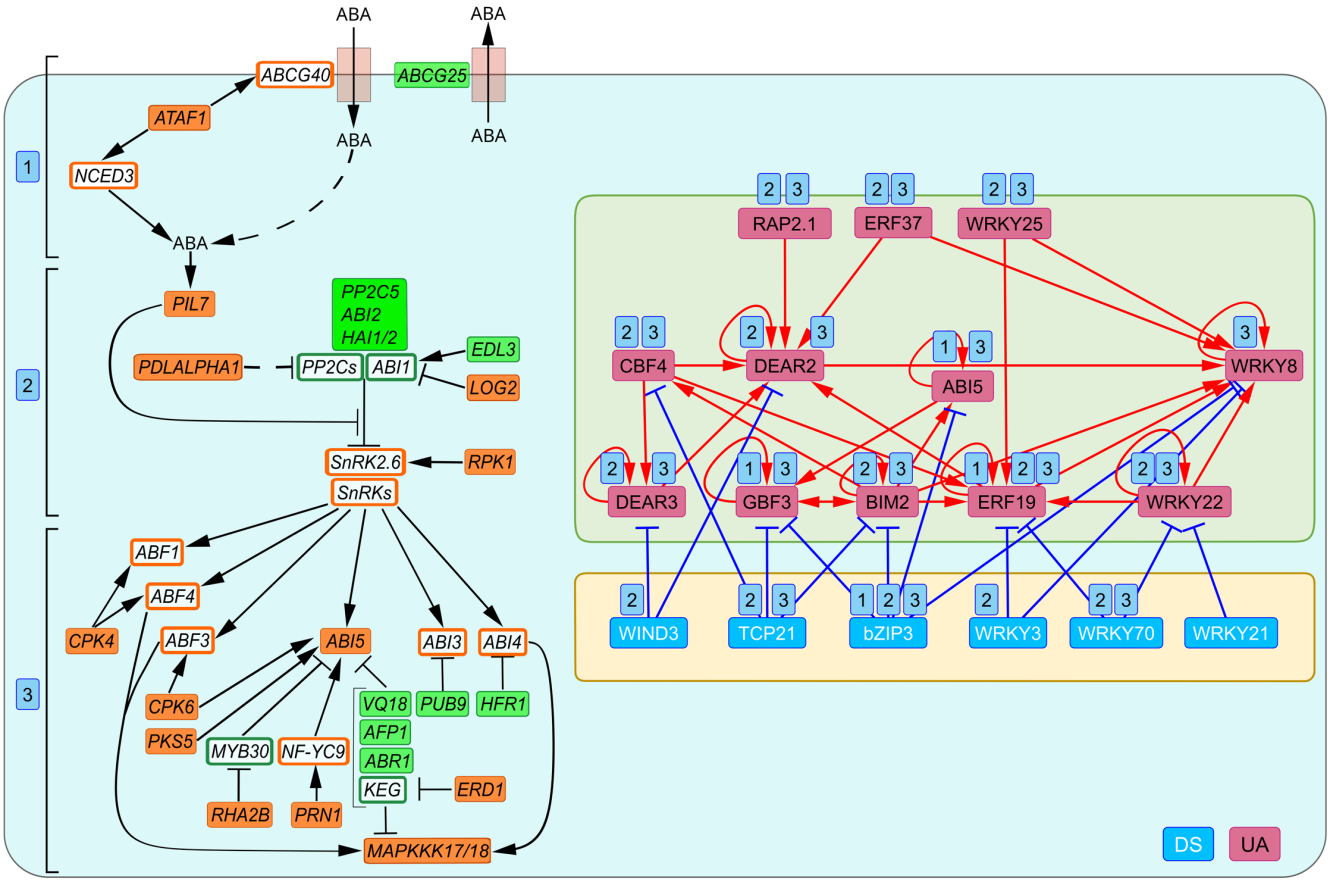


Рис. 4. Транскрипционная регуляция уровня и сигнального пути АБК в условиях раннего солевого стресса.

Зеленые и оранжевые прямоугольники обозначают аДЭГ, которые соответственно кодируют репрессоры и активаторы уровня АБК в клетке и сигнального пути АБК, являющиеся потенциальными мишенями ТФ из РСТФ. Белые прямоугольники в зеленой и оранжевой рамках соответствуют генам-репрессорам и генам-активаторам уровня АБК в клетке и сигнального пути АБК, которые не являются потенциальными мишенями ТФ из РСТФ. Числа в голубых прямоугольниках обозначают различные стадии контроля уровня и сигнального пути АБК: 1 – стадия контроля уровня АБК; 2 – стадия восприятия АБК рецепторами; 3 – стадия активации главных ТФ ответа на АБК. Аббревиатуры для генов транспорта, биосинтеза и передачи сигнала АБК: *ATP-BINDING CASETTE G25/40* (*ABCG25/40*), *PYR1 LIKE 7* (*PYL7*), *PROTEIN PHOSPHATASES TYPE 2C* (*PP2Cs*), *ABA INSENSITIVE1/2/3/4/5* (*ABI1/2/3/4/5*), *SNF1-RELATED PROTEIN KINASE* (*SnRKs*), *ABSCISIC ACID RESPONSIVE ELEMENT-BINDING FACTOR1/3/4* (*ABF1/3/4*), *CALCIUM-DEPENDENT PROTEIN KINASE 4/6* (*CPK4/6*), *ABI FIVE BINDING PROTEIN 1* (*AFP1*), *KEEP ON GOING* (*KEG*), *ENHANCED DISEASE RESISTANCE 1* (*EDR1*), *NUCLEAR FACTOR Y9* (*NF-YC9*), *PLANT U-BOX/ARM-REPEAT* (*ATPUB-ARM*) *E3 LIGASE 9* (*PUB9*), *ABA REPRESSOR 1* (*ABR1*), *VQ PROTEIN 18* (*VQ18*), *HIGHLY ABA-INDUCED PP2C GENE 1/2* (*HAI1/2*), *ARABIDOPSIS THALIANA ACTIVATING FACTOR1* (*ATAF1*), *EID1-LIKE 3* (*EDL3*), *LONG HYPOCOTYL IN FAR-RED 1* (*HFR1*), *LOSS OF GDU2* (*LOG2*), *MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 17/18* (*MAPKKK17/18*), *PHOSPHOLIPASE D ALPHA 1* (*PLDALPHA1*), *PIRIN 1* (*PRN1*), *RING-H2 FINGER PROTEIN 2B* (*RHA2B*), *RECEPTOR-LIKE PROTEIN KINASE 1* (*RPK1*), *CALCINEURIN B-LIKE PROTEIN-INTERACTING PROTEIN KINASES/SOS2-LIKE PROTEIN KINASE* (*PKS5*), *MYB DOMAIN PROTEIN 30* (*MYB30*), *NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 3* (*NCED3*).

рующие экспортер АБК из клетки (Park et al., 2016) и ТФ, активирующий гены биосинтеза АБК *NCED3* (Jensen et al., 2013) и импортера АБК *ABCG40* (Kang et al., 2015).

На следующей стадии (стадия 2 на рис. 4) АБК связывается и активирует группу рецепторов PYRABACTIN RESISTANCE1/PYR1 LIKE/REGULATORY COMPONENTS OF ABA RECEPTORS (PYR/PYL/RCAR) (Fidler et al., 2022), среди которых РСТФ ответа на солевой стресс контролирует *PYL7*, причем это наиболее строго регулируемый ген сигнального пути АБК, так как его экспрессию контролирует половина ТФ из состава РСТФ (9 из 18). В нормальных условиях активность *PYL7* подавляют bZIP3 и WIND3. Причем, если bZIP3 до начала солевого стресса ингибирует активность 11 из 24 генов сигнального пути АБК, включаемых солевым стрессом, то WIND3 – специфический супрессор *PYL7*. Солевой стресс активирует *PYL7* с помощью семи ТФ, образующих регуляторный контур, хабом которого является *DEAR2*, активируемый напрямую пятью ТФ (CBF4, DEAR3, ERF19, ERF37, RAP2.1), и шестой ТФ (WRKY22) стимулирует его через ERF19. В сигнальном пути АБК PYR/PYL/RCAR рецепторы ингибируют активность PP2C фосфатаз, тем самым предотвращая дефосфорилирование SnRK2 киназ (Fidler et al., 2022).

Прямыми мишенями РСТФ ответа на солевой стресс являются PP2C фосфатазы PP2C5, ABI2 и HAI/2, а также активатор SNRK2.6 – RPK1 (Shang et al., 2020), регуляторы PP2C фосфатаз – EDL3 (Koops et al., 2011), LOG2 (Pan W. et al., 2020) и ген фосфолипазы PLDALPHA1, продукт которой (фосфатидная кислота) ингибирует активность некоторых PP2C фосфатаз (Ndathe, Kato, 2024).

Третья стадия передачи сигнала АБК (стадия 3 на рис. 4) начинается с активации SnRK2 киназами главных ТФ ответа на АБК. Примечательно, что один из них, ABI5, также

входит в РСТФ ответа на солевой стресс. За исключением *ABI5* и инициаторов MAPK каскада *MAPKKK17/18* (Zhou M. et al., 2021; Zhao et al., 2023), все остальные мишени РСТФ на этом этапе представляют собой гены, кодирующие регуляторы активности главных ТФ ответа на АБК. Среди них киназы CPK4/6, PKS5, EDR1 (Zhu et al., 2007; Wawrzynska et al., 2008; Zhou X. et al., 2015; Zhang H. et al., 2020), транскрипционные факторы ABR1 (Sanyal, Pandey, 2024) и HFR1 (Wang Z. et al., 2024), транскрипционные регуляторы VQ18 (Pan J. et al., 2018) и PRN1 (Warpeha et al., 2007), компоненты комплексов по деградации белков PUB9 (Samuel et al., 2008), AFP1 (Lopez-Molina et al., 2003), RHA2B (Li H. et al., 2011).

Необходимо отметить, что в контроле третьей стадии сигнального пути участвует половина супрессоров РСТФ ответа на солевой стресс и все активаторы из этой сети. Причем оба хаба РСТФ, WRKY8 и DEAR2, транскрипция которых в этой сети усиливается семью и шестью ТФ активаторами соответственно, задействованы на этой стадии. Более того, в то время как DEAR2 имеет мишени и на стадии 2, WRKY8 специфичен для стадии 3. Было показано, что при вирусной инфекции WRKY8 контролирует сигнальный путь АБК как подавляемый инфекцией активатор транскрипции *ABI4* (Chen L. et al., 2013). Мы продемонстрировали, что при солевом стрессе WRKY8 контролирует сигнальный путь АБК через усиление транскрипции *CPK6*. Киназа CPK6, как было выявлено ранее, стимулирует активность ТФ ABF4 и ABI5 через их фосфорилирование (Zhang H. et al., 2020). Это дает возможность предположить, что один и тот же ТФ может иметь разные мишени в сигнальном пути АБК при различных стрессах.

Использование программы PlantReg позволило определить мишени РСТФ ответа на солевой стресс в генах ответа на АБК и показать, что регуляция солевым стрессом ответа на этот гормон осуществляется в основном за счет регуляции им сигнального пути АБК, в котором наиболее строгий контроль происходит на стадии активации главных ТФ ответа на АБК, ABF1/3/4 и ABI3/4/5. Более того, ABI5, один из главных ТФ сигнального пути АБК, является также одним из ТФ в РСТФ ответа на солевой стресс, где его активность до стресса подавляется bZIP3, а при стрессе стимулируется BIM2. Сам же ABI5 активирует *GBF3*, который, как и *BIM2*, подавляется bZIP3 до стресса. При этом GBF3 и BIM2 взаимно активируют друг друга. Таким образом, BIM2, bZIP3, GBF3 и ABI5 образуют четкий регуляторный контур в реконструированной нами РСТФ (см. рис. 3, *а* и 4).

Интересно, что в генной сети ответа на АБК в работе (Aerts et al., 2024) ТФ, составляющие этот регуляторный контур (BIM2, bZIP3, GBF3 и ABI5), принадлежат к группе наиболее ранних регуляторов и имеют большое число общих мишеней, т. е. контролируют одни и те же гены. Наряду с BIM2, bZIP3, GBF3 и ABI5, реконструированная нами РСТФ ответа на солевой стресс пересекается с генной сетью ответа на абсцизовую кислоту из статьи (Aerts et al., 2024) еще по трем ТФ: CBF4, DEAR2 и WRKY3. Мы обнаружили, что DEAR2 является хабом активации при солевом стрессе второго и третьего этапа сигнального пути АБК. Более того, CBF4, DEAR2 и WRKY3 – это компоненты сети, соединяющей ее центральный активирующий регуляторный контур (BIM2, GBF3 и ABI5) со вторым хабом активации сигнала АБК, WRKY8.

WRKY3 наряду с bZIP3 подавляет WRKY8 до стресса (см. рис. 3, *а*). При стрессе происходит последовательная активация BIM2, CBF4, DEAR2 и WRKY8. Таким образом, сравнение результатов PlantReg с генной сетью ответа на абсцизовую кислоту (Aerts et al., 2024) позволило выделить ТФ, которые являются основными регуляторами ответа на АБК. Оставшиеся ТФ, RAP2.1, ERF19/37, DEAR3, TCP21, WRKY8/22/25/70, возможно, вовлечены в контроль сигнального пути АБК только при солевом стрессе.

Заключение

Программа PlantReg показала свою эффективность в системном анализе результатов полногеномных экспериментов по выявлению дифференциальной активности генов. Она позволяет наряду с функциональной аннотацией ДЭГ находить среди них мишени ТФ и, исходя из этого, определять ТФ, регулирующие определенные биологические процессы. Совмещение результатов PlantReg с программами, реконструирующими РСТФ (например, CisCross-FindTFnet), дает возможность разбивать РСТФ на подсети, контролирующие каждый из процессов, выделять ключевые ТФ в этих процессах и даже на отдельных стадиях этих процессов. Подходы и методы, разработанные при создании PlantReg, могут быть с успехом использованы для реконструкции механизмов транскрипционной регуляции биологических процессов у различных видов живых организмов.

Список литературы / References

Aerts N., Hickman R., Van Dijken A.J., Kaufmann M., Snoek B.L., Pieterse C.M., Van Wees S.C. Architecture and dynamics of the abscisic acid gene regulatory network. *Plant J*. 2024;119(5):2538-2563. doi 10.1111/tpj.16899

Azameti M.K., Tanuja N., Kumar S., Rathinam M., Imoro A.W.M., Singh P.K., Gaikwad K., Sreevathsa R., Dalal M., Arora A., Rai V., Padaria J.C. Transgenic tobacco plants overexpressing a wheat salt stress root protein (TaSSRP) exhibit enhanced tolerance to heat stress. *Mol. Biol. Rep*. 2024;51(1):791. doi 10.1007/s11033-024-09755-4

Chaffai R., Ganesan M., Cherif A. Mechanisms of plant response to heat stress: recent insights. In: Plant Adaptation to Abiotic Stress: From Signaling Pathways and Microbiomes to Molecular Mechanisms. Singapore: Springer, 2024;83-105. doi 10.1007/978-981-97-0672-3_5

Chang H., Wu T., Shalmani A., Xu L., Li C., Zhang W., Pan R. Heat shock protein HvHSP16.9 from wild barley enhances tolerance to salt stress. *Physiol. Mol. Biol. Plants*. 2024;30(5):687-704. doi 10.1007/s12298-024-01455-4

Chen J.W., Shrestha L., Green G., Leier A., Marquez-Lago T.T. The hitchhikers' guide to RNA sequencing and functional analysis. *Brief. Bioinform.* 2023;24(1):bbac529. doi 10.1093/bib/bbac529

Chen L., Zhang L., Li D., Wang F., Yu D. WRKY8 transcription factor functions in the TMV-cg defense response by mediating both abscisic acid and ethylene signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA*. 2013;110(21):E1963-E1971. doi 10.1073/pnas.1221347110

Defoort J., Van de Peer Y., Vermeirssen V. Function, dynamics and evolution of network motif modules in integrated gene regulatory networks of worm and plant. *Nucleic Acids Res*. 2018;46(13):6480-6503. doi 10.1093/nar/gky468

Fidler J., Graska J., Gietler M., Nykiel M., Prabucka B., Rybarczyk-Płońska A., Muszyńska E., Morkunas I., Labudda M. PYR/PYL/RCAR receptors play a vital role in the abscisic-acid-dependent responses of plants to external or internal stimuli. *Cells*. 2022;11(8):1352. doi 10.3390/cells11081352

Ghorbani R., Alemzadeh A., Razi H. Microarray analysis of transcriptional responses to salt and drought stress in *Arabidopsis thaliana*. *Heliyon*. 2019;5(11):e02614. doi 10.1016/j.heliyon.2019.e02614

Halkier B.A., Gershenzon J. Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* 2006;57(1):303-333. doi 10.1146/annurev.arplant.57.032905.105228

He X., Zhang J. Why do hubs tend to be essential in protein networks? *PLoS Genet*. 2006;2(6):88. doi 10.1371/journal.pgen.0020088

Hecker D., Lauber M., Behjati Ardakani F., Ashrafiyan S., Manz Q., Kersting J., Hoffmann M., Schulz M.H., List M. Computational tools for inferring transcription factor activity. *Proteomics*. 2023;23(23-24):2200462. doi 10.1002/pmic.202200462

Hossain M.A., Henríquez-Valencia C., Gómez-Páez M., Medina J., Orellana A., Vicente-Carbajosa J., Zouhar J. Identification of novel components of the unfolded protein response in Arabidopsis. *Front. Plant Sci*. 2016;7:650. doi 10.3389/fpls.2016.00650

Hu Y., Chen L., Wang H., Zhang L., Wang F., Yu D. Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. *Plant J*. 2013;74(5):730-745. doi 10.1111/tpj.12159

Jensen M.K., Lindemose S., De Masi F., Reimer J.J., Nielsen M., Perera V., Workman C.T., Turck F., Grant M.R., Mundy J., Petersen M., Skriver K. ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene *NCED3* in *Arabidopsis thaliana*. *FEBS Open Bio*. 2013;3:321-327. doi 10.1016/j.fob.2013.07.006

Jiang Y., Deyholos M.K. Functional characterization of Arabidopsis NaCl-inducible *WRKY25* and *WRKY33* transcription factors in abiotic stresses. *Plant Mol. Biol*. 2009;69(1-2):91-105. doi 10.1007/s11103-008-9408-3

Joseph B., Jini D. Salinity induced programmed cell death in plants: challenges and opportunities for salt-tolerant plants. *J. Plant Sci*. 2010;5(4):376-390. doi 10.3923/jps.2010.376.390

Kang J., Yim S., Choi H., Kim A., Lee K.P., Lopez-Molina L., Martinoia E., Lee Y. Abscisic acid transporters cooperate to control seed germination. *Nat. Commun*. 2015;6(1):8113. doi 10.1038/ncomms9113

Kesawat M.S., Satheesh N., Kherawat B.S., Kumar A., Kim H.U., Chung S.M., Kumar M. Regulation of reactive oxygen species during salt stress in plants and their crosstalk with other signaling molecules – current perspectives and future directions. *Plants*. 2023;12(4):864. doi 10.3390/plants12040864

Khan M.I.R., Syeed S., Nazar R., Anjum N.A. An insight into the role of salicylic acid and jasmonic acid in salt stress tolerance. In: Khan N., Nazar R., Iqbal N., Anjum N. (Eds). Phytohormones and Abiotic Stress Tolerance in Plants. Berlin; Heidelberg: Springer, 2012;277-300. doi 10.1007/978-3-642-25829-9_12

Ko D.K., Kim J.Y., Thibault E.A., Brandizzi F. An IRE1-proteasome system signalling cohort controls cell fate determination in unresolved proteotoxic stress of the plant endoplasmic reticulum. *Nat. Plants*. 2023;9(8):1333-1346. doi 10.1038/s41477-023-01480-3

Kolmykov S., Yevshin I., Kulyashov M., Sharipov R., Kondrakhin Y., Makeev V.J., Kulakovskiy I.V., Kel A., Kolpakov F. GTRD: an integrated view of transcription regulation. *Nucleic Acids Res*. 2021;49:104-111. doi 10.1093/nar/gkaa1057

Koops P., Pelser S., Ignatz M., Klose C., Marrocco-Selden K., Kretsch T. EDL3 is an F-box protein involved in the regulation of abscisic acid signalling in *Arabidopsis thaliana*. *J. Exp. Bot*. 2011;62(15):5547-5560. doi 10.1093/jxb/err236

Kumar M., Kumari P., Reddy C.R.K., Jha B. Salinity and desiccation induced oxidative stress acclimation in seaweeds. *Adv. Bot. Res.* 2014;71:91-123. doi 10.1016/B978-0-12-408062-1.00004-4

Lavrekha V.V., Levitsky V.G., Tsukanov A.V., Bogomolov A.G., Grigorovich D.A., Omelyanchuk N., Ubogoeva E.V., Zemlyanskaya E.V., Mironova V. CisCross: A gene list enrichment analysis to predict upstream regulators in *Arabidopsis thaliana*. *Front. Plant Sci*. 2022;13:942710. doi 10.3389/fpls.2022.942710

Li H., Jiang H., Bu Q., Zhao Q., Sun J., Xie Q., Li C. The Arabidopsis RING finger E3 ligase RHA2b acts additively with RHA2a in regulating abscisic acid signaling and drought response. *Plant Physiol*. 2011;156(2):550-563. doi 10.1104/pp.111.176214

Li J., Brader G., Kariola T., Tapio Palva E. WRKY70 modulates the selection of signaling pathways in plant defense. *Plant J*. 2006;46(3):477-491. doi 10.1111/j.1365-313X.2006.02712.x

Li J., Besseau S., Törönen P., Sipari N., Kollist H., Holm L., Palva E.T. Defense-related transcription factors WRKY70 and WRKY54 modulate osmotic stress tolerance by regulating stomatal aperture in *Arabidopsis*. *New Phytol*. 2013;200(2):457-472. doi 10.1111/nph.12378

Li P., Li X., Jiang M. CRISPR/Cas9-mediated mutagenesis of WRKY3 and WRKY4 function decreases salt and Me-JA stress tolerance in *Arabidopsis thaliana*. *Mol. Biol. Rep*. 2021;48(8):5821-5832. doi 10.1007/s11033-021-06541-4

Liu J.X., Srivastava R., Che P., Howell S.H. Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant J*. 2007;51(5):897-909. doi 10.1111/j.1365-313X.2007.03195.x

Liu Y., Ji X., Zheng L., Nie X., Wang Y. Microarray analysis of transcriptional responses to abscisic acid and salt stress in *Arabidopsis thaliana*. *Int. J. Mol. Sci*. 2013;14(5):9979-9998. doi 10.3390/ijms14059979

Lopez-Molina L., Mongrand S., Kinoshita N., Chua N.-H. AFP is a novel negative regulator of ABA signaling that promotes ABI5 protein degradation. *Genes Dev*. 2003;17:410-418. doi 10.1101/gad.1055803

Lu Y., Fricke W. Salt stress – regulation of root water uptake in a whole-plant and diurnal context. *Int. J. Mol. Sci*. 2023;24(9):8070. doi 10.3390/ijms24098070

Morant M., Ekstrøm C., Ulvskov P., Kristensen C., Rudemo M., Olsen C.E., Hansen J., Jørgensen K., Jørgensen B., Møller B.L., Bak S. Metabolomic, transcriptional, hormonal, and signaling cross-talk in *superroot2*. *Mol. Plant*. 2010;3(1):192-211. doi 10.1093/mp/ssp098

Ndathe R., Kato N. Phosphatidic acid produced by phospholipase Dα1 and Dδ is incorporated into the internal membranes but not involved in the gene expression of *RD29A* in the abscisic acid signaling network in *Arabidopsis thaliana*. *Front. Plant Sci*. 2024;15:1356699. doi 10.3389/fpls.2024.1356699

Omelyanchuk N.A., Lavrekha V.V., Bogomolov A.G., Dolgikh V.A., Sidorenko A.D., Zemlyanskaya E.V. Computational reconstruction of the transcription factor regulatory network induced by auxin in *Arabidopsis thaliana* L. *Plants*. 2024;13(14):1905. doi 10.3390/plants13141905

O'Malley R.C., Huang S.S.C., Song L., Lewsey M.G., Bartlett A., Nery J.R., Galli M., Gallavotti A., Ecker J.R. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*. 2016;165:1280-1292. doi 10.1016/j.cell.2016.04.038

Pan J., Wang H., Hu Y., Yu D. Arabidopsis VQ18 and VQ26 proteins interact with ABI5 transcription factor to negatively modulate ABA response during seed germination. *Plant J.* 2018;95(3):529-544. doi 10.1111/tpj.13969

Pan W., Lin B., Yang X., Liu L., Xia R., Li J., Wu Y., Xie Q. The UBC27-AIRP3 ubiquitination complex modulates ABA signaling by promoting the degradation of ABI1 in Arabidopsis. *Proc. Natl. Acad. Sci. USA*. 2020;117(44):27694-27702. doi 10.1073/pnas.2007366117

Park Y., Xu Z.-Y., Kim S.Y., Lee J., Choi B., Lee J., Kim H., Sim H.-J., Hwang I. Spatial regulation of ABCG25, an ABA exporter, is an important component of the mechanism controlling cellular ABA levels. *Plant Cell*. 2016;28(10):2528-2544. doi 10.1105/tpc.16.00359

Peng Y., Zhu H., Wang Y., Kang J., Hu L., Li L., Zhu K., Yan J., Bu X., Wang X., Zhang Y., Sun X., Ahammed G.J., Jiang C., Meng S.,

Liu Y., Sun Z., Qi M., Li T., Wang F. Revisiting the role of light signaling in plant responses to salt stress. *Hortic. Res*. 2024;uhae262. doi 10.1093/hr/uhae262

Qi C., Zhang H., Liu Y., Wang X., Dong D., Yuan X., Li X., Zhang X., Li X., Zhang N., Guo Y.-D. CsSNAT positively regulates salt tolerance and growth of cucumber by promoting melatonin biosynthesis. *Environ. Exp. Bot*. 2020;175:104036. doi 10.1016/j.envexpbot.2020.104036

Rodriguez M.C., Mehta D., Tan M., Uhrig R.G. Quantitative proteome and PTMome analysis of *Arabidopsis thaliana* root responses to persistent osmotic and salinity stress. *Plant Cell Physiol*. 2021;62(6):1012-1029. doi 10.1093/pcp/pcab076

Sakuma Y., Liu Q., Dubouzet J.G., Abe H., Shinozaki K., Yamaguchi-Shinozaki K. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun*. 2002;290(3);998-1009. doi 10.1006/bbrc.2001.6299

Samuel M.A., Mudgil Y., Salt J.N., Delmas F., Ramachandran S., Chilelli A., Goring D.R. Interactions between the S-domain receptor kinases and AtPUB-ARM E3 ubiquitin ligases suggest a conserved signaling pathway in Arabidopsis. *Plant Physiol*. 2008;147(4):2084-2095. doi 10.1104/pp.108.123380

Sanyal S.K., Pandey G.K. ERF111/ABR1: An AP2 domain transcription factor serving as a hub for multiple signaling pathways. *J. Plant Growth Regul*. 2024. doi 10.1007/s00344-023-11225-3

Serraj R., Roy G., Drevon J.J. Salt stress induces a decrease in the oxygen uptake of soybean nodules and in their permeability to oxygen diffusion. *Physiol. Plant*. 1994;91(2):161-168. doi 10.1111/j.1399-3054.1994.tb00414.x

Shamloo-Dashtpagerdi R., Aliakbari M., Lindlöf A., Tahmasebi S. A systems biology study unveils the association between a melatonin biosynthesis gene, O-methyl transferase 1 (OMT1) and wheat (*Triticum aestivum* L.) combined drought and salinity stress tolerance. *Planta*. 2022;255(5):99. doi 10.1007/s00425-022-03885-4

Shang Y., Yang D., Ha Y., Shin H.Y., Nam K.H. Receptor-like protein kinases RPK1 and BAK1 sequentially form complexes with the cytoplasmic kinase OST1 to regulate ABA-induced stomatal closure. *J. Exp. Bot*. 2020;71(4):1491-1502. doi 10.1093/jxb/erz489

Sharma M., Irfan M., Kumar A., Kumar P., Datta A. Recent insights into plant circadian clock response against abiotic stress. *J. Plant Growth Regul*. 2022;41(8):3530-3543. doi 10.1007/s00344-021-10531-y

Szepesi Á., Csiszár J., Gémes K., Horváth E., Horváth F., Simon M.L., Tari I. Salicylic acid improves acclimation to salt stress by stimulating abscisic aldehyde oxidase activity and abscisic acid accumulation, and increases Na^+ content in leaves without toxicity symptoms in *Solanum lycopersicum* L. *J. Plant Physiol*. 2009;166(9):914-925. doi 10.1016/j.jplph.2008.11.012

Tan S., Sha Y., Sun L., Li Z. Abiotic stress-induced leaf senescence: regulatory mechanisms and application. *Int. J. Mol. Sci.* 2023;24(15):11996. doi 10.3390/ijms241511996

Wang L.Y., Li J., Gong B., Wang R.H., Chen Y.L., Yin J., Yang C., Lin J.-T., Liu H.-Z., Yang Y., Li J., Li C., Yao N. Orosomucoid proteins limit endoplasmic reticulum stress in plants. *New Phytol*. 2023;240(3):1134-1148. doi 10.1111/nph.19200

Wang M., Xu Q., Yuan M. The unfolded protein response induced by salt stress in *Arabidopsis*. *Methods Enzymol*. 2011;489:319-328. doi 10.1016/B978-0-12-385116-1.00018-2

Wang Z., Mao Y., Liang L., Pedro G.C., Zhi L., Li P., Hu X. HFR1 antagonizes ABI4 to coordinate cytosolic redox status for seed germination under high-temperature stress. *Physiol. Plant*. 2024;176(4):e14490. doi 10.1111/ppl.14490

Warpeha K.M., Upadhyay S., Yeh J., Adamiak J., Hawkins S.I., Lapik Y.R., Anderson M.B., Kaufman L.S. The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in Arabidopsis. *Plant Physiol*. 2007;143(4):1590-1600. doi 10.1104/pp.106.089904

Wawrzynska A., Christiansen K.M., Lan Y., Rodibaugh N.L., Innes R.W. Powdery mildew resistance conferred by loss of the ENHANCED DISEASE RESISTANCE1 protein kinase is suppressed by a missense mutation in *KEEP ON GOING*, a regulator of abscisic acid signaling. *Plant Physiol*. 2008;148(3):1510-1522. doi 10.1104/pp.108.127605

Wu T., Goh H., Azodi C.B., Krishnamoorthi S., Liu M.J., Urano D. Evolutionarily conserved hierarchical gene regulatory networks for plant salt stress response. *Nat. Plants*. 2021a;7(6):787-799. doi 10.1038/s41477-021-00929-7

Wu T., Hu E., Xu S., Chen M., Guo P., Dai Z., Feng T., Zhou L., Tang W., Zhan L.I., Fu X. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* (*Camb*). 2021b;2(3):100141. doi 10.1016/j.xinn.2021.100141

Yu G., Wang L., Han Y., He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287. doi 10.1089/omi.2011.0118

Yuan K., Rashotte A.M., Wysocka-Diller J.W. ABA and GA signaling pathways interact and regulate seed germination and seedling development under salt stress. *Acta Physiol. Plant*. 2011;33:261-271. doi 10.1007/s11738-010-0542-6

Zhang H., Liu D., Yang B., Liu W.Z., Mu B., Song H., Chen B., Li Y., Ren D., Deng H., Jiang Y.Q. Arabidopsis CPK6 positively regulates ABA signaling and drought tolerance through phosphorylating ABA-responsive element-binding factors. *J. Exp. Bot*. 2020;71(1):188-203. doi 10.1093/jxb/erz432

Zhang L., Li Q., Shen J., Xue J., Han Y. Transcriptional regulatory networks in response to salt and drought stress in *Arabidopsis thaliana*. *J. Med. Plants Res*. 2012;6(6):950-958. doi 0.5897/JMPR11.240

Zhang L., Zhang X., Fan S. Meta-analysis of salt-related gene expression profiles identifies common signatures of salt stress responses in *Arabidopsis*. *Plant Syst. Evol*. 2017;303:757-774. doi 10.1007/s00606-017-1407-x

Zhao G., Cheng Q., Zhao Y., Wu F., Mu B., Gao J., Yang L., Yan J., Zhang H., Cui X., Chen Q., Lu F., Ao Q., Amdouni A., Jiang Y.-Q., Yang B. The abscisic acid-responsive element binding factors MAPKKK18 module regulates abscisic acid-induced leaf senescence in *Arabidopsis*. *J. Biol. Chem*. 2023;299(4):103060. doi 10.1016/j.jbc.2023.103060

Zhou M., Zhang J., Shen J., Zhou H., Zhao D., Gotor C., Romero L.C., Fu L., Li Z., Yang J., Shen W., Yuan X., Xie Y. Hydrogen sulfide-linked persulfidation of ABI4 controls ABA responses through the transactivation of MAPKKK18 in *Arabidopsis*. *Mol. Plant*. 2021;14(6):921-936. doi 10.1016/j.molp.2021.03.007

Zhou X., Hao H., Zhang Y., Bai Y., Zhu W., Qin Y., Yuan F., Zhao F., Wang M., Hu J., Xu H., Guo A., Zhao H., Zhao Y., Cao C., Yang Y., Schumaker K.S., Guo Y., Xie C.G. SOS2-LIKE PROTEIN KINASE5, an SNF1-RELATED PROTEIN KINASE3-type protein kinase, is important for abscisic acid responses in Arabidopsis through phosphorylation of ABSCISIC ACID-INSENSITIVE5. *Plant Physiol*. 2015;168(2):659-676. doi 10.1104/pp.114.255455

Zhu S.Y., Yu X.C., Wang X.J., Zhao R., Li Y., Fan R.C., Shang Y., Du S.Y., Wang X.F., Wu F.Q., Xu Y.H., Zhang X.Y., Zhang D.P. Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in *Arabidopsis*. *Plant Cell*. 2007;19(10):3019-3036. doi 10.1105/tpc.107.050666

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 17.10.2024. После доработки 20.11.2024. Принята к публикации 20.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):960-973

doi 10.18699/vjgb-24-103

Поиск перспективных генетических маркеров, ассоциированных с молекулярными механизмами снижения устойчивости риса к *Rhizoctonia solani* при избытке азотных удобрений, методом реконструкции и анализа генных сетей

Е.А. Антропова [1, 2] ✉, А.Р. Волянская [1, 2], А.В. Адамовская [1, 2], П.С. Деменков [1, 2, 3, 4], И.В. Яцык [1, 2, 4], Т.В. Иванисенко [1, 2, 3, 4], Ю.Л. Орлов [1, 3, 5, 6], Х. Чао [7], М. Чэнь [7], В.А. Иванисенко [1, 2, 3, 4]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Исследовательский центр в сфере искусственного интеллекта Новосибирского национального исследовательского государственного университета, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[4] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
[5] Аграрно-технологический институт Российского университета дружбы народов им. Патриса Лумумбы, Москва, Россия
[6] Центр цифровой медицины, Первый Московский государственный медицинский университет им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия
[7] Отдел биоинформатики, Колледж естественных наук, Чжэцзянский университет, Ханчжоу, Китай
✉ nzhenia@bionet.nsc.ru

Аннотация. Азотные удобрения, повышающие урожайность риса, при избытке могут снижать устойчивость растений к заболеваниям, в частности к ризоктониозу, вызываемому *Rhizoctonia solani*. Этот патоген способен уничтожить до 50 % урожая, однако механизмы, лежащие в основе снижения устойчивости при избытке азота, остаются малоизученными. Данное исследование направлено на выявление потенциальных генов-маркеров для повышения устойчивости риса к *R. solani* в условиях избытка азота. Применен комплексный биоинформатический подход, включающий анализ дифференциальной экспрессии генов, реконструкцию генных сетей, анализ перепредставленности биологических процессов, филостратиграфический анализ и анализ коэкспрессии некодирующих РНК. Использованы когнитивная система Smart crop, ANDSystem, база данных ncPlantDB и другие биоинформатические ресурсы. Анализ молекулярно-генетической сети взаимодействий выявил три потенциальных механизма, объясняющих снижение устойчивости риса к *R. solani* при избытке азота: OsGSK2-опосредованный путь, путь OsMYB44-OsWRKY6-OsPR1 и путь SOG1-Rad51-PR1/PR2. Идентифицированы потенциальные маркеры для селекции: 7 генов, контролирующих ответы риса на широкий круг стрессов, и 11 генов-модуляторов иммунной системы. Особое внимание уделено ключевым участникам регуляторных путей в условиях избытка азота. Анализ некодирующих РНК выявил 30 микроРНК, мишенями которых являются гены из реконструированной генной сети. Для двух микроРНК (Osa-miR396 и Osa-miR7695) обнаружено около 7400 тыс. уникальных длинных некодирующих РНК (днРНК) с различными индексами коэкспрессии. Выделены топ-50 днРНК с наибольшим индексом коэкспрессии для каждой микроРНК, что открывает новые перспективы в изучении регуляторных механизмов устойчивости риса к патогенам. Полученные результаты создают теоретическую основу для экспериментальных работ по созданию новых сортов риса с повышенной устойчивостью к патогенам в условиях избыточного азотного питания.

Ключевые слова: *Oryza sativa*; *Rhizoctonia solani*; биоинформатика растений; дифференциально экспрессируемые гены; генетическая регуляция; ассоциативные генные сети; база знаний Smart crop; программно-информационная система ANDSystem; азотные удобрения; ответ на грибную инфекцию.

Для цитирования: Антропова Е.А., Волянская А.Р., Адамовская А.В., Деменков П.С., Яцык И.В., Иванисенко Т.В., Орлов Ю.Л., Чао Х., Чэнь М., Иванисенко В.А. Поиск перспективных генетических маркеров, ассоциированных с молекулярными механизмами снижения устойчивости риса к *Rhizoctonia solani* при избытке азотных удобрений, методом реконструкции и анализа генных сетей. *Вавиловский журнал генетики и селекции*. 2024;28(8):960-973. doi 10.18699/vjgb-24-103

Финансирование. Работа ЕАА, АРВ, АВА, ПСД, ИВЯ, ТВИ, ЮЛО и ВАИ поддержана российско-китайским грантом Российского научного фонда № 23-44-00030. Работа ХЧ и МЧ поддержана Национальным фондом естественных наук Китая (№ 32261133526).

Computational identification of promising genetic markers associated with molecular mechanisms of reduced rice resistance to *Rhizoctonia solani* under excess nitrogen fertilization using gene network reconstruction and analysis methods

E.A. Antropova [1, 2] ✉, A.R. Volyanskaya [1, 2], A.V. Adamovskaya [1, 2], P.S. Demenkov [1, 2, 3, 4], I.V. Yatsyk [1, 2, 4], T.V. Ivanisenko [1, 2, 3, 4], Y.L. Orlov [1, 3, 5, 6], Ch. Haoyu [7], M. Chen [7], V.A. Ivanisenko [1, 2, 3, 4]

© Антропова Е.А., Волянская А.Р., Адамовская А.В., Деменков П.С., Яцык И.В., Иванисенко Т.В., Орлов Ю.Л., Чао Х., Чэнь М., Иванисенко В.А., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Artificial Intelligence Research Center, Novosibirsk State University, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
[4] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
[5] Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia
[6] Digital Health Center, I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenovskiy University), Moscow, Russia
[7] Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China
nzhenia@bionet.nsc.ru

Abstract. Although nitrogen fertilizers increase rice yield, their excess can impair plant resistance to diseases, particularly sheath blight caused by *Rhizoctonia solani*. This pathogen can destroy up to 50 % of the crop, but the mechanisms underlying reduced resistance under excess nitrogen remain poorly understood. This study aims to identify potential marker genes to enhance rice resistance to *R. solani* under excess nitrogen conditions. A comprehensive bioinformatics approach was applied, including differential gene expression analysis, gene network reconstruction, biological process overrepresentation analysis, phylostratigraphic analysis, and non-coding RNA co-expression analysis. The Smart crop cognitive system, ANDSystem, the ncPlantDB database, and other bioinformatics resources were used. Analysis of the molecular genetic interaction network revealed three potential mechanisms explaining reduced resistance of rice to *R. solani* under excess nitrogen: the OsGSK2-mediated pathway, the OsMYB44-OsWRKY6-OsPR1 pathway, and the SOG1-Rad51-PR1/PR2 pathway. Potential markers for breeding were identified: 7 genes controlling rice responses to various stresses and 11 genes modulating the immune system. Special attention was given to key participants in regulatory pathways under excess nitrogen conditions. Non-coding RNA analysis revealed 30 miRNAs targeting genes of the reconstructed gene network. For two miRNAs (Osa-miR396 and Osa-miR7695), about 7,400 unique long non-coding RNAs (lncRNAs) with various co-expression indices were found. The top 50 lncRNAs with the highest co-expression index for each miRNA were highlighted, opening new perspectives for studying regulatory mechanisms of rice resistance to pathogens. The results provide a theoretical basis for experimental work on creating new rice varieties with increased pathogen resistance under excessive nitrogen nutrition. This study opens prospects for developing innovative strategies in rice breeding aimed at optimizing the balance between yield and disease resistance in modern agrotechnical conditions.

Key words: *Oryza sativa*; *Rhizoctonia solani*; plant bioinformatics; differentially expressed genes; genetic regulation; associative gene networks; Smart crop knowledge base; ANDSystem software and information system; nitrogen fertilizer; fungal response.

For citation: Antropova E.A., Volyanskaya A.R., Adamovskaya A.V., Demenkov P.S., Yatsyk I.V., Ivanisenko T.V., Orlov Y.L., Haoyu Ch., Chen M., Ivanisenko V.A. Computational identification of promising genetic markers associated with molecular mechanisms of reduced rice resistance to *Rhizoctonia solani* under excess nitrogen fertilization using gene network reconstruction and analysis methods. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):960-973. doi 10.18699/vjgb-24-103

Введение

Рис (*Oryza sativa* L.) – одна из самых хозяйственно ценных культур в мире, составляющая основную часть рациона питания около половины мирового населения. Азотные удобрения находят широкое применение при выращивании риса. На их долю приходится около 80–90 % прибавки урожая, получаемой от минеральных удобрений (Кумейко и др., 2013). Однако, помимо положительного эффекта, азотные удобрения ухудшают устойчивость риса к заболеваниям. Избыток азотных удобрений является одним из основных факторов, способствующих развитию у риса такого заболевания как ризоктониоз, вызываемого грибком *Rhizoctonia solani* Kühn. Ризоктониоз наносит серьезный ущерб этой сельскохозяйственной культуре, приводя к потерям урожая до 50 % (Senapati et al., 2022).

Чувствительность растений к патогенным инфекциям на фоне избытка азотных удобрений обусловлена комплексом факторов, связанных как с быстрым ростом и развитием, так и с изменениями защитных реакций растений. Избыток азота приводит к ряду физиологических изменений, которые могут повышать восприимчивость растений к патогенам. В частности, ускоренный рост может вызывать ослабление клеточных структур, включая снижение прочности клеточных стенок и уменьшение толщины кутикулы, что облегчает проникновение патогенов (Hückelhoven, 2007; Rose et al., 2018). Кроме того, избыточное азотное питание способно вызывать изменения в микробиоме растений и стимулировать рост патогенных микроорганизмов в ризосфере (Xiong et al., 2021).

На молекулярно-генетическом уровне ключевую роль в формировании устойчивости к патогенам играют сложные регуляторные сети, включающие фитогормоны, транскрипционные факторы и некодирующие РНК. Эти компоненты участвуют в комплексных механизмах ответа на стрессы, затрагивающих иммунные процессы растений.

Фитогормоны, такие как салициловая кислота, брассиностероиды, жасмоновая кислота, гиббереллины, абсцизовая кислота, ауксины и этилен, имеют особое значение в ответе на патогенные инфекции (Yang J. et al., 2019). Примечательно, что некоторые из этих фитогормонов, в частности салициловая и абсцизовая кислоты, также вовлечены в метаболизм азотных соединений, регулируя экспрессию генов, связанных с азотным обменом (Xing et al., 2023). Это наблюдение позволяет предположить, что интерференция сигнальных путей фитогормонов может служить механизмом, посредством которого избыток азота влияет на устойчивость растений к патогенам.

Некодирующие РНК (нкРНК) представляют собой разнообразную группу молекул РНК, которые не транслируются в белки, но выполняют важные регуляторные функции в клетке. Среди них выделяют несколько основных типов: микроРНК, малые интерферирующие РНК (миРНК), Piwi-взаимодействующие РНК (пиРНК), рибосомные РНК (рРНК), транспортные РНК (тРНК) и длинные некодирую-

щие РНК (днРНК). днРНК представляют особый интерес, так как они играют значительную роль в регуляции генов, влияя на стабильность и трансляцию мРНК и участвуя в сигнальных путях. В частности, в работе Supriya с сотрудниками (2024) показано, что днРНК принимают участие в ответе риса на инфекцию *R. solani*.

Несмотря на их важность, днРНК остаются наименее изученными среди некодирующих РНК (Statello et al., 2021). Это связано с их разнообразием, сложностью функций и механизмов действия, а также с техническими трудностями их идентификации и характеризации. Одним из подходов для изучения функциональной роли нкРНК является анализ их коэкспрессии с кодирующими белки генами, а также с другими типами нкРНК, для которых установлена их функция. Наиболее полным ресурсом по коэкспрессии нкРНК, включая днРНК риса, является база данных ncPlantDB (https://bis.zju.edu.cn/ncPlantDB/).

Изучение взаимодействий между такими регуляторными элементами как фитогормоны, транскрипционные факторы и нкРНК представляет собой перспективное направление исследований в контексте азотного обмена и устойчивости к патогенам. Оно может привести к более глубокому пониманию механизмов, лежащих в основе индуцированной азотом восприимчивости растений к заболеваниям, и потенциально открыть новые пути для повышения устойчивости сельскохозяйственных культур в условиях интенсивного азотного питания.

В области компьютерной системной биологии метод генных сетей является широко используемым подходом к изучению сложных молекулярно-генетических процессов (Колчанов и др., 2013). Для автоматической реконструкции генных сетей в ИЦиГ СО РАН была разработана когнитивная система ANDSystem, которая с использованием методов искусственного интеллекта осуществляет экстракцию знаний из баз данных и текстов научных публикаций (Ivanisenko et al., 2015, 2019). ANDSystem успешно применялась для реконструкции ассоциативных генных сетей и интерпретации геномных, протеомных и метаболомных данных в различных областях биомедицины и агробиологии. С использованием данной программно-информационной системы были реконструированы важные молекулярно-генетические механизмы различных патологических процессов и биологических явлений, включая астму (Bragina et al., 2014; Saik et al., 2018; Zolotareva et al., 2019), лимфедему (Saik et al., 2019), туберкулез (Bragina et al., 2016), гепатит С (Saik et al., 2016), коронавирусную инфекцию (Ivanisenko et al., 2022), болезнь Хантингтона (Bragina et al., 2023), глиому (Rogachev et al., 2021), послеоперационный делирий (Иванисенко В.А. и др., 2023) и др.

В области биологии растений ANDSystem позволила получить новые данные о молекулярных механизмах функционирования клеточной стенки листьев *Arabidopsis thaliana* L. при ответе на засуху (Volyanskaya et al., 2023). Настройка методов извлечения знаний ANDSystem на область биологии картофеля привела к созданию специализированной базы знаний SOLANUM TUBEROSUM, содержащей информацию о генетической регуляции метаболических путей данной культуры (Иванисенко Т.В. и др., 2018), с использованием которой была проведена приоритизация генов картофеля, вовлеченных в формирование агрономически ценных признаков (Деменков и др., 2019).

Цель настоящей работы заключалась в комплексном биоинформатическом анализе молекулярных механизмов ответа риса на *R. solani* в условиях избытка азота. Исследование включало реконструкцию генных сетей при помощи базы знаний Smart crop – специализированной версии ANDSystem, настроенной на предметную область биологии риса, а также применение биоинформатических методов анализа перепредставленности биологических процессов, филостратиграфического анализа эволюционного возраста генов и анализа коэкспрессии нкРНК.

Материалы и методы

Исследование состояло из нескольких последовательных этапов (рис. 1). На первом этапе на основе анализа транскриптомных данных были идентифицированы гены, дифференциально экспрессирующиеся при инфицировании *R. solani*, а также гены, дифференциальная экспрессия которых наблюдалась в условиях избытка азота. Второй этап включал реконструкцию регуляторных генных сетей с участием идентифицированных генов. На третьем этапе был проведен структурно-функциональный анализ полученных сетей, включающий оценку показателей центральности узлов, анализ обогащения биологических процессов и определение эволюционного возраста генов. Далее был выполнен анализ регуляции трансляции генов сети микроРНК и исследована коэкспрессия днРНК. Заключительный этап был направлен на выявление потенциальных маркеров устойчивости к *R. solani* в условиях избытка азота.

Общедоступные данные по экспрессии генов. Публично доступные транскриптомные данные о реакции *O. sativa* на избыток азотных удобрений, а также на патоген *R. solani*, были собраны из баз данных NCBI GEO (Gene Expression Omnibus) и NCBI SRA (Sequence Read Archive) (https://www.ncbi.nlm.nih.gov/sra) (табл. 1). Для анализа транскриптома *O. sativa* в условиях избытка азота было найдено одно исследование, включающее в себя три эксперимента. В данной работе растения обрабатывались избытком удобрения – нитратом аммония (NH_4NO_3) в концентрациях, превышающих нормальную в 4, 16 и 64 раза.

В анализ дифференциальной экспрессии генов *O. sativa* при инфицировании *R. solani* были включены данные пяти исследований с временными сериями, содержащие в общей сложности 21 эксперимент.

Анализ транскриптомных данных. Для извлечения файлов в формате FASTQ был использован пакет SRA Toolkit (v3.1.0). Контроль качества прочтений был произведен с помощью FastQC (v0.12.0). Фильтрация и удаление нуклеотидов низкого качества были проведены в программе Trimmomatic (https://github.com/usadellab/Trimmomatic). В качестве порога использовали длину прочтений в 15 п.н. и показатель качества последовательности Phred < Q20. Прочтения были картированы на референсный геном *O. sativa* Japonica Group (IRGSP-1.0), депонированный из базы данных EnsemblPlants (https://plants.ensembl.org/index.html) при помощи инструмента

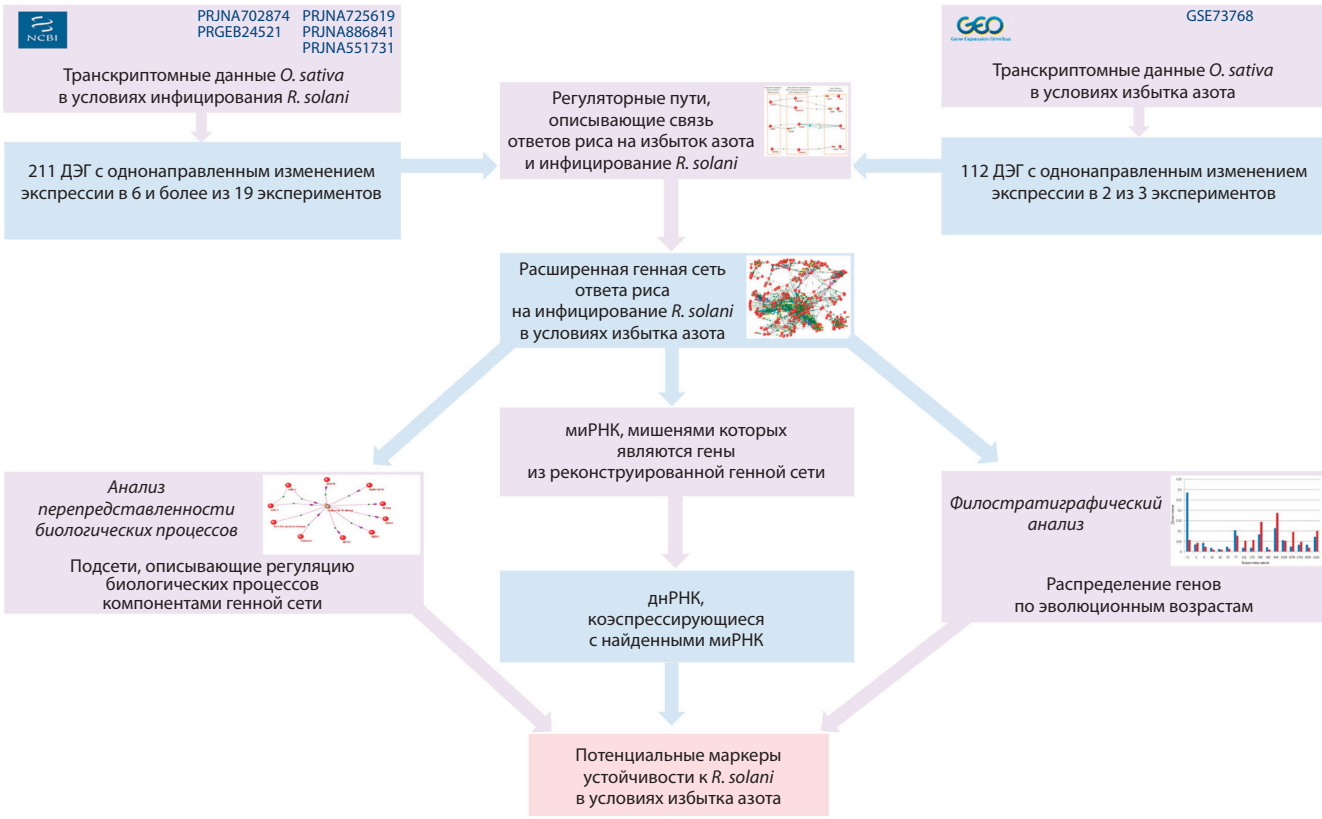


Рис. 1. Схема этапов исследования.

Таблица 1. Перечень использованных в работе публично доступных данных RNA-seq и ДНК-микрочипов

Стресс	Дизайн	ID проекта	Подсемейство	Генотип	Орган	Тип данных
Избыток азота (NH_4NO_3)	3 эксперимента: 3 концентрации (x64, x16, и x4 от нормальной концентрации)	GSE73768	ssp. *japonica*	cv. Nipponbare	Побег	микрочип
Инфицирование *R. solani*	3 эксперимента: 3 временные точки (1, 2 и 5 дней после инфекции)	PRJNA725619	ssp. *indica*	var. BPT- 5204	Лист	RNA-seq
Инфицирование *R. solani*	4 эксперимента: 2 временные точки для 2 сортов (1 и 2 дня после инфекции)	PRJNA886841	ssp. *japonica*	var. Lemont, var. GD66	Лист	RNA-seq
Инфицирование *R. solani*	2 эксперимента: 1 временная точка для 2 сортов (3 дня после инфекции)	PRJNA551731	ssp. *japonica*	cv. Yanhui-888, cv. Jingang-30	Лист	RNA-seq
Инфицирование *R. solani*	6 экспериментов: 3 временные точки для 2 сортов (1, 3 и 5 дней после инфекции)	PRJEB24521	ssp. *japonica*	var. Cocodrie, line MCR	Лист	RNA-seq
Инфицирование *R. solani*	6 экспериментов: 3 временные точки для 2 сортов (1, 2 и 3 дня после инфекции)	PRJNA702874	ssp. *indica*	Line PAU-ShB8, line PR114	Лист	RNA-seq

HISAT2 (v2.2.1). Для конвертации выходных файлов картирования формата SAM в бинарный формат BAM использовали SAMtools (v1.20). Для квантификации был использован HTSeq (v2.0.2). Нормализация количества прочтений и анализ дифференциальной экспрессии генов осуществлялся с помощью инструмента edgeR (4.0.16), реализованного в проекте Bioconductor (https://www.bioconductor.org/). Для нормализации был использован метод TMM (Trimmed Mean of M-values). Поправка на множественное тестирование была введена с использованием FDR (false discovery rate).

Данные ДНК-микрочипов аналиировали с помощью пакета limma (v3.58.1) из проекта Bioconductor. Чтение сырых файлов ДНК-микрочипа платформы Agilent было выполнено с использованием read.images. Далее была проведена коррекция фонового шума и квантильная

№	Схема шаблона
Ш1	$Ген_N$ —*экспр*→ $Белок_N$ —*рег.акт., асс., рег.тран., PPi, расщ., рег.дегр.*→ $Белок_{гр}$
Ш2	$Ген_N$ —*экспр*→ $Белок_N$ —*рег.эксп., асс., PPi*→ $Ген_{гр}$
Ш3	$Ген_N$ —*экспр*→ $Белок_N$ —*рег.акт., асс., рег.тран., PPi, расщ., рег.дегр.*→ $Белок_п$ —*рег.акт., асс., рег.тран., PPi, расщ., рег.дегр.*→ $Белок_{гр}$
Ш4	$Ген_N$ —*экспр*→ $Белок_N$ —*рег.акт., асс., рег.тран., PPi, расщ., рег.дегр.*→ $Белок_п$ —*рег.эксп., асс., PPi*→ $Ген_{гр}$
Ш5	$Ген_N$ —*экспр*→ $Белок_N$ —*рег.эксп., асс., PPi*→ $Ген_п$ —*экспр*→ $Белок_п$ —*рег.акт., асс., рег.тран., PPi, расщ., рег.дегр.*→ $Белок_{гр}$
Ш6	$Ген_N$ —*экспр*→ $Белок_N$ —*рег.эксп., асс., PPi*→ $Ген_п$ —*экспр*→ $Белок_п$ —*рег.эксп., асс., PPi*→ $Ген_{гр}$

Рис. 2. Схема шаблонов, использованных для поиска молекулярно-генетических путей в базе знаний Smart crop.

Обозначения: Ш – шаблон; $Ген_N$ – ДЭГ риса при избытке азотных удобрений; $Белок_N$ – белковые продукты ДЭГ при избытке азотных удобрений; $Ген_п$ – гены, кодирующие белки-посредники; $Белок_п$ – белки-посредники; $Ген_{гр}$ – ДЭГ риса при ответе на инфекцию *R. solani*; $Белок_{гр}$ – белковые продукты ДЭГ риса при ответе на инфекцию *R. solani*.

нормализация данных. Пакет biomaRt (v2.58.2) (https://bioconductor.org/packages/release/bioc/html/biomaRt.html) был применен для сопоставления идентификаторов зондов ДНК-микрочипа с идентификаторами генов *Ensembl*. Анализ дифференциальной экспрессии генов был осуществлен с использованием пакета limma. Для идентификации дифференциально экспрессирующихся генов был использован порог FDR < 0.05.

База знаний Smart crop. В данном исследовании была использована специализированная база знаний Smart crop, представляющая собой адаптированную версию программно-информационной системы ANDSystem, ориентированной на область генетики и селекции риса и пшеницы. Адаптация системы включала настройку трех ключевых модулей ANDSystem для эффективного решения поставленных задач. Первым стал модуль предметно-ориентированной онтологии, который был расширен специальными словарями. Эти словари охватывали широкий спектр объектов исследования, которые можно разбить на группу молекулярно-генетических объектов (гены, белки, метаболиты, нкРНК и микроРНК), их функциональные характеристики (биологические процессы, генетические биомаркеры, QTL-полиморфизмы), фенотипические характеристики (сорта растений, селекционно значимые качества, фенотипические признаки, болезни), биотические и абиотические факторы (патогены, вредители и др.). Для формирования этих словарей были использованы различные базы данных и онтологии, такие как NCBI Gene (https://www.ncbi.nlm.nih.gov/gene/), ChEBI (https://www.ebi.ac.uk/chebi/), MirBase (https://www.mirbase.org/), Gene Ontology (https://cropontology.org/), Wheat Ontology, Rice Ontology и другие (Chao et al., 2023). Так, например, словарь генов из группы молекулярно-генетических объектов содержит наименования около 627 тысяч генов, включая их общепринятые названия и синонимы. Биологические процессы, относящиеся к функциональным характеристикам, содержат более 122 тысяч наименований. Словарь патогенов, входящих в группу биотических факторов, содержит около 755 наименований.

Вторым важным компонентом стал модуль экстракции информации из фактографических баз данных, который был настроен на автоматизированное извлечение данных из специализированных источников в области биологии растений. К таким источникам относились Oryzabase, (https://shigen.nig.ac.jp/rice/oryzabase/), GrainGenes (https://wheat.pw.usda.gov/GG3/), ASPNet и др. Третьим выступил модуль интеллектуального анализа текстов (text mining) с использованием семантико-лингвистических шаблонов и методов искусственного интеллекта. Он был адаптирован для эффективного извлечения знаний из текстовых источников, таких как научные статьи и патенты в области биологии растений. На основе анализа научных публикаций, выполненного с использованием данного модуля, было экстрагировано более 4 млн взаимодействий между объектами, представленными в словарях.

Реконструкцию генных сетей производили с помощью «Мастера запросов» и «Мастера путей» программного модуля ANDVisio (Demenkov et al., 2011), представляющего собой пользовательский интерфейс в системах ANDSystem и Smart crop. Структура шаблонов, использованных при поиске регуляторных путей в базе знаний Smart Crop с помощью «Мастера путей», приведена на рис. 2.

Центральность узлов в генной сети оценивалась с помощью показателя сетевой связности (Network connectivity), определяемого как число связей данного узла с другими узлами сети.

Анализ перепредставленности биологических процессов Онтологии гена был выполнен с помощью ресурса PANTHER (https://pantherdb.org/).

Анализ длинных некодирующих РНК. Анализ коэкспрессии между микроРНК и днРНК проводился с использованием базы данных ncPlantDB (https://bis.zju.edu.cn/ncPlantDB/).

Филостратиграфический анализ. Эволюционный возраст генов был определен с использованием базы данных GenOrigin (http://chenzxlab.hzau.edu.cn/) (Tong et al., 2021), которая содержит информацию об эволюционном возрасте генов различных организмов, установленном методом филостратиграфического анализа. Для оценки статистической значимости различий в распределении генов разного возраста между полным набором белок-кодирующих генов риса и генами реконструированной сети был применен гипергеометрический тест. Вероятность наблю-

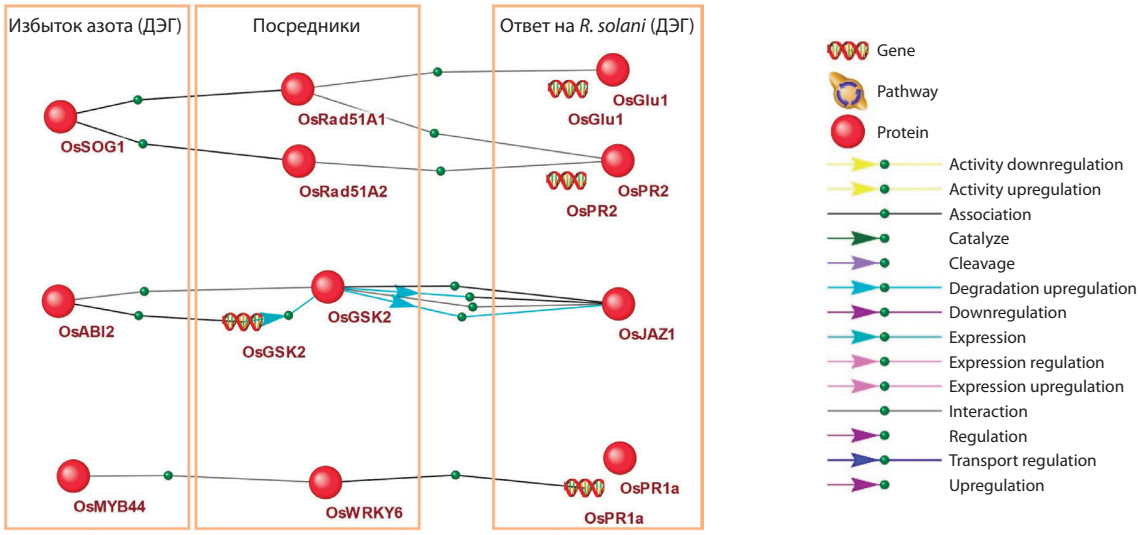


Рис. 3. Регуляторные пути, описывающие связь между ДЭГ при ответе риса на избыток азота и инфицирование *R. solani*.

дения m и более генов определенного возрастного интервала среди M генов сети рассчитывалась с помощью функции hypergeom.pmf библиотеки scipy. Анализ проводился для 17 возрастных интервалов, представленных в базе данных GenOrigin. При расчетах использовались следующие параметры: N – общее количество белок-кодирующих генов риса, n – число генов риса в заданном возрастном интервале, M – число генов в генной сети, m – количество генов сети в анализируемом возрастном интервале. Различия считались статистически значимыми при p-value < 0.05.

Результаты и обсуждение

Выявление устойчивых дифференциально экспрессирующихся генов

Для поиска дифференциально экспрессирующихся генов (ДЭГ) риса в условиях избытка азота было проанализировано 3 эксперимента, а в условиях влияния грибка *R. solani* – 21 эксперимент, транскриптомные данные которых удалось найти в открытых источниках. Мы рассмотрели гены с однонаправленным изменением экспрессии в разных экспериментах (одновременное понижение или повышение), которые в дальнейшем будем называть устойчивыми ДЭГ.

В случае избытка азота было обнаружено только 5 генов, которые являлись устойчивыми ДЭГ во всех трех экспериментах (*Os09g0538000*, *Os05g0162000*, *Os09g0537700*, *Os04g0664900*, *Os06g0113800*). При условии встречаемости ДЭГ в двух из трех экспериментов число таких генов составило 112. Они и были взяты для дальнейшего анализа.

Анализ дифференциальной экспрессии генов в условиях инфицирования *R. solani* показал, что в двух экспериментах из 21 статистически достоверные ДЭГ не были выявлены. При анализе оставшихся 19 экспериментов не было выявлено ни одного гена, который являлся бы ДЭГ в каждом эксперименте. Оказалось, что только 2 гена являлись устойчивыми ДЭГ в половине и более экспериментов (*Os04g0180500* и *Os09g0255600*). При рассмотрении трети экспериментов (6 и более из 19) в число устойчивых ДЭГ вошло 211 генов. Число устойчивых ДЭГ для четверти экспериментов (5 и более из 19) составило 463 гена. Для дальнейшего анализа мы выбрали значение порога для определения устойчивых ДЭГ, равное одной трети экспериментов (6 и более из 19), поскольку при этом значении выборки устойчивых ДЭГ в условиях избытка азота и влияния грибка оказались сопоставимы по размеру.

Реконструкция молекулярно-генетических путей, описывающих взаимосвязь ответов риса на избыток азота и инфицирование *R. solani*

С помощью программы ANDVisio, которая является пользовательским интерфейсом для баз знаний Smart crop и ANDSystem, был проведен поиск молекулярно-генетических путей в глобальной генной сети Smart crop (см. рис. 2), связывающих группу отобранных 112 устойчивых ДЭГ при ответе на избыток азота и 211 устойчивых ДЭГ при ответе на инфекцию *R. solani*. В результате такого поиска было найдено несколько регуляторных путей, которые включали в себя 3 белка, кодируемых ДЭГ при ответе на избыток азота, 4 ДЭГ и кодируемых ими белков при ответе на инфицирование *R. solani*, а также 4 белка, являющихся посредниками во взаимодействиях между рассмотренными ДЭГ (рис. 3).

Молекулярно-генетический путь OsABI2-OsGSK2-OsJAZ1

Одним из важных реконструированных путей (см. рис. 3), потенциально объясняющих механизм снижения устойчивости риса к грибку при избытке азота, является путь OsABI2-OsGSK2-OsJAZ1. Белок OsABI2 (PP2C06, protein phosphatase 2C6) является продуктом дифференциально экспрессируемого при избытке азота гена *Os01g0583100*. Его экспрессия снижается при 16- и 64-кратном превышении концентрации азотного удобрения (Приложение 1[1]).

[1] Приложения 1–5 см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx34.xlsx

Известно, что ABI2 – один из основных участников сигнального пути абсцизовой кислоты (ABA) (Sun et al., 2011), которая является важным гормоном, регулирующим закрывание устьиц, старение листьев, покой почек, ингибирование прорастания семян, ингибирование роста растений, ответные реакции на засуху, засоление, действие токсичных металлов (Chen et al., 2020; Kumar S. et al., 2022). Было показано, что OsABI2 участвует в ответе риса на избыток железа (Junior et al., 2015), его экспрессия усиливается при засухе у подсолнечника (Shen et al., 2023) и также отмечается при засухе в корнях и стебле риса (Sircar et al., 2022). Представленность этого белка в реконструированном регуляторном пути может свидетельствовать о его участии в модуляции реакции риса на патоген при избытке азота. OsABI2 может оказывать регуляторное влияние на OsJAZ1 (jasmonate ZIM-domain protein 1), важного фактора ответа на патогены, через посредника OsGSK2.

Согласно нашему анализу, OsJAZ1 (*Os10g0392400*) является ДЭГ с повышенным уровнем экспрессии в 7 из 19 экспериментов по влиянию *R. solani* на транскриптом риса (Приложение 2). Было показано, что грибок *Verticillium dahliae*, вызывающий вертициллёзное увядание (так называемый «вилт») у арабидопсиса и хлопка, индуцирует фосфорилирование JAZ1 через GSK2 (glycogen synthase kinase 2, brassinosteroid insensitive 2, BIN2), и это способствует дальнейшей деградации JAZ1 (Song Y. et al., 2021). Авторы отмечают, что в данном случае GSK2 является отрицательным регулятором устойчивости к грибку – его конститутивная экспрессия ослабляла устойчивость, а нокдаун гена *GSK2* повышал устойчивость к *V. dahliae*. Интересно отметить, что *OsGSK2* (*Os05g0207500*) является ДЭГ в двух из 19 проанализированных экспериментах по изучению влияния *R. solani* на транскриптом, в которых его экспрессия была снижена (см. Приложение 2). *OsGSK2* также являлся ДЭГ при ответе на избыток азота в эксперименте с наибольшей концентрацией азотного удобрения, превышающей нормальную концентрацию в 64 раза.

В нашей сети связь между ABI2 и GSK2 имеет тип “interaction” (физическое взаимодействие). У арабидопсиса ABI1 и ABI2 дефосфорилируют белок GSK2, что приводит к подавлению его киназной активности и снижению стабильности. Рассмотренные взаимодействия между участниками регуляторного пути согласуются с данными Wang H. с сотрудниками (2018), в которых показано, что сигнальный путь ABA подавляет сигнальный путь брассиностероидов (BR). BR представляют собой важную группу гормонов растений, в ряде случаев выполняющих антагонистическую роль по отношению к действию ABA. Например, BR стимулируют прорастание семян, в то время как ABA способствует их покою (Steber, McCourt, 2001). ABA действует противоположно BR в регуляции наклона листьев *O. sativa* через ген биосинтеза BR *OsD11* и сигнальные гены *OsGSK2* и *OsDLT* (Li et al., 2019).

Молекулярно-генетический путь MYB44-WRKY6-PR1

Другой важный регуляторный путь начинается с белка OsMYB44 – продукта гена *Os09g0106700*, дифференциально экспрессируемого при избытке азота. Отметим, что он является ДЭГ в двух из трех экспериментов, когда экспрессия гена снижена при 16- и 64-кратном превышении концентрации азотного удобрения (см. Приложение 1). Транскрипционный фактор (ТФ) MYB44 принимает важное участие в регуляции жизни растения (развитии корней, соматическом эмбриогенезе, старении листьев и др.) и ответе на биотические и абиотические стрессы, такие как засуха, холод, недостаток фосфата и азота, заражение патогенными организмами (Wang F. et al., 2024). Интересно отметить разнонаправленное действие MYB44 при защите от инфекций. Он усиливал защитную реакцию на патогенную бактерию *Pseudomonas syringae* pv. *tomato*, индуцированную салициловой кислотой, но снижал зависимый от жасмоновой кислоты защитный ответ против грибка *Alternaria brassicicola*, вызывающего болезнь черной пятнистости (Shim et al., 2013). В рассматриваемом пути MYB44 образует регуляторный комплекс с другим ТФ – WRKY6 (*Os03g0798500*), регулирующим транспорт неорганического фосфата у картофеля (Zhou et al., 2017). ТФ WRKY6, как и MYB44 у *A. thaliana*, действует как положительный регулятор сигнализации ABA. Семейство ТФ WRKY участвует в формировании устойчивости растений к широкому кругу стрессов, в частности OsWRKY6 необходим для защиты риса от *Xanthomonas oryzae* pv. *oryzae*, вызывающего бактериальную полосатость (Im et al., 2024). OsWRKY6 активирует экспрессию *OsPR1* (Im et al., 2022) – конечного звена рассматриваемого регуляторного пути.

Молекулярно-генетический путь SOG1-Rad51-PR1/PR2

Данный путь включает в себя три звена: SOG1 (suppressor of gamma response1), RAD51 (DNA repair protein RAD51) и гены *PR1* (pathogenesis-related protein 1) и *PR2* (pathogenesis-related protein 2) (см. рис. 3). SOG1 – ТФ растений, аналог белка p53 животных, играющий решающую роль в регуляции транскрипции генов, вовлеченных в программируемую клеточную гибель, восстановление повреждений ДНК, а также в ответы на абиотические стрессы и патогенные инфекции (Ogita et al., 2018; Yoshiyama, Kimura, 2018). Согласно нашему анализу транскрипционных данных, *SOG1* (*Os06g0267500*) является ДЭГ при избытке азота. Уровень его экспрессии повышается в двух из трех экспериментов: при 16- и 64-кратном превышении концентрации азотного удобрения (см. Приложение 1).

SOG1 регулирует транскрипцию *OsRad51* (Ogita et al., 2018; Yoshiyama, Kimura, 2018), выступающего в роли посредника в рассматриваемом пути. RAD51 – регуляторный белок иммунного ответа растений. В числе прочих его прямых мишеней есть представители семейства белков *PR1* и *PR2*, связанные с патогенезом (Wang S. et al., 2010). Эти гены оказались в числе ДЭГ ответа на инфекцию *R. solani* (см. Приложение 2).

Экспрессия *PR1* (*Os07g0129200*) повышалась в 6 из 19 экспериментов по влиянию *R. solani* на транскриптом. В геноме риса обнаружено семь генов, названных *PR2* (Yokotani et al., 2014). Согласно нашим данным, экспрессия трех из них (*Os07g0539900*, *Os01g0940700* и *Os01g0940800*) повышалась в 7 из 19 экспериментов.

Следует отметить, что гены *PR1* и *PR2* также оказались в числе ДЭГ по результатам нашего анализа транскриптомных данных серии экспериментов по исследованию

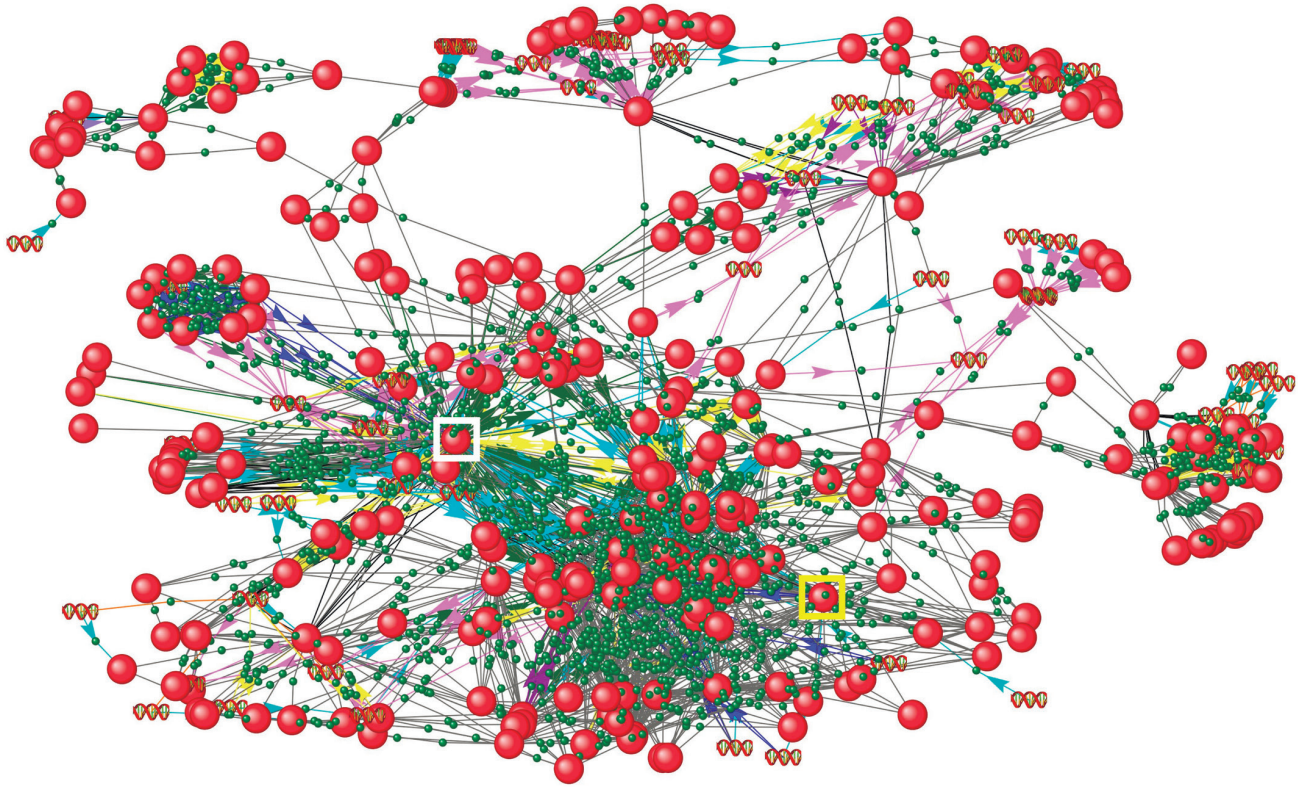


Рис. 4. Расширенная генная сеть ответа риса на инфицирование *R. solani* в условиях избытка азота.

Сеть включает как исходные регуляторные пути, так и вновь идентифицированные компоненты (гены и белки). Желтым и белым квадратами выделены белки JAZ1 и GSK2 соответственно. Обозначения генов, белков и типов их взаимодействий аналогичны представленным на рис. 3.

избытка азота. Их экспрессия статистически значимо изменялась в одном из трех экспериментов, в котором концентрация азотных удобрений была максимальной.

Реконструкция расширенной генной сети ответа риса на инфицирование *R. solani* в условиях избытка азота

Для выявления более широкого круга потенциальных участников механизмов снижения устойчивости риса к грибку *R. solani* при избытке азота мы провели реконструкцию расширенной генной сети, в основу которой были положены рассмотренные выше регуляторные пути. Реконструкция осуществлялась автоматически с помощью функционального модуля программы ANDVisio. Этот инструмент позволяет расширять исходную сеть, добавляя новые компоненты (гены, белки, метаболиты и др.) на основе данных об их взаимодействиях, содержащихся в базе знаний Smart crop. Для 15 участников исходных регуляторных путей (см. рис. 3) в базе знаний содержалась информация об их взаимодействиях с 358 новыми белками и генами. Реконструированная таким образом сеть содержала 61 ген, 271 белок, 2359 взаимодействий (рис. 4). Для выявления ключевых участников данной сети был проведен анализ центральности вершин с использованием индекса Network connectivity, указывающего на количество ближайших соседей. Наибольшим значением индекса обладал белок OsGSK2 – участник исходных регуляторных путей, опосредующий взаимодействия между дифференциально экспрессирующимися генами. В числе первой тройки по величине индекса Network connectivity также оказался Jaz1. Следует отметить, что ген, кодирующий Jaz1, являлся устойчивым ДЭГ при ответе на инфекцию *R. solani*.

Идентификация днРНК, потенциально регулирующих выявленные молекулярно-генетических пути

Для поиска днРНК, потенциально вовлеченных в регуляцию генной сети ответа риса на грибную инфекцию в условиях избытка азота, мы проанализировали базу данных ncPlantDB. Эта база содержит информацию о коэкспрессии днРНК с микроРНК, полученную в результате анализа single-cell данных.

Основываясь на данных базы знаний Smart crop, мы нашли 30 микроРНК, мишенями которых являются гены из реконструированной генной сети (табл. 2). В базе данных ncPlantDB были обнаружены связи для Osa-miR396 и Osa-miR7695 с днРНК с различными индексами степени коэкспрессии. Для двух вариантов Osa-miR396 (Osa-miR396b и Osa-miR396c) число таких нкРНК составило ~4000. Для Osa-miR7695 было выявлено около 3500 коэкспрессионных связей с днРНК. Суммарное число уникальных днРНК составило ~7400.

Среди выявленных днРНК особое внимание стоит уделить тем, которые обладают наибольшим индексом коэкспрессии. К таким днРНК можно отнести топ-50, ранжированных по индексу коэкспрессии, в частности, группу выявленных в метаксилеме риса днРНК, обладающих одинаковым индексом коэкспрессии с Osa-miR396b, генами-мишенями которой являются *GFR1* и *GFR3*: LNC-

Таблица 2. МикроРНК, регулирующие гены стрессовых ответов в реконструированной генной сети

№	МикроРНК	Ген-мишень	Литература
1–3	Osa-miR156	*OsMPKs*, *OsSPL14*	Xie et al., 2006; Kumar K. et al., 2022; Song L. et al., 2021
4	Osa-miR159	*OsGAMYB*	Kumar K. et al., 2022
5	Osa-miR162	*OsDCL1*	Kumar K. et al., 2022
6–8	Osa-miR166	*EIN 2*	Song L. et al., 2021; Kumar K. et al., 2022
9	Osa-miR167	*ARF12*	Kumar K. et al., 2022
10	Osa-miR319	*OsTCP21*	Song L. et al., 2021; Kumar K. et al., 2022
11–12	Osa-miR393	*AFB2/TIR*	Song L. et al., 2021
13–21	Osa-miR396	*OsGRFs*	Song L. et al., 2021
22	Osa-miR398	*SOD*, *CSD1*, *CSD2*	Song L. et al., 2021; Kumar K. et al., 2022
23	Osa-miR408	*OsAAE3*	Charagh et al., 2024
24–29	Osa-miR444	*MADS23/27a/57*	Kumar K. et al., 2022; Song L. et al., 2021
30	Osa-miR7695	*OsNramp6*	Kumar K. et al., 2022; Song L. et al., 2021

Примечание. МикроРНК одного семейства объединены в группы.

Os08g15450, LNC-Os04g61735, LNC-Os05g27975, LNC-Os05g62500 и др. (Приложение 3).

Поиск функций этих днРНК по литературным данным не дал результатов. Поэтому связь днРНК с генной сетью может иметь особое значение для дальнейшей характеризации их функций.

Филостратиграфический анализ

Применение методов филостратиграфического анализа для оценки эволюционного возраста генов является перспективным подходом к исследованию особенностей эволюции генных сетей и идентификации их ключевых компонентов (Мустафин и др., 2021). В настоящей работе этот подход был использован для анализа этапов эволюции, на которых возникли гены, участвующие в реконструированной сети ответа на грибковую инфекцию в условиях повышенных концентраций азотных удобрений.

Анализ распределения эволюционного возраста генов показал, что в реконструированной сети присутствуют гены разного возраста, среди которых можно выделить несколько наиболее представленных групп (рис. 5). Возрастные интервалы, в пределах которых количество генов статистически значимо превышало ожидаемое по случайным причинам, соответствовали следующим временным точкам, представленным на графике (см. рис. 5): 1 – 132 млн лет ($p = 1.85 \cdot 10^{-3}$), 2 – 170 млн лет ($p = 9.16 \cdot 10^{-4}$) и 3 – 1578 млн лет ($p = 5.41 \cdot 10^{-7}$).

Первая группа, включающая 11 генов возрастом около 132 млн лет, вероятно, возникла на эволюционном этапе появления однодольных растений (Friis et al., 2004). К представителям этой группы принадлежит ТФ OFP3 (ovate family protein 3). Специфичное для растений семейство OFP принимает участие в регуляции клеточной плюрипотентности, морфогенеза и роста у *A. thaliana* (Wang F. et al., 2016). Более того, существует предположение, что изменения в сетях регуляции транскрипционных факторов являются существенной чертой эволюции однодольных растений (Vincentz et al., 2004).

В пределах второго рассматриваемого интервала (170 млн лет) оказался возраст 12 генов. К этому периоду относят появление цветковых растений (van der Kooi, Ollerton, 2020). В этот интервал попали члены семейства

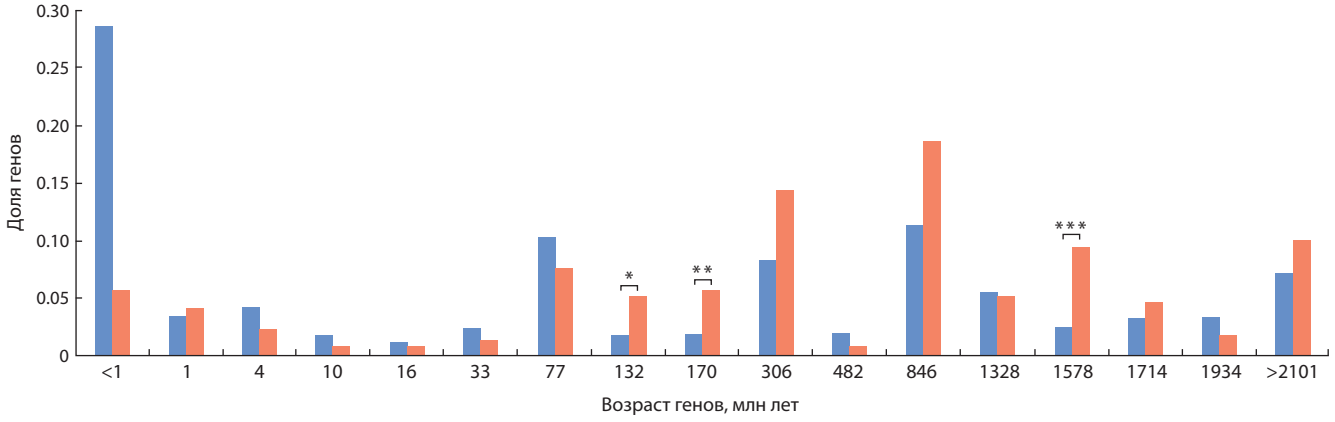


Рис. 5. Распределение эволюционного возраста генов в реконструированной генной сети.

По оси абсцисс отложены центральные точки возрастных интервалов (млн лет) согласно базе данных GenOrigin, по оси ординат – доля генов в каждом возрастном интервале. Синим цветом показано распределение для полного набора белок-кодирующих генов риса, красным – для генов реконструированной сети. Звездочками отмечены возрастные интервалы со статистически значимыми различиями в представленности генов: * $p = 1.85 \cdot 10^{-3}$, ** $p = 9.16 \cdot 10^{-4}$, *** $p = 5.41 \cdot 10^{-7}$, гипергеометрический тест.

Таблица 3. Результаты анализа перепредставленности биологических процессов для генов расширенной сети ответа риса на инфицирование *R. solani* в условиях избытка азота

Термин из Gene Ontology	*p*-value	FDR	Количество генов
Response to hormone	1.09E-38	2.87E-36	44
Hormone-mediated signaling pathway	2.08E-37	4.75E-35	38
Response to chemical	2.15E-30	3.38E-28	47
Response to stress	1.41E-22	1.69E-20	50
Regulation of defense response	4.12E-16	4.41E-14	13
Seed germination	5.17E-10	3.76E-08	6
Response to water deprivation	4.47E-09	2.74E-07	8
Defense response	3.36E-07	1.51E-05	18
Cellular response to abiotic stimulus	3.70E-06	1.46E-04	5
Defense response to fungus	1.81E-03	3.96E-02	4

Примечание. Анализ выполнен с помощью ресурса PANTHER. Представлены наиболее значимые биологические процессы, связанные с ответом на различные биотические и абиотические факторы.

ТФ WRKY (WRKY6, 40 и 46), участвующие в молекулярных механизмах регуляции цветения (Song H. et al., 2024). Важно отметить, что WRKY6 также является участником исходных регуляторных путей.

В третью группу вошли 20 генов, возраст которых оказался в пределах третьего интервала (1578 млн лет), соответствующего появлению красных и зеленых водорослей (Zhang S. et al., 2021). Одним из представителей этой группы является ген *PHT1* (*phosphate transporter1*), продукт которого участвует в поглощении и транспорте неорганического фосфата (Wang X. et al., 2014). Развитие механизмов усвоения фосфора могло иметь значение в эволюции растений, поскольку увеличение доступности фосфата в океанах связывают с ростом более крупных эукариотических организмов (Zhang S. et al., 2021).

Можно отметить еще одну особенность генной сети: доля «молодых» генов, возраст которых меньше 1 млн лет, оказалась меньше, чем их доля в полном геноме. В число «молодых» генов, оказавшихся в этом интервале, попадает 12 генов, многие из которых в той или иной степени связаны с иммунными ответами: *OsPR5* (*OS01G0122000*), *OsNAC6* (*Os01g0672100*), *similar to histone H4* (*OS01G0835900*), *OsMPK3* (*OS02G0148100*), *R2R3-MYB* (*OS02G0641300*), *R2R3-MYB* (*OS06G0205100*), *OsPR1b* (*OS07G0127700*), *histone H4* (*OS07G0549900*), *R2R3MYB-domain protein* (*OS12G0564100*).

Полученные данные могут способствовать более глубокому пониманию механизмов функционирования реконструированной генной сети и послужить основой для дальнейшего выбора маркеров для селекции растений, устойчивых к патогенам в условиях повышенной концентрации азотных удобрений.

Поиск потенциальных мишеней для маркер-ориентированной селекции

Чтобы найти потенциальные мишени для маркер-ориентированной селекции, был проведен анализ функциональной значимости генов на уровне биологических процессов. С помощью ресурса PANTHER была проанализирована перепредставленность терминов Онтологии гена для расширенной генной сети. В результате было выявлено 239 статистически значимых биологических процессов (Приложение 4), включая ключевые сигнальные пути и ответы на абиотические и биотические стрессы, в том числе на грибковые инфекции (табл. 3).

Хотя анализ перепредставленности биологических процессов предоставляет важные сведения о функциональной значимости генной сети, для отбора эффективных маркеров необходимо понимание конкретных регуляторных механизмов. База знаний Smart crop содержит информацию о регуляторных взаимодействиях между генами и биологическими процессами, что позволяет идентифицировать потенциальные маркеры не только по их ассоциации с ключевыми процессами, но и по их регуляторному потенциалу.

Для поиска потенциальных маркеров генная сеть была дополнена регуляторными связями с биологическими процессами с использованием ANDVisio (Приложение 5). Регуляторные связи между генами и процессами были классифицированы как положительные (upregulation), отрицательные (downregulation) или без указания знака направленности (regulation). На рис. 6 представлены регуляторные сети для процессов “response to stress” и “innate immune system”, играющих ключевую роль в механизмах стрессового ответа.

Следует отметить, что “response to stress” оказался перепредставлен среди генов расширенной сети ответа риса на инфицирование *R. solani* в условиях избытка азота (см. табл. 3). Важными регуляторами данного процесса являются три белка (см. рис. 6, *а*): BZR1 (brassinazole resistant 1), серин-треониновая протеинкиназа SAPK4 (Ser/Thr protein kinase см. рис. 6, *а*) и ТФ SOG1 (OsSOG1 см. рис. 6, *а*, *б*). Известно, что BZR1 опосредует передачу сигналов брассиностероидов, подавляя транскрипцию генов стрессового ответа (Yang Y.X. et al., 2015; Cao et al., 2024). SAPK4 регулирует экспрессию генов в ответ на солевой стресс у риса (Diédhiou et al., 2008). SOG1 контролирует ответ растений на стрессы, вызывающие повреждение ДНК (Ogita et al., 2018; Yoshiyama, Kimura, 2018). SOG1 является компонентом исходных регуляторных путей, что позволяет отнести его к особо важным потенциальным маркерам. Все рассмотренные белки можно

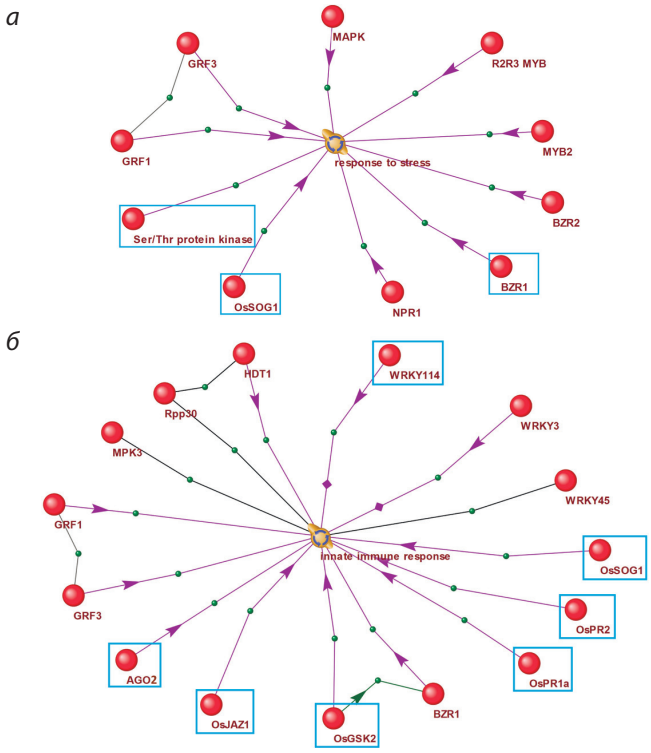


Рис. 6. Регуляция биологических процессов "response to stress" (*а*) и "innate immune response" (*б*) белками, являющимися компонентами генной сети ответа риса на патогенный грибок в условиях избытка азота.

Связи между объектами, отмеченные черными линиями, обозначают ассоциацию; фиолетовыми стрелками – регуляторные воздействия. Голубыми прямоугольниками выделены белки, обсуждаемые в тексте.

классифицировать как маркеры, контролирующие ответы на широкий спектр стрессовых воздействий. Такая характеристика делает их особенно ценными для дальнейших исследований и потенциального применения в биотехнологии растений.

Процесс "innate immune system" (врожденная иммунная система) интересен тем, что регулируется тринадцатью участниками генной сети, которые могут рассматриваться как перспективные маркеры, ассоциированные с устойчивостью к патогенам (см. рис. 6, *б*). Ключевыми регуляторами этого процесса являются белки WRKT114 и AGO2, а также компоненты описанных выше молекулярно-генетических путей (GSK, PR1, PR2, JAZ1 и SOG1). WRKT114 активирует иммунный ответ при инфицировании *X. oryzae* pv. *oryzae* (Son et al., 2020). AGO2 регулирует врожденный иммунитет через микроРНК-опосредованную супрессию генов-мишеней при инфицировании *P. syringae* pv. *tomato* (Zhang X. et al., 2011). Остальные компоненты также вносят существенный вклад в регуляцию иммунного ответа растений (Song Y. et al., 2021; Johnson et al., 2023; Javed et al., 2024).

Характеризация генов-маркеров по эволюционному возрасту

Оценка эволюционного возраста генов может дать важную информацию при планировании селекционных программ, позволяя прогнозировать специфичность, функциональную консервативность и фенотипические эффекты генов-кандидатов. Применение анализа эволюционного возраста генов при планировании экспериментов иллюстрирует работа по интрогрессии гена *Xa21* риса. Этот ген обеспечивает устойчивость к бактериальному ожогу риса, вызываемому *X. oryzae* pv. *oryzae*. *Xa21* был изолирован из дикого вида *Oryza longistaminata* и является эволюционно молодым геном, специфичным для рода *Oryza*. Введение гена *Xa21* в культивируемые сорта риса привело к созданию линий с высокой устойчивостью к болезни без негативного влияния на урожайность и качество зерна (Song W.Y. et al., 1995; Wang G.L. et al., 1996).

Другим примером является модификация гена *ERF922* для увеличения устойчивости риса к грибковым патогенам с использованием CRISPR/Cas9. *ERF922* – эволюционно молодой ген, участвующий в регулировании иммунного ответа риса. Его нокаут привел к повышению устойчивости к «взрыву риса» без негативного влияния на рост растения (Wang F. et al., 2016).

Проведенный нами филостратиграфический анализ генной сети выявил, что средний эволюционный возраст потенциальных генов-маркеров группы "innate immune response" составляет 605 млн лет, что существенно меньше аналогичного показателя для группы "response to stress" (1270 млн лет). Эти данные подтверждают представления об эволюционной молодости иммунных механизмов (Han, 2019). В группе "innate immune response" возрастной диапазон простирается от OsPR1a (менее 1 млн лет) до OsGSK2 (более 2101 млн лет), тогда как в группе "response to stress" – от OsSOG1 (306 млн лет) до Ser/Thr protein kinase (1714 млн лет).

Известно, что гены с большим эволюционным возрастом функционируют в наиболее фундаментальных процессах (Wolf et al., 2009; Domazet-Lošo, Tautz, 2010). Вариации в этих генах могут влиять на множество фенотипических признаков, что может усложнять селекцию по целевым свойствам. В связи с этим для маркер-ориентированной селекции наиболее перспективными представляются эволюционно молодые гены сети: *OsPR5*, *OsNAC6*, *OsMPK3*, *R2R3-MYB*, *OsPR1b* и *histone H4*.

Заключение

В данной работе для поиска потенциальных генов-маркеров, направленных на повышение устойчивости риса к *R. solani* в условиях избытка азота, был применен системный подход, включающий широкий круг биоинформатических методов, реализованных в когнитивной системе Smart crop, ANDSystem и других известных биоинформатических ресурсах. Системный анализ, реализованный в виде конвейерной обработки данных, включал: 1 – исследование дифференциальной экспрессии генов; 2 – реконструкцию и анализ генных сетей; 3 – анализ перепредставленности биологических процессов; 4 – анализ эволюции генных сетей методом филостратиграфического анализа; 5 – анализ омиксных данных по коэкспрессии некодирующих РНК.

Анализ молекулярно-генетической сети взаимодействий, связывающих ответы риса на избыток азота и инфицирование *R. solani*, позволил предложить механизмы, объясняющие снижение устойчивости риса к грибку на

фоне повышенных концентраций азотных удобрений. Выделены три потенциальных пути. Первый – опосредован через OsGSK2, который может связывать ответы растения на избыток азота и на инфекцию *R. solani*. При повышенном уровне OsGSK2 способен снижать устойчивость растений к грибку, что показано на примере *V. dahliae*, поражающего арабидопсис и хлопок. По нашим данным, уровень активного OsGSK2 может быть повышен при избытке азота из-за снижения экспрессии его ингибитора OsABI2. Второй – путь OsMYB44-OsWRKY6-OsPR1. Все участники этого пути имеют отношение к защите растений от биотических стрессов. Третий – путь SOG1-Rad51-PR1/PR2: от транскрипционного фактора SOG1 через регулятор транскрипции генов иммунного ответа Rad51 к генам *PR1* и *PR2*, незаменимым участникам ответа на патогены.

Реконструкция расширенной генной сети позволила идентифицировать потенциальные маркеры для селекции, направленной на повышение устойчивости к патогенам, таким как *R. solani*, в условиях избытка азота. Найденные маркеры разделены на две группы: маркеры, контролирующие ответы риса на широкий круг стрессов (7 генов) и маркеры-модуляторы иммунной системы (11 генов).

Среди наиболее важных маркеров выделяются гены, являющиеся ключевыми участниками регуляторных путей, лежащих в основе генной сети ответа риса на патоген *R. solani* в условиях избытка азота (*OsGSK2*, *JAZ1*, *PR1*/*PR2*, *SOG1*).

Полученные теоретические результаты могут послужить фундаментом для дальнейших экспериментальных работ по созданию новых сортов риса, обладающих повышенной устойчивостью к патогенам в условиях избыточного содержания азотных удобрений. Проведенные исследования открывают перспективы для разработки инновационных стратегий в селекции риса, направленных на оптимизацию баланса между урожайностью и устойчивостью к заболеваниям в современных агротехнических условиях.

Список литературы / References

Деменков П.С., Сайк О.В., Иванисенко Т.В., Колчанов Н.А., Кочетов А.В., Иванисенко В.А. Приоритизация генов картофеля, вовлеченных в формирование селекционно-значимых признаков, с использованием базы знаний SOLANUM TUBEROSUM. *Вавиловский журнал генетики и селекции.* 2019;23(3):312-319. doi 10.18699/VJ19.501
[Demenkov P.S., Saik O.V., Ivanisenko T.V., Kolchanov N.A., Kochetov A.V., Ivanisenko V.A. Prioritization of potato genes involved in the formation of agronomically valuable traits using the SOLANUM TUBEROSUM knowledge base. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2019;23(3):312-319. doi 10.18699/VJ19.501]

Иванисенко В.А., Басов Н.В., Макарова А.А., Вензель А.С., Рогачев А.Д., Деменков П.С., Иванисенко Т.В., Клещев М.А., Гайслер Е.В., Мороз Г.Б., Плеско В.В., Сотникова Ю.С., Патрушев Ю.В., Ломиворотов В.В., Колчанов Н.А., Покровский А.Г. Применение генных сетей к анализу результатов метаболомного скрининга плазмы крови пациентов с послеоперационным делирием. *Вавиловский журнал генетики и селекции*. 2023;27(7):768-775. doi 10.18699/VJGB-23-89
[Ivanisenko V.A., Basov N.V., Makarova A.A., Venzel A.S., Rogachev A.D., Demenkov P.S., Ivanisenko T.V., Kleshchev M.A., Gaisler E.V., Moroz G.B., Plesko V.V., Sotnikova Y.S., Patrushev Y.V., Lomivorotov V.V., Kolchanov N.A., Pokrovsky A.G. Gene networks for use in metabolomic data analysis of blood plasma from patients with postoperative delirium. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):768-775. doi 10.18699/VJGB-23-89]

Иванисенко Т.В., Сайк О.В., Деменков П.С., Хлесткин В.К., Хлесткина Е.К., Колчанов Н.А., Иванисенко В.А. База знаний SOLANUM TUBEROSUM: раздел по молекулярно-генетической регуляции метаболических путей. *Вавиловский журнал генетики и селекции*. 2018;22(1):8-17. doi 10.18699/VJ18.325
[Ivanisenko T.V., Saik O.V., Demenkov P.S., Khlestkin V.K., Khlestkina E.K., Kolchanov N.A., Ivanisenko V.A. The SOLANUM TUBEROSUM knowledge base: the section on molecular-genetic regulation of metabolic pathways. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2018;22(1):8-17. doi 10.18699/VJ18.325 (in Russian)]

Колчанов Н.А., Игнатьева Е.В., Подколодная О.А., Лихошвай В.А., Матушкин Ю.Г. Генные сети. *Вавиловский журнал генетики и селекции*. 2013;17(4/2):833-850
[Kolchanov N.A., Ignatyeva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/2):833-850 (in Russian)]

Кумейко Ю.В., Паращенко В.Н., Кремзин Н.М. Применение ингибитора нитрификации для снижения потерь азота и повышения эффективности азотного удобрения в рисоводстве. *Сб. науч. трудов Ставропольского НИИ животноводства и кормопроизводства*. 2013;3(6):144-147
[Kumeiko Yu.V., Paraschenko V.N., Kremzin N.M. Application of nitrification inhibitor to reduce nitrogen losses and increase the efficiency of nitrogen fertilizers in rice growing. *Sbornik Nauchnykh Trudov Stavropolskogo NII Zhivotnovodstva i Kormoproizvodstva = Proceedings of the Stavropol Research Institute of Animal Husbandry and Forage Production*. 2013;3(6):144-147 (in Russian)]

Мустафин З.С., Лашин С.А., Матушкин Ю.Г. Филостратиграфический анализ генных сетей заболеваний человека. *Вавиловский журнал генетики и селекции*. 2021;25(1):46-56. doi 10.18699/VJ21.006
[Mustafin Z.S., Lashin S.A., Matushkin Yu.G. Phylostratigraphic analysis of gene networks of human diseases. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(1):46-56. doi 10.18699/VJ21.006]

Bragina E.Y., Tiys E.S., Freidin M.B., Koneva L.A., Demenkov P.S., Ivanisenko V.A., Kolchanov N.A., Puzyrev V.P. Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis. *Immunogenetics*. 2014;66(7-8):457-465. doi 10.1007/s00251-014-0786-1

Bragina E.Y., Tiys E.S., Rudko A.A., Ivanisenko V.A., Freidin M.B. Novel tuberculosis susceptibility candidate genes revealed by the reconstruction and analysis of associative networks. *Infect. Genet. Evol*. 2016;46:118-123. doi 10.1016/j.meegid.2016.10.030

Bragina E.Y., Gomboeva D.E., Saik O.V., Ivanisenko V.A., Freidin M.B., Nazarenko M.S., Puzyrev V.P. Apoptosis genes as a key to identification of inverse comorbidity of Huntington's disease and cancer. *Int. J. Mol. Sci*. 2023;24(11):9385. doi 10.3390/ijms24119385

Cao X., Wei Y., Shen B., Liu L., Mao J. Interaction of the transcription factors BES1/BZR1 in plant growth and stress response. *Int. J. Mol. Sci*. 2024;25(13):6836. doi 10.3390/ijms25136836

Chao H., Zhang S., Hu Y., Ni Q., Xin S., Zhao L., Ivanisenko V.A., Orlov Y.L., Chen M. Integrating omics databases for enhanced crop breeding. *J. Integr. Bioinform*. 2023;20(4):20230012. doi 10.1515/jib-2023-0012

Charagh S., Hui S., Wang J., Raza A., Zhou L., Xu B., Zhang Y., Sheng Z., Tang S., Hu S., Hu P. Unveiling innovative approaches to mitigate metals/metalloids toxicity for sustainable agriculture. *Physiol. Plant*. 2024;176(2):e14226. doi 10.1111/ppl.14226

Chen K., Li G.J., Bressan R.A., Song C.P., Zhu J.K., Zhao Y. Abscisic acid dynamics, signaling, and functions in plants. *J. Integr. Plant Biol*. 2020;62(1):25-54. doi 10.1111/jipb.12899

Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol*. 2011-2012;11(3-4):149-161. doi 10.3233/ISB-2012-0449

Diédhiou C.J., Popova O.V., Dietz K.J., Golldack D. The SNF1-type serine-threonine protein kinase SAPK4 regulates stress-responsive gene expression in rice. *BMC Plant Biol*. 2008;8:49. doi 10.1186/1471-2229-8-49

Domazet-Lošo T., Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*. 2010;468(7325):815-818. doi 10.1038/nature09632

Friis E.M., Pedersen K.R., Crane P.R. Araceae from the Early Cretaceous of Portugal: evidence on the emergence of monocotyledons. *Proc. Natl. Acad. Sci. USA*. 2004;101(47):16565-16570. doi 10.1073/pnas.0407174101

Han G.Z. Origin and evolution of the plant immune system. *New Phytol*. 2019;222(1):70-83. doi 10.1111/nph.15596

Hückelhoven R. Cell wall-associated mechanisms of disease resistance and susceptibility. *Annu. Rev. Phytopathol*. 2007;45(1):101-127. doi 10.1146/annurev.phyto.45.062806.094325

Im J.H., Choi C., Park S.R., Hwang D.J. The OsWRKY6 transcriptional cascade functions in basal defense and Xa1-mediated defense of rice against *Xanthomonas oryzae* pv. *oryzae*. *Planta*. 2022;255(2):47. doi 10.1007/s00425-022-03830-5

Im J.H., Choi C., Jung M.Y., Park S.R., Hwang D.J. The *OsICS1* is directly regulated by OsWRKY6 and increases resistance against *Xanthomonas oryzae* pv. *oryzae*. *Planta*. 2024;259(6):124. doi 10.1007/s00425-024-04405-2

Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol*. 2015;9(Suppl.2):S2. doi 10.1186/1752-0509-9-S2-S2

Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl.1):34. doi 10.1186/s12859-018-2567-6

Ivanisenko V.A., Gaisler E.V., Basov N.V., Rogachev A.D., Cheresiz S.V., Ivanisenko T.V., Demenkov P.S., Mishchenko E.L., Khripko O.P., Khripko Y.I., Voevoda S.M., Karpenko T.N., Velichko A.J., Voevoda M.I., Kolchanov N.A., Pokrovsky A.G. Plasma metabolomics and gene regulatory networks analysis reveal the role of nonstructural SARS-CoV-2 viral proteins in metabolic dysregulation in COVID-19 patients. *Sci. Rep*. 2022;12(1):19977. doi 10.1038/s41598-022-24170-0

Javed T., Wang W., Yang B., Shen L., Sun T., Gao S.J., Zhang S. Pathogenesis related-1 proteins in plant defense: regulation and functional diversity. *Crit. Rev. Biotechnol.* 2024;1-9. doi 10.1080/07388551.2024.2344583

Johnson L.Y.D., Major I.T., Chen Y., Yang C., Vanegas-Cano L.J., Howe G.A. Diversification of JAZ-MYC signaling function in immune metabolism. *New Phytol*. 2023;239(6):2277-2291. doi 10.1111/nph.19114

Junior A.T.D., Farias D.D., dos Santos R.S., do Amaral M.N., Arge L.W.P., Oliveira D.D.C., Silveira S.F.S., Sousa R.O., Braga E.J.B., Maia L.C., Oliveira A.C. The quest for more tolerant rice: How high concentrations of iron affect alternative splicing? *Transcriptomics*. 2015;3:2. doi 10.4172/2329-8936.1000122

Kumar K., Mandal S.N., Neelam K., de Los Reyes B.G. MicroRNA-mediated host defense mechanisms against pathogens and herbivores in rice: balancing gains from genetic resistance with trade-offs to productivity potential. *BMC Plant Biol*. 2022;22(1):351. doi 10.1186/s12870-022-03723-5

Kumar S., Shah S.H., Vimala Y., Jatav H.S., Ahmad P., Chen Y., Siddique K.H.M. Abscisic acid: metabolism, transport, crosstalk with other plant growth regulators, and its role in heavy metal stress mitigation. *Front. Plant Sci*. 2022;13:972856. doi 10.3389/fpls.2022.972856

Li Q.F., Lu J., Zhou Y., Wu F., Tong H.N., Wang J.D., Yu J.W., Zhang C.Q., Fan X.L., Liu Q.Q. Abscisic acid represses rice lamina joint inclination by antagonizing brassinosteroid biosynthesis and signaling. *Int. J. Mol. Sci*. 2019;20(19):4908. doi 10.3390/ijms20194908

Ogita N., Okushima Y., Tokizawa M., Yamamoto Y.Y., Tanaka M., Seki M., Makita Y., Matsui M., Okamoto-Yoshiyama K., Sakamoto T., Kurata T., Hiruma K., Saijo Y., Takahashi N., Umeda M. Identifying the target genes of SUPPRESSOR OF GAMMA RESPONSE 1, a master transcription factor controlling DNA damage response in *Arabidopsis*. *Plant J*. 2018;94(3):439-453. doi 10.1111/tpj.13866

Rogachev A.D., Alemasov N.A., Ivanisenko V.A., Ivanisenko N.V., Gaisler E.V., Oleshko O.S., Cheresiz S.V., Mishinov S.V., Stupak V.V., Pokrovsky A.G. Correlation of metabolic profiles of plasma and cerebrospinal fluid of high-grade glioma patients. *Metabolites*. 2021;11(3):133. doi 10.3390/metabo11030133

Rose J.K.C., Catalá C., Gonzalez-Carranza Z.H., Roberts J.A. Cell wall disassembly. In: Annual Plant Reviews online. Vol. 8. The Plant Cell Wall. Wiley, 2018;264-324. doi 10.1002/9781119312994.apr0075

Saik O.V., Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Interactome of the hepatitis C virus: literature mining with ANDSystem. *Virus Res*. 2016;218:40-48. doi 10.1016/j.virusres.2015.12.003

Saik O.V., Demenkov P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Dosenko V.E., Zolotareva O.I., Choynzonov E.L., Hofestaedt R., Ivanisenko V.A. Search for new candidate genes involved in the comorbidity of asthma and hypertension based on automatic analysis of scientific literature. *J. Integr. Bioinform*. 2018;15(4):20180054. doi 10.1515/jib-2018-0054

Saik O.V., Nimaev V.V., Usmonov D.B., Demenkov P.S., Ivanisenko T.V., Lavrik I.N., Ivanisenko V.A. Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema using an analysis of associative gene networks with ANDSystem. *BMC Med. Genomics*. 2019;12(Suppl.2):47. doi 10.1186/s12920-019-0492-9

Senapati M., Tiwari A., Sharma N., Chandra P., Bashyal B.M., Ellur R.K., Bhowmick P.K., Bollinedi H., Vinod K.K., Singh A.K., Krishnan S.G. *Rhizoctonia solani* Kühn pathophysiology: status and prospects of sheath blight disease management in rice. *Front. Plant Sci.* 2022;13:881116. doi 10.3389/fpls.2022.881116

Shen J., Wang X., Song H., Wang M., Niu T., Lei H., Qin C., Liu A. Physiology and transcriptomics highlight the underlying mechanism of sunflower responses to drought stress and rehydration. *iScience*. 2023;26(11):108112. doi 10.1016/j.isci.2023.108112

Shim J.S., Jung C., Lee S., Min K., Lee Y.W., Choi Y., Lee J.S., Song J.T., Kim J.K., Choi Y.D. AtMYB44 regulates *WRKY70* expression and modulates antagonistic interaction between salicylic acid and jasmonic acid signaling. *Plant J.* 2013;73(3):483-495. doi 10.1111/tpj.12051

Sircar S., Musaddi M., Parekh N. NetREx: Network-based Rice Expression Analysis Server for abiotic stress conditions. *Database* (*Oxford*). 2022;2022:baac060. doi 10.1093/database/baac060

Son S., An H.K., Seol Y.J., Park S.R., Im J.H. Rice transcription factor WRKY114 directly regulates the expression of *OsPR1a* and *Chitinase* to enhance resistance against *Xanthomonas oryzae* pv. *oryzae*. *Biochem. Biophys. Res. Commun*. 2020;533(4):1262-1268. doi 10.1016/j.bbrc.2020.09.141

Song H., Duan Z., Zhang J. WRKY transcription factors modulate flowering time and response to environmental changes. *Plant Physiol. Biochem*. 2024;210:108630. doi 10.1016/j.plaphy.2024.108630

Song L., Fang Y., Chen L., Wang J., Chen X. Role of non-coding RNAs in plant immunity. *Plant Commun*. 2021;2(3):100180. doi 10.1016/j.xplc.2021.100180

Song W.Y., Wang G.L., Chen L.L., Kim H.S., Pi L.Y., Holsten T., Gardner J., Wang B., Zhai W.X., Zhu L.H., Fauquet C., Ronald P. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science*. 1995;270(5243):1804-1806. doi 10.1126/science.270.5243.1804

Song Y., Zhai Y., Li L., Yang Z., Ge X., Yang Z., Zhang C., Li F., Ren M. BIN2 negatively regulates plant defence against *Verticillium dahliae* in *Arabidopsis* and cotton. *Plant Biotechnol. J.* 2021;19(10):2097-2112. doi 10.1111/pbi.13640.

Statello L., Guo C.J., Chen L.L., Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol*. 2021;22(2):96-118. doi 10.1038/s41580-020-00315-9

Steber C.M., McCourt P. A role for brassinosteroids in germination in *Arabidopsis*. *Plant Physiol.* 2001;125(2):763-769. doi 10.1104/pp.125.2.763.

Sun H.L., Wang X.J., Ding W.H., Zhu S.Y., Zhao R., Zhang Y.X., Xin Q., Wang X.F., Zhang D.P. Identification of an important site for function of the type 2C protein phosphatase ABI2 in abscisic acid signalling in *Arabidopsis*. *J. Exp. Bot*. 2011;62(15):5713-5725. doi 10.1093/jxb/err274

Supriya P., Srividya G.K., Solanki M., Manvitha D., Prakasam V., Balakrishnan M., Neeraja C.N., Srinivasa Rao Ch, Sundaram R.M., Mangrauthia S.K. Identification and expression analysis of long non-coding RNAs of rice induced during interaction with *Rhizoctonia solani*. *Physiol. Mol. Plant Pathol.* 2024;134:102389. doi 10.1016/j.pmpp.2024.102389

Tong Y.-B., Shi M.-W., Qian S.H., Chen Y.-J., Luo Z.-H., Tu Y.-X., Xiong Y.-L., Geng Y.-J., Chen C., Chen Z.-X. GenOrigin: a comprehensive protein-coding gene origination database on the evolutionary timescale of life. *J. Genet. Genomics*. 2021;48(12):1122-1129. doi 10.1016/j.jgg.2021.03.018

van der Kooi C.J., Ollerton J. The origins of flowering plants and pollinators. *Science*. 2020;368(6497):1306-1308. doi 10.1126/science.aay3662

Vincentz M., Cara F.A., Okura V.K., da Silva F.R., Pedrosa G.L., Hemerly A.S., Capella A.N., Marins M., Ferreira P.C., França S.C., Grivet L., Vettore A.L., Kemper E.L., Burnquist W.L., Targon M.L., Siqueira W.J., Kuramae E.E., Marino C.L., Camargo L.E., Carrer H., Coutinho L.L., Furlan L.R., Lemos M.V., Nunes L.R., Gomes S.L., Santelli R.V., Goldman M.H., Bacci M. Jr, Giglioti E.A., Thiemann O.H., Silva F.H., Van Sluys M.A., Nobrega F.G., Arruda P., Menck C.F. Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. *Plant Physiol*. 2004;134(3):951-959. doi 10.1104/pp.103.033878

Volyanskaya A.R., Antropova E.A., Zubairova U.S., Demenkov P.S., Venzel A.S., Orlov Y.L., Makarova A.A., Ivanisenko T.V., Gorshkova T.A., Aglyamova A.R., Kolchanov N.A., Chen M., Ivanisenko V.A. Reconstruction and analysis of the gene regulatory network for cell wall function in *Arabidopsis thaliana* L. leaves in response to water deficit. *Vavilov J. Genet. Breed*. 2023;27(8):1031-1041. doi 10.18699/VJGB-23-118

Wang F., Wang C., Liu P., Lei C., Hao W., Gao Y., Liu Y.G., Zhao K. Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene *OsERF922*. *PLoS One*. 2016;11(4):e0154027. doi 10.1371/journal.pone.0154027

Wang F., Yang F., Zhu D., Saniboere B., Zhou B., Peng D. MYB44 plays key roles in regulating plant responses to abiotic and biotic stress, metabolism, and development. *J. Plant Biochem. Biotechnol*. 2024;33(4):462-473. doi 10.1007/s13562-023-00864-y

Wang G.L., Song W.Y., Ruan D.L., Sideris S., Ronald P.C. The cloned gene, *Xa21*, confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in transgenic plants. *Mol. Plant-Microbe Interact*. 1996;9(9):850-855. doi 10.1094/mpmi-9-0850

Wang H., Tang J., Liu J., Hu J., Liu J., Chen Y., Cai Z., Wang X. Abscisic acid signaling inhibits brassinosteroid signaling through dampening the dephosphorylation of BIN2 by ABI1 and ABI2. *Mol. Plant*. 2018;11:315-325. doi 10.1016/j.molp.2017.12.013

Wang S., Durrant W.E., Song J., Spivey N.W., Dong X. *Arabidopsis* BRCA2 and RAD51 proteins are specifically involved in defense gene transcription during plant immune responses. *Proc. Natl. Acad. Sci. USA*. 2010;107(52):22716-22721. doi 10.1073/pnas.1005978107

Wang X., Wang Y., Piñeros M.A., Wang Z., Wang W., Li C., Wu Z., Kochian L.V., Wu P. Phosphate transporters OsPHT1;9 and OsPHT1;10 are involved in phosphate uptake in rice. *Plant Cell Environ*. 2014;37(5):1159-1170. doi 10.1111/pce.12224

Wolf Y.I., Novichkov P.S., Karev G.P., Koonin E.V., Lipman D.J. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. *Proc. Natl. Acad. Sci. USA*. 2009;106(18):7273-7280. doi10.1073/pnas.0901808106

Xie K., Wu C., Xiong L. Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol.* 2006;142(1):280-293. doi 10.1104/pp.106.084475

Xing J., Cao X., Zhang M., Wei X., Zhang J., Wan X. Plant nitrogen availability and crosstalk with phytohormones signallings and their biotechnology breeding application in crops. *Plant Biotechnol. J*. 2023;21(7):1320-1342. doi 10.1111/pbi.13971

Xiong Q., Hu J., Wei H., Zhang H., Zhu J. Relationship between plant roots, rhizosphere microorganisms, and nitrogen and its special focus on rice. *Agriculture*. 2021;11(3):234. doi 10.3390/agriculture11030234

Yang J., Duan G., Li C., Liu L., Han G., Zhang Y., Wang C. The crosstalks between jasmonic acid and other plant hormone signaling highlight the involvement of jasmonic acid as a core component in plant response to biotic and abiotic stresses. *Front. Plant Sci*. 2019;10:1349. doi 10.3389/fpls.2019.01349

Yang Y.X., Ahammed G.J., Wu C., Fan S.Y., Zhou Y.H. Crosstalk among jasmonate, salicylate and ethylene signaling pathways in plant disease and immune responses. *Curr. Protein Pept. Sci.* 2015;16(5):450-461. doi 10.2174/1389203716666150330141638

Yokotani N., Tsuchida-Mayama T., Ichikawa H., Mitsuda N., Ohme-Takagi M., Kaku H., Minami E., Nishizawa Y. OsNAC111, a blast disease-responsive transcription factor in rice, positively regulates the expression of defense-related genes. *Mol. Plant-Microbe Interact*. 2014;27(10):1027-1034. doi 10.1094/MPMI-03-14-0065-R

Yoshiyama K.O., Kimura S. Ser-Gln sites of SOG1 are rapidly hyperphosphorylated in response to DNA double-strand breaks. *Plant Signal. Behav*. 2018;13(6):e1477904. doi 10.1080/15592324.2018.1477904

Zhang S., Su J., Ma S., Wang H., Wang X., He K., Wang H., Canfield D.E. Eukaryotic red and green algae populated the tropical ocean 1400 million years ago. *Precambrian Res*. 2021;357:106166. doi 10.1016/j.precamres.2021.106166

Zhang X., Zhao H., Gao S., Wang W.C., Katiyar-Agarwal S., Huang H.D., Raikhel N., Jin H. *Arabidopsis* Argonaute 2 regulates innate immunity via miRNA393*-mediated silencing of a Golgi-localized SNARE gene *MEMB12*. *Mol. Cell*. 2011;42(3):356-366. doi 10.1016/j.molcel.2011.04.010

Zhou X., Zha M., Huang J., Li L., Imran M., Zhang C. StMYB44 negatively regulates phosphate transport by suppressing expression of *PHOSPHATE1* in potato. *J. Exp. Bot.* 2017;68(5):1265-1281. doi 10.1093/jxb/erx026

Zolotareva O., Saik O.V., Königs C., Bragina E.Y., Goncharova I.A., Freidin M.B., Dosenko V.E., Ivanisenko V.A., Hofestädt R. Comorbidity of asthma and hypertension may be mediated by shared genetic dysregulation and drug side effects. *Sci. Rep.* 2019;9(1):16302. doi 10.1038/s41598-019-52762-w

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 27.09.2024. После доработки 14.11.2024. Принята к публикации 14.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):974-981

doi 10.18699/vjgb-24-104

Методы реконструкции генных регуляторных сетей на основе транскриптомных данных отдельных клеток

М.А. Рыбаков[1, 2], Н.А. Омельянчук[1], Е.В. Землянская[1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
ezemlyanskaya@bionet.nsc.ru

Аннотация. Генные регуляторные сети – интерпретируемые графовые модели регуляции экспрессии генов – являются важным инструментом для понимания и исследования механизмов, которые клетки реализуют в процессе развития и при ответе на различные внутренние и внешние стимулы. Исторически первый подход для реконструкции генных регуляторных сетей основывался на анализе литературных сведений, в том числе обобщенных в базах данных. В настоящее время основной способ системной реконструкции генных регуляторных сетей – анализ омиксных (в первую очередь транскриптомных) данных; разработан ряд математических подходов для решения этой задачи. Развитие технологий получения омиксных данных для отдельных клеток сделало возможным проведение широкомасштабных молекулярно-генетических исследований с беспрецедентно высоким уровнем разрешения. В частности, появилась возможность реконструировать генные регуляторные сети для отдельных клеточных типов и для различных стадий развития клеток. Однако технические и биологические особенности омиксных данных отдельных клеток требуют специальных программ для решения этой задачи. В обзоре описаны подходы и программы, которые разработаны и используются для построения генных регуляторных сетей по транскриптомным данным отдельных клеток (scRNA-seq). Разбираются преимущества применения транскриптомных данных для отдельных клеток по сравнению с транскриптомами многоклеточных образцов, а также их недостатки в рамках решения задачи реконструкции регуляторных генных сетей. Существенное внимание уделяется повышению точности генных регуляторных сетей, построенных по транскриптомным данным отдельных клеток с помощью привлечения других омиксных данных, в первую очередь данных по сайтам связывания транскрипционных факторов и профилирования районов открытого хроматина (scATAC-seq). Рассматриваются вопросы применимости получаемых сетей в молекулярно-генетических исследованиях, приводятся примеры успешного использования генных регуляторных сетей, реконструированных различными методами с применением омиксных данных отдельных клеток для решения конкретных биологических задач. Обсуждаются перспективные направления развития этой области.

Ключевые слова: регуляторная генная сеть; данные для отдельных клеток; секвенирование РНК; scRNA-seq; scATAC-seq.

Для цитирования: Рыбаков М.А., Омельянчук Н.А., Землянская Е.В. Методы реконструкции генных регуляторных сетей на основе транскриптомных данных отдельных клеток. *Вавиловский журнал генетики и селекции*. 2024; 28(8):974-981. doi 10.18699/vjgb-24-104

Финансирование. Работа выполнена в рамках бюджетного проекта FWNR-2022-0020.

Reconstruction of gene regulatory networks from single cell transcriptomic data

M.A. Rybakov[1, 2], N.A. Omelyanchuk[1], E.V. Zemlyanskaya[1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
ezemlyanskaya@bionet.nsc.ru

Abstract. Gene regulatory networks (GRNs) – interpretable graph models of gene expression regulation – are a pivotal tool for understanding and investigating the mechanisms utilized by cells during development and in response to various internal and external stimuli. Historically, the first approach for the GRN reconstruction was based on the analysis of published data (including those summarized in databases). Currently, the primary GRN inference approach is the analysis of omics (mainly transcriptomic) data; a number of mathematical methods have been adapted for that. Obtaining omics data for individual cells has made it possible to conduct large-scale molecular genetic studies with an extremely high resolution. In particular, it has become possible to reconstruct GRNs for individual cell types and for various cell states. However, technical and biological features of single-cell omics data require specific approaches for

© Рыбаков М.А., Омельянчук Н.А., Землянская Е.В., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

GRN inference. This review describes the approaches and programs that are used to reconstruct GRNs from single-cell RNA sequencing (scRNA-seq) data. We consider the advantages of using scRNA-seq data compared to bulk RNA-seq, as well as challenges in GRN inference. We pay specific attention to state-of-the-art methods for GRN reconstruction from single-cell transcriptomes recruiting other omics data, primarily transcription factor binding sites and open chromatin profiles (scATAC-seq), in order to increase inference accuracy. The review also considers the applicability of GRNs reconstructed from single-cell omics data to recover and characterize various biological processes. Future perspectives in this area are discussed.

Key words: gene regulatory network; single-cell data; RNA sequencing; scRNA-seq; scATAC-seq.

For citation: Rybakov M.A., Omelyanchuk N.A., Zemlyanskaya E.V. Reconstruction of gene regulatory networks from single cell transcriptomic data. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024; 28(8):974-981. doi 10.18699/vjgb-24-104

Введение

Генная сеть – это управляющая формированием определенных признаков группа координированно экспрессирующихся генов, которые взаимодействуют друг с другом через кодируемые ими РНК и белки, а также продукты активности белков (Колчанов и др., 2013). Генные сети – центральный объект системной биологии. С целью более глубокого исследования отдельных аспектов выделяют специализированные типы генных сетей. Особое место среди них занимают генные регуляторные сети (ГРС), которые описывают осуществляемое транскрипционными факторами (ТФ) управление экспрессией генов – ключевой механизм гибкой реализации генетической информации (Huynh-Thu, Sanguinetti, 2019). Генные регуляторные сети визуализируют в виде графа взаимодействий между ТФ и регулируемыми ими генами (рис. 1, *а*) (Badia-i-Mompel et al., 2023). Каждая вершина в ГРС представляет собой ген (причем некоторые из них кодируют ТФ), а каждое ребро соответствует регуляторным отношениям между генами, кодирующими ТФ, и другими генами (эти отношения могут отражать реальные молекулярные взаимодействия ТФ с промоторами генов-мишеней или только их статистическую взаимосвязь). Ребро может иметь знак, указывающий, описывает ли оно активацию или ингибирование транскрипции, а также вес, отражающий, насколько сильно влияние регулятора. Таким образом, ГРС являются моделями логики регуляторных событий между генами в ходе выполнения клеточных программ (Tieri, Castiglione, 2021). Они представляют действенную альтернативу классическому моделированию в дифференциальных уравнениях в тех случаях, когда информация по кинетике процессов недоступна.

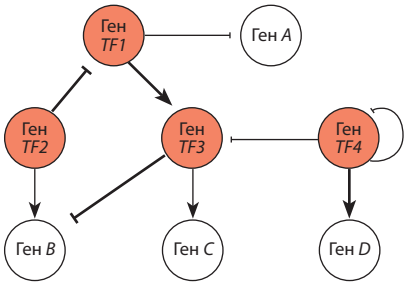
Генные регуляторные сети можно построить на основе информации о ТФ и их генах-мишенях из опубликованных статей или вывести *de novo* из транскриптомных данных (Badia-i-Mompel et al., 2023). Для решения этой задачи широко применялись транскриптомные даные клеточных популяций, получаемые секвенированием РНК (RNA-seq). В них уровень экспрессии каждого гена суммирован по всем клеткам в образце ткани или органа, взятом для секвенирования. Эти данные можно представить в виде так называемой матрицы экспрессии, в которой приводятся значения уровней экспрессии для каждого гена (им соответствуют строки) в различных образцах (им соответствуют столбцы) (см. рис. 1, *б*). Исходя из того, что уровни экспрессии генов, представленные в этих матрицах, являются результатом регуляции, осуществляемой путем связывания ТФ с промоторами генов-мишеней, можно построить математическую модель, которая объясняла бы наблюдаемую экспрессию генов (Mercatelli et al., 2020; Nguyen et al., 2021). На этой предпосылке основано большинство современных методов реконструкции ГРС по транскриптомным данным (Mercatelli et al., 2020). В настоящее время реконструкция ГРС по данным RNA-seq – одно из направлений системной биологии, в рамках которого разработано большое количество методов и компьютерных программ (Nguyen et al., 2019; Mercatelli et al., 2020).

В то же время описанный выше подход имеет слабые стороны. Во-первых, транскриптомные данные не содержат в явном виде сведения о конкретных регуляторных событиях (например, о связывании ТФ с промотором регулируемого им гена), все связи выводятся математически по уровням экспрессии генов. В результате в ГРС могут быть реконструированы несуществующие (ошибочные) связи. Привлечение данных, которые напрямую описывают управление транскрипцией (например, полногеномных профилей открытого хроматина или сайтов связывания ТФ), может значительно улучшить точность построенных ГРС (Sönmezer et al., 2020; Isbel et al., 2022). Во-вторых, данные RNA-seq не учитывают гетерогенность клеточных популяций, тогда как экспрессия генов может кардинально различаться в клетках разного типа. Эта проблема решается секвенированием транскриптомов отдельных клеток (scRNA-seq) (Tang et al., 2009).

Транскриптомные данные отдельных клеток представляют собой матрицу экспрессии, в которой строки соответствуют генам, а столбцы – клеткам (см. рис. 1, *в*), которые могут быть объединены по клеточным типам с помощью специально разработанных подходов (Luecken, Theis, 2019). Данные scRNA-seq открывают возможности исследовать биологические процессы на уровне отдельных типов клеток и новые перспективы для реконструкции и анализа ГРС (Nguyen et al., 2021). ГРС для отдельных типов клеток позволят обнаружить регуляторные контуры, специфичные для их состояний или степеней дифференцировки.

В настоящем обзоре мы рассматриваем методы реконструкции ГРС на основе транскриптомных данных отдельных клеток, подробно останавливаемся на привлечении для этого других омиксных данных, в первую очередь по сайтам связывания ТФ и полногеномным профилям открытого хроматина. Особое внимание уделяется описанию биологических результатов, которых удалось достичь с применением этих подходов.

а



б

Образцы

	s1	s2	s3	s4
Ген *TF1*	3	2	2	3
Ген *TF2*	6	4	5	6
Ген *TF3*	6	3	4	5
Ген *TF4*	3	2	3	4
Ген *A*	2	1	2	2
Ген *B*	2	1	1	3
Ген *C*	1	1	2	1
Ген *D*	1	1	1	0

в

Клетки

	a	b	c	d	e	f	g	h	i	j
Ген *TF1*	1	0	0	0	0	0	1	0	0	1
Ген *TF2*	0	1	0	1	1	0	0	2	1	0
Ген *TF3*	1	0	0	0	0	1	2	0	1	1
Ген *TF4*	1	1	0	1	0	0	0	0	0	0
Ген *A*	0	0	0	1	0	0	0	1	0	0
Ген *B*	0	0	0	1	1	0	0	0	0	0
Ген *C*	0	0	0	0	0	1	0	0	0	0
Ген *D*	0	1	0	0	0	0	0	0	0	0

Рис. 1. Генная регуляторная сеть и транскриптомные данные, на основании которых она может быть построена.

а – визуализация графовой модели ГРС; *б* – матрица экспрессии, построенная по данным RNA-seq для нескольких образцов (s1–s4); *в* – матрица экспрессии, построенная по данным scRNA-seq для одного образца. В вершинах графа (*а*) находятся гены, ребра отображают наличие регуляторной связи, ее направление, тип (активация или ингибирование транскрипции) и величину (чем больше вес ребра, тем сильнее влияние регулятора на транскрипцию). Красные вершины графа соответствуют генам, кодирующим ТФ, белые – другим генам. В ГРС ребра исходят только из вершин, которые соответствуют генам, кодирующим ТФ. На панели (*в*) разными цветами обозначены разные типы клеток.

Транскриптомы отдельных клеток как источник информации для ГРС

Помимо того, что транскриптомные данные отдельных клеток открывают возможность реконструировать ГРС для индивидуальных типов клеток, они имеют и другие преимущества по сравнению с транскриптомами клеточных популяций для реконструкции ГРС. Поскольку количество взаимодействий в ГРС, как правило, достаточно велико, для их качественной реконструкции необходимо использовать большее число транскриптомных профилей (столбцы в матрице экспрессии, см. рис. 1). Это не всегда достижимо для данных RNA-seq (см. рис. 1, *б*) (Altay, 2012), в то время как данные scRNA-seq содержат представительный (от нескольких сотен до нескольких тысяч) набор транскриптомов (см. рис. 1, *в*) (Luecken, Theis, 2019).

Генные регуляторные сети призваны описывать динамику регуляции экспрессии генов в различных биологических процессах, включая дифференцировку клеток и их реакции на различные внутренние и внешние стимулы. Для их наиболее точной реконструкции по данным RNA-seq требуется временной ряд образцов. В отличие от RNA-seq, данные scRNAseq, полученные в результате секвенирования одного образца, могут содержать информацию об изменении экспрессии генов во времени, если клетки в образце участвуют в одном и том же непрерывном биологическом процессе (например, дифференцировке) и находятся на разных его стадиях (Saelens et al., 2019; Hou et al., 2023). В таком случае вычислительное размещение клеток вдоль псевдовременной траектории (где порядок клеток определяется расстоянием между их транскриптомами) позволяет в хорошем приближении реконструировать динамику экспрессии генов в ходе процесса.

Однако не следует забывать, что в некоторых образцах клетки могут находиться в статичном состоянии или участвовать в многочисленных независимых процессах, что делает невозможной реконструкцию имеющих биологический смысл псевдовременных траекторий (Pratapa et al., 2020). Поэтому при выборе метода для реконструкции ГРС крайне важно определить, присутствует ли псевдовременная информация в наборе транскриптомов отдельных клеток, поскольку некоторые методы разработаны исключительно для данных с клеточной динамикой, в то время как другие подходят только для статических данных. Существуют также методы, применимые к обоим типам данных.

Одновременно транскриптомные профили отдельных клеток имеют свои особенности, которые затрудняют их анализ в целом и реконструкцию ГРС в частности (Wagner et al., 2016; Nguyen et al., 2021). К ним относятся кратковременная активация или низкий уровень экспрессии некоторых генов, изменения экспрессии генов в зависимости от стадии клеточного цикла и другие факторы. Широкое применение технологии scRNA-seq в биологии вызвало разработку большого числа алгоритмов для анализа генерируемых ею данных, по разному решающих эти проблемы.

Реконструкция ГРС по данным scRNA-seq

В этом разделе мы описываем основные категории популярных алгоритмов, используемых для построения ГРС по данным scRNA-seq: анализ корреляции и взаимной информации, регрессию, байесовские и логические сети, математическое моделирование на основе дифференциальных уравнений (рис. 2). Нужно отметить, что в ряде исследований, где проводилась оценка эффективности соответствующих инструментов с использованием как смоделированных, так и реальных данных scRNA-seq, ни один метод не был признан универсально лучшим (Chen, Mar, 2018; Blencowe et al., 2019; Pratapa et al., 2020). Такая вариативность может быть обусловлена тем, что каждый из этих методов подходит для конкретных типов и источников данных, для которых он был разработан.

Корреляция

Корреляция Пирсона, общепризнанный статистический индекс для расчета ассоциации между двумя переменными, была применена для измерения коэкспрессии генов, кодирующих ТФ, и их потенциальных мишеней в наборах данных RNA-seq и scRNA-seq (Hong et al., 2013; Nguyen et al., 2021). Будучи симметричной в своих аргументах, корреляция не предсказывает направленность регуляторного взаимодействия. Она может определять ассоциации между парами генов, которые не обязательно

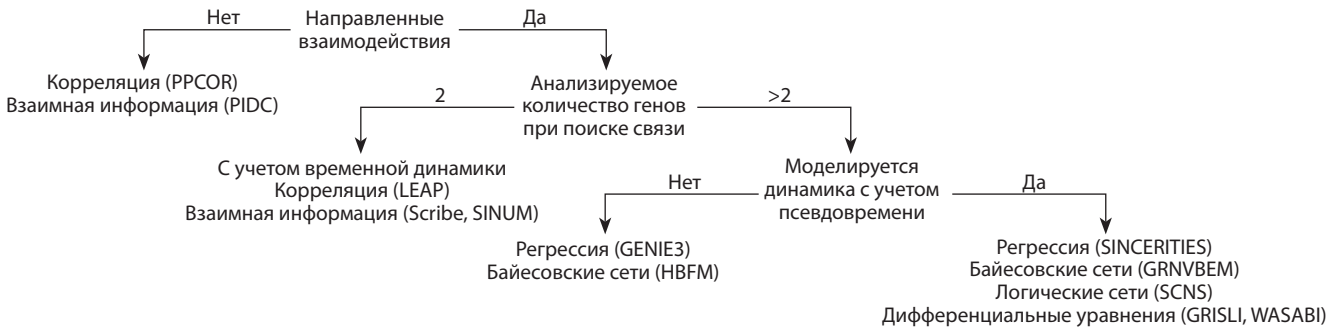


Рис. 2. Основные категории популярных алгоритмов, используемых для построения ГРС по данным scRNA-seq.

имеют прямую регуляторную связь. Такие методы, как PPCOR (Kim, 2015), учитывают влияние других генов, вычисляют получастные корреляции. LEAP (Specht, Li, 2017), алгоритм, разработанный специально для анализа данных отдельных клеток, определяет максимальную корреляцию Пирсона между каждой парой генов в окнах с переменным запаздыванием при условии, что клетки расположены в псевдовременном порядке. Поскольку этот тип корреляции несимметричный, он способен реконструировать направленные генные регуляторные сети. В результате тестирования данной программы на транскриптомах 564 отдельных дендритных клеток мыши LEAP выявил несколько тысяч ранее неизвестных связей между генами (Shalek et al., 2014).

Взаимная информация

Информационно-теоретические подходы используют взаимную информацию, которая измеряет снижение энтропии для одной переменной (например, уровня экспрессии одного гена) с учетом значения другой переменной (уровня экспрессии другого гена) (Chan et al., 2017; Qiu et al., 2020; Chang et al., 2024). Чтобы уменьшить ложные срабатывания, возникающие из-за косвенных связей между двумя генами, такие методы, как PIDC (Chan et al., 2017), используют частичное разложение информации (partial information decomposition, PID) для вычисления пропорционального уникального вклада (proportional unique contribution, PUC) для пары генов, который не может быть объяснен экспрессией третьего гена. Поскольку эта связь симметрична, реконструируемые ребра ненаправленны.

Метод PIDC был успешно применен для реконструкции ГРС по транскриптомам отдельных клеток для трех процессов у мыши: дифференцировки мегакариоцитов и эритроцитов от общего предшественника, раннего эмбриогенеза и эмбрионального гематопоэза. Во всех трех примерах PIDC нашел ранее неизвестные связи, эффективно выделил модули генов разных стадий дифференцировки и предполагаемые взаимодействия генов, осуществляющие переход между стадиями. При системной оценке 12 основанных на разных типах моделирования программ построения ГРС метод PIDC был указан как один из самых эффективных (Pratapa et al., 2020).

Алгоритм Scribe (Qiu et al., 2020) использует псевдовремя для вычисления ограниченной направленной информации (restricted directed information, RDI). Эта величина измеряет взаимную информацию между предшествующим уровнем экспрессии гена, кодирующего ТФ, и текущим уровнем экспрессии гена-мишени, обусловленную экспрессией регулятора ранее в псевдовременном ряду. Поскольку взаимная информация между предшествующим и текущим выражением несимметрична, Scribe может выводить направленные ребра. Этот алгоритм применялся как для верификациии существования отдельных связей в разных генных сетях, так и для реконструкции генной сети раннего эмбриогенеза *Caenorhabditis elegans*, где с его помощью была выведена известная иерархия регулирования транскрипции генов.

Третья программа, SINUM, также оценивающая взаимную информацию между любыми двумя генами и определяющая, являются ли они зависимыми или независимыми в конкретной клетке, была апробирована на различных данных и показала высокую эффективность в определении клеточных типов, их маркерных генов и связей между генами (Chang et al., 2024). Также с помощью программы были установлены изменения в ассоциациях генов в ходе дифференцировки эмбриональных стволовых клеток человека в эндодерму.

Регрессия

Генные регуляторные сети можно реконструировать, моделируя экспрессию каждого гена как функцию уровня экспрессии других генов и применяя методы, основанные на регрессии, для решения полученной системы уравнений (Huynh-Thu et al., 2010; Gao et al., 2017; Moerman et al., 2018). Алгоритм GENIE3 использует метод случайного леса, основанный на ансамбле регрессионных деревьев (Huynh-Thu et al., 2010). Вес ребра от ТФ к гену-мишени возникает из значимости ТФ в прогнозировании экспрессии гена-мишени, усредненной по всем деревьям регрессии в случайном лесу. Алгоритм GENIE3 был разработан и широко применялся для работы на данных массового секвенирования транскриптомов клеточных пупуляций. Программное обеспечение GRNBoost2 улучшает масштабируемость GENIE3, особенно с точки зрения эффективной обработки больших наборов данных отдельных клеток (Moerman et al., 2018). Оба инструмента, GENIE3 и GRNBoost2, продемонстрировали свою эффективность для реконструкции ГРС транскриптомов отдельных клеток, показав хорошее пересечение с реальными биологическими взаимодействиями (Kang et al., 2021).

Алгоритм SINCERITIES был создан специально для транскриптомов отдельных клеток, он решает регрессион-

ную модель, основанную на изменениях распределения уровней экспрессии каждого гена во времени и в псевдовремени (Gao et al., 2017). Алгоритмы GRNBoost2 и SINCERITIES были названы среди самых эффективных при системной оценке 12 основанных на разных типах моделирования программ построения ГРС (Pratapa et al., 2020). Тем не менее недавний сравнительный анализ на разных данных по различным показателям выявил, что GRNBoost2 в целом работает лучше, чем SINCERITIES, и более точно идентифицирует хабы в ГРС (Stock et al., 2024).

Байесовские сети

Еще один способ вывода ГРС заключается в моделировании регуляторных взаимодействий в байесовской сети. Алгоритм GRNVBEM работает с данными, представляющими собой временной ряд, т. е. клетки предварительно должны быть ранжированы по псевдовремени (Sanchez-Castillo et al., 2017). Затем он моделирует кратность изменения экспрессии гена между последовательными временными точками как линейную комбинацию экспрессии регуляторов гена в предшествующей точке в байесовской сети. Реконструкция ГРС раннего эмбриогенеза мыши и клеток почек *Danio rerio* с помощью этого метода позволила выделить хабы и сформировать гипотезы о регуляторах дифференцировки.

Метод HBFM основан на коэкспрессии генов с использованием разреженной иерархической модели байесовского фактора для снижения влияния высокой межклеточной изменчивости и шума в наборах данных отдельных клеток на прогнозируемую сеть (Sekula et al., 2020). Результаты работы программы показали значительное совпадение с известными и предсказанными белок-белковыми взаимодействиями из базы данных STRING.

Логические сети

В то время как ранее представленные методы предсказывают сети, описывающие регуляторные эффекты отдельных ТФ, они не учитывают логические правила, управляющие комбинаторным эффектом нескольких ТФ на экспрессию гена-мишени (Nguyen et al., 2021). Например, регуляторные механизмы могут включать активацию гена только в присутствии нескольких определенных ТФ или, в качестве альтернативы, его ингибирование другим ТФ, независимо от дополнительных факторов. Булевы сети способны характеризовать эти комбинации взаимодействий, представляя активное или неактивное состояние гена как двоичную переменную, дискретизированную с использованием порога экспрессии гена, и объединяя эти состояния с помощью операций AND, OR и NOT для объяснения экспрессии всех генов в системе.

Программа SCNS вычисляет логические правила, объясняющие прогрессию экспрессии генов от одной точки псевдовремени к другой (Woodhouse et al., 2018). На основе применения программы к транскриптомам клеток ранних стадий зародыша человека была построена коровая ГРС преимплантационного развития эмбриона. Алгоритм LogicNet использует вероятностную непрерывную логику для построения булевой сети, в которой экспрессия генов моделируется как непрерывная, а не двоичная переменная между 0 и 1 для построения ГРС с направленными и знаковыми ребрами (Malekpour et al., 2020). С помощью LogicNet были построены ГРС раннего эмбриогенеза мыши.

Дифференциальные уравнения

Наличие псевдовременной информации в данных scRNA-seq позволяет моделировать экспрессию генов с помощью обыкновенных дифференциальных уравнений (ОДУ) (Nguyen et al., 2021). Здесь скорость изменения экспрессии гена-мишени является функцией экспрессии гена, кодирующего регулирующий его ТФ. Решая эту систему уравнений, можно определить регуляторные связи на основе веса каждого ТФ в функции, описывающей изменение экспрессии гена. Алгоритм SCODE делает упрощающее предположение о том, что изменения в экспрессии генов можно определить как линейную комбинацию пространств уменьшенных размерностей для эффективного решения менее сложной системы уравнений с использованием линейной регрессии (Matsumoto et al., 2017). В качестве альтернативы GRISLI оценивает скорость, с которой экспрессия каждого гена меняется в соответствии с динамическим процессом в каждой клетке (Aubin-Frankowski, Vert, 2020). Впоследствии он упрощает систему уравнений, основываясь на предположении о том, что искомая ГРС имеет мало регуляторных ребер относительно числа генов в сети, тем самым сводя задачу к проблеме разреженной регрессии.

Ценным свойством GRISLI является то, что он позволяет клеткам следовать множественным траекториям дифференцировки, в то время как большинство методов допускают лишь линейную, неветвящуюся траекторию. Алгоритм DynGENIE3 применяет подход случайного леса GENIE3 для решения системы ОДУ, в которой изменение экспрессии одного гена определяется как потенциально нелинейная комбинация экспрессии других генов (Huynh-Thu, Geurts, 2018).

Другой класс подходов основан на наблюдении о том, что вариации экспрессии генов от клетки к клетке могут возникать из-за стохастической природы молекулярных регуляторных взаимодействий (Nguyen et al., 2021). Модель кусочно-детерминированного марковского процесса (PDMP) задает ОДУ для экспрессии гена как функцию стохастического двухстадийного марковского процесса, указывающего, активирована ли транскрипция гена, а не напрямую как функцию экспрессии регулирующих ТФ (Herbach et al., 2017).

Для каждого гена функция вероятности, представляющая переходы между активным и неактивным состояниями, включает вес каждого потенциального регулятора. Метод PDMP использует оценку максимального правдоподобия для определения этих весов и, таким образом, выводит ребра ГРС. Алгоритм WASABI реализует альтернативную оценку максимального правдоподобия, основанную на концепции, что наблюдаемому увеличению или уменьшению экспрессии гена должно предшествовать изменение количества регулирующего его активность белка в более раннем временном окне (Bonnaffoux et al., 2019). С применением WASABI для реконструкции ГРС дифференцировки эритроцитов у птиц обнаружены

ее необычные свойства: отсутствие хабов, распределенная структура сети и контроль экспрессии большинства генов непосредственно фактором, вызывающим дифференцировку.

Уточнение ГРС, реконструированных по данным scRNA-seq, путем распознавания сайтов связывания ТФ

Несмотря на широкое использование данных scRNA-seq для вывода ГРС, точность реконструкции реального регуляторного механизма на основании их остается неудовлетворительной (Chen, Mar, 2018; Pratapa et al., 2020). Эта проблема возникает потому, что программы для реконструкции ГРС из данных транскриптома основаны на предположении о том, что выявляемые взаимосвязи между уровнями экспрессии ТФ-кодирующего гена и его потенциальных генов-мишеней подразумевают прямое регулирование транскрипции. Однако наблюдаемые ассоциации могут быть вызваны другими биологическими явлениями или даже случайными причинами. Транскриптомные данные не содержат прямой информации о регуляторных событиях (например, о связывании ТФ с регуляторными районами генов). Таким образом, сложно найти различие между прямым и косвенным регулированием, основываясь исключительно на данных scRNA-seq.

Для решения этих проблем и повышения эффективности вывода ГРС необходимо привлекать дополнительные данные, напрямую характеризующие факторы, вовлеченные в регуляцию транскрипции. Например, геномные последовательности, содержащие регуляторные коды, могут быть использованы для определения потенциальных сайтов связывания ТФ. В этом случае наличие мотива, связывающего ТФ и расположенного в регуляторном районе гена-мишени, свидетельствует в пользу прямой регуляции между ТФ и геном-мишенью.

Для такого способа уточнения ГРС, идентифицированных с помощью алгоритма GENIE3, конвейер SCENIC использует базу данных мотивов связывания ТФ (Aibar et al., 2017). Он включает взаимодействия в сеть только в том случае, если мотивы, описывающие сайты связывания ТФ, обогащены в промоторных областях генов-мишеней. Более поздняя версия pySCENIC использует распараллеливание для повышения эффективности (Van de Sande et al., 2020). Конвейер SCENIC показал свою эффективность в определении типов клеток мозга мыши и человека (включая даже те, что были представлены двумя-шестью клетками), а также стадий развития раковых опухолей, которые более трудно различимы, чем клеточные типы (Aibar et al., 2017; Van de Sande et al., 2020). В дополнение для каждого типа клеток и стадий опухоли был установлен специфический набор ТФ, включая ранее неизвестные онкологические маркеры. Роль части из них в прогрессировании опухоли была подтверждена экспериментально в этих же исследованиях.

Интеграция данных scRNA-seq и scATAC-seq для реконструкции ГРС

ДНК в геноме упакована в нуклеосомы, базовые структурные единицы хроматина, которые препятствуют связыванию ТФ с ДНК, предотвращая транскрипцию генов (Parmar, Padinhateeri, 2020). Активация гена возможна, если только его регуляторный район свободен от нуклеосом. Нуклеосомная упаковка ДНК – это регулируемый процесс, она может различаться в зависимости от условий и клеточных типов. Технология scATAC-seq (single cell Assay for Transposase-Accessible Chromatin using sequencing) позволяет в отдельных клетках идентифицировать открытые участки хроматина, т. е. доступные для связывания ТФ регуляторные районы ДНК (Buenrostro et al., 2015). Таким образом, данные scATAC-seq могут способствовать более точной реконструкции прямой регуляторной связи между ТФ и их генами-мишенями в ГРС.

Было показано, что интеграция данных RNA-seq и ATAC-seq (или других эпигеномных данных) для многоклеточных образцов значительно повышает точность построения ГРС (Qin et al., 2014; Wang et al., 2015; Ackermann et al., 2016). Эта методология также применима к данным секвенирования отдельных клеток. Однако из-за специфичности транскриптомных и эпигеномных профилей по типу клеток и состоянию комбинирование данных RNA-seq с данными ATAC-seq или ChIP-seq, как правило, требует, чтобы оба набора данных были получены из клеток одного и того же типа в идентичных условиях.

На сегодняшний день разработаны технологии, позволяющие проводить одновременное секвенирование транскриптома и эпигенома для одной и той же клетки (Angermueller et al., 2016; Hu et al., 2016; Chen et al., 2019). Альтернативой является интегрированное исследование данных scRNA-seq и scATAC-seq, полученных с различных биологических образцов одинаковой природы. В этом случае дополнительную проблему для реконструкции ГРС составляет поиск соответствия между кластерами клеток, представляющих один и тот же тип, условие или состояние, по двум видам данных секвенирования. Для решения этой задачи разрабатываются методы так называемой диагональной интеграции (Argelaguet et al., 2021).

Поскольку по сравнению с другими методами профилирования эпигенома отдельных клеток scATAC-seq используется наиболее часто, было разработано несколько биоинформатических инструментов для объединения данных scRNA-seq и scATAC-seq при реконструкции ГРС (Loers, Vermeirssen, 2024). ГРС, реконструируемые на основе этих данных, получили специальное название «энхансерные ГРС» (эГРС). Метод STREAM реконструирует эГРС на базе совместно профилированных scRNA-seq и scATAC-seq данных с помощью модели задачи леса Штейнера, гибридного бикластерного конвейера и субмодульной оптимизации для вывода генной сети (Li et al., 2024). Метод STREAM был апробирован на данных отдельных клеток человека из органов с патологиями (болезни Альцгеймера и лимфоцитарной лимфомы) и показал свою эффективность в реконструкции связей ТФ–открытый сайт связывания–ген вдоль псевдовременной траектории и выявлении специфичных для этих болезней регуляций транскрипции генов.

Существуют и программы, которые используют результаты предварительного раздельного анализа данных scRNA-seq и scATAC-seq. Например, scMTNI на вход берет схему дифференцировки клеток, результаты scRNA-seq и

априорные сети на основе данных scATAC-seq для каждого типа клеток (Zhang et al., 2023). Применение scMTNI к данным scRNA-seq и scATAC-seq по репрограммированию клеток у мышей и дифференцировке гемопоэтических клеток у человека позволило построить эГРС как для линейных, так и для разветвленных путей дифференцировки и определить регуляторы и другие компоненты эГРС, специфичные для их разных этапов.

Заключение

Выявление взаимоотношений генов в процессе регуляции их экспрессии – ключ к пониманию механизмов, обеспечивающих реализацию генетической информации в тот или иной фенотипический признак. Построение ГРС по омиксным данным отдельных клеток предоставляет уникальную возможность системно исследовать механизмы клеточной дифференцировки, поскольку теоретически позволяет воссоздавать регуляторные генные сети для отдельных типов клеток и даже на отдельных стадиях их развития. На сегодняшний день разработан целый ряд методов для реконструкции таких ГРС, многие из них доступны пользователям в виде компьютерных программ. Несмотря на перспективность данного подхода, его потенциал далеко не полностью реализован. Не все доступные методы удобны для работы и для интерпретации их результатов.

Актуальной проблемой является также развитие методов верификации реконструируемых ГРС. Вероятно, поэтому использование данных моделей в конкретных биологических исследованиях пока ограничено, и можно говорить лишь об отдельных примерах успешного применения клеточных ГРС для решения биологических задач. Дальнейшее развитие технологий молекулярно-генетического исследования отдельных клеток и компьютерных методов анализа генерируемых ими данных (в частности, с целью реконструкции ГРС и их анализа) позволит существенно сократить разрыв между нашими знаниями о молекулярных детерминантах признаков (в том числе на клеточном уровне) и транскрипционных каскадах, запускаемых под воздействием внешних или внутренних стимулов. Прорывные открытия, сделанные с помощью ГРС, реконструированных на основании омиксных данных для отдельных клеток, возможно, ждут нас в будущем.

Список литературы / References

Колчанов Н.А., Игнатьева Е.В., Подколодная О.А., Лихошвай В.А., Матушкин Ю.Г. Генные сети. *Вавиловский журнал генетики и селекции*. 2013;17(4/2):833-850
[Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2013;17(4/2):833-850 (in Russian)]

Ackermann A.M., Wang Z., Schug J., Naji A., Kaestner K.H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab*. 2016;5(3):233-244. doi 10.1016/j.molmet.2016.01.002

Aibar S., González-Blas C.B., Moerman T., Huynh-Thu V.A., Imrichova H., Hulselmans G., Rambow F., Marine J., Geurts P., Aerts J., Van Den Oord J., Atak Z.K., Wouters J., Aerts S. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods*. 2017;14(11):1083-1086. doi 10.1038/nmeth.4463

Altay G. Empirically determining the sample size for large-scale gene network inference algorithms. *IET Syst. Biol*. 2012;6(2):35-43. doi 10.1049/iet-syb.2010.0091

Angermueller C., Clark S.J., Lee H.J., Macaulay I.C., Teng M.J., Hu T.X., Krueger F., Smallwood S.A., Ponting C.P., Voet T., Kelsey G., Stegle O., Reik W. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods*. 2016;13(3):229-232. doi 10.1038/nmeth.3728

Argelaguet R., Cuomo A.S.E., Stegle O., Marioni J.C. Computational principles and challenges in single-cell data integration. *Nat. Biotechnol*. 2021;39(10):1202-1215. doi 10.1038/s41587-021-00895-7

Aubin-Frankowski P., Vert J. Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference. *Bioinformatics*. 2020;36(18):4774-4780. doi 10.1093/bioinformatics/btaa576

Badia-i-Mompel P., Wessels L., Müller-Dott S., Trimbour R., Flores R.O.R., Argelaguet R., Saez-Rodriguez J. Gene regulatory network inference in the era of single-cell multi-omics. *Nat. Rev. Genet*. 2023;24(11):739-754. doi 10.1038/s41576-023-00618-5

Blencowe M., Arneson D., Ding J., Chen Y.W., Saleem Z., Yang X. Network modeling of single-cell omics data: challenges, opportunities, and progresses. *Emerg. Top. Life Sci*. 2019;3(4):379-398. doi 10.1042/ETLS20180176

Bonnaffoux A., Herbach U., Richard A., Guillemin A., Gonin-Giraud S., Gros P., Gandrillon O. WASABI: a dynamic iterative framework for gene regulatory network inference. *BMC Bioinformatics*. 2019;20(1):220. doi 10.1186/s12859-019-2798-1

Buenrostro J.D., Wu B., Litzenburger U.M., Ruff D., Gonzales M.L., Snyder M.P., Chang H.Y., Greenleaf W.J. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. 2015;523(7561):486-490. doi 10.1038/nature14590

Chan T.E., Stumpf M.P., Babtie A.C. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst*. 2017;5(3):251-267.e3. doi 10.1016/j.cels.2017.08.014

Chang L., Hao T., Wang W., Lin C. Inference of single-cell network using mutual information for scRNA-seq data analysis. *BMC Bioinformatics*. 2024;25(S2):292. doi 10.1186/s12859-024-05895-3

Chen S., Mar J.C. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics*. 2018;19(1):232. doi 10.1186/s12859-018-2217-z

Chen S., Lake B.B., Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol*. 2019;37(12):1452-1457. doi 10.1038/s41587-019-0290-0

Gao N.P., Ud-Dean S.M.M., Gandrillon O., Gunawan R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics*. 2017;34(2):258-266. doi 10.1093/bioinformatics/btx575

Herbach U., Bonnaffoux A., Espinasse T., Gandrillon O. Inferring gene regulatory networks from single-cell data: a mechanistic approach. *BMC Syst. Biol*. 2017;11(1):105. doi 10.1186/s12918-017-0487-0

Hong S., Chen X., Jin L., Xiong M. Canonical correlation analysis for RNA-seq co-expression networks. *Nucleic Acids Res*. 2013;41(8):e95. doi 10.1093/nar/gkt145

Hou W., Ji Z., Chen Z., Wherry E.J., Hicks S.C., Ji H. A statistical framework for differential pseudotime analysis with multiple single-cell RNA-seq samples. *Nat. Commun*. 2023;14(1):7286. doi 10.1038/s41467-023-42841-y

Hu Y., Huang K., An Q., Du G., Hu G., Xue J., Zhu X., Wang C., Xue Z., Fan G. Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol*. 2016;17(1):88. doi 10.1186/s13059-016-0950-z

Huynh-Thu V.A., Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Sci. Rep*. 2018;8(1):3384. doi 10.1038/s41598-018-21715-0

Huynh-Thu V.A., Sanguinetti G. Gene regulatory network inference: An introductory survey. *Methods Mol. Biol.* 2019;1883:1-23. doi 10.1007/978-1-4939-8882-2_1

Huynh-Thu V.A., Irrthum A., Wehenkel L., Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*. 2010;5(9):e12776. doi 10.1371/journal.pone.0012776

Isbel L., Grand R.S., Schübeler D. Generating specificity in genome regulation through transcription factor sensitivity to chromatin. *Nat. Rev. Genet*. 2022;23(12):728-740. doi 10.1038/s41576-022-00512-6

Kang Y., Thieffry D., Cantini L. Evaluating the reproducibility of single-cell gene regulatory network inference algorithms. *Front. Genet*. 2021;12:617282. doi 10.3389/fgene.2021.617282

Kim S. ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods*. 2015;22(6):665-674. doi 10.5351/CSAM.2015.22.6.665

Li Y., Ma A., Wang Y., Guo Q., Wang C., Fu H., Liu B., Ma Q. Enhancer-driven gene regulatory networks inference from single-cell RNA-seq and ATAC-seq data. *Brief. Bioinform*. 2024;25(5):bbae369. doi 10.1093/bib/bbae369

Loers J.U., Vermeirssen V. A single-cell multimodal view on gene regulatory network inference from transcriptomics and chromatin accessibility data. *Brief. Bioinform*. 2024;25(5):bbae382. doi 10.1093/bib/bbae382

Luecken M.D., Theis F.J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol*. 2019;15(6):e8746. doi 10.15252/msb.20188746

Malekpour S.A., Alizad-Rahvar A.R., Sadeghi M. LogicNet: probabilistic continuous logics in reconstructing gene regulatory networks. *BMC Bioinformatics*. 2020;21(1):318. doi 10.1186/s12859-020-03651-x

Matsumoto H., Kiryu H., Furusawa C., Ko M.S.H., Ko S.B.H., Gouda N., Hayashi T., Nikaido I. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics*. 2017;33(15):2314-2321. doi 10.1093/bioinformatics/btx194

Mercatelli D., Scalambra L., Triboli L., Ray F., Giorgi F.M. Gene regulatory network inference resources: A practical overview. *Biochim. Biophys. Acta Gene Regul. Mech*. 2020;1863(6):194430. doi 10.1016/j.bbagrm.2019.194430

Moerman T., Santos S.A., González-Blas C.B., Simm J., Moreau Y., Aerts J., Aerts S. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*. 2018;35(12):2159-2161. doi 10.1093/bioinformatics/bty916

Nguyen H., Shrestha S., Tran D., Shafi A., Draghici S., Nguyen T. A comprehensive survey of tools and software for active subnetwork identification. *Front. Genet*. 2019;10:155. doi 10.3389/fgene.2019.00155

Nguyen H., Tran D., Tran B., Pehlivan B., Nguyen T. A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data. *Brief. Bioinform*. 2021;22(3):bbaa190. doi 10.1093/bib/bbaa190

Parmar J.J., Padinhateeri R. Nucleosome positioning and chromatin organization. *Curr. Opin. Struct. Biol*. 2020;64:111-118. doi 10.1016/j.sbi.2020.06.021

Pratapa A., Jalihal A.P., Law J.N., Bharadwaj A., Murali T.M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods*. 2020;17(2):147-154. doi 10.1038/s41592-019-0690-6

Qin J., Hu Y., Xu F., Yalamanchili H.K., Wang J. Inferring gene regulatory networks by integrating ChIP-seq/chip and transcriptome data via LASSO-type regularization methods. *Methods*. 2014;67(3):294-303. doi 10.1016/j.ymeth.2014.03.006

Qiu X., Rahimzamani A., Wang L., Ren B., Mao Q., Durham T., McFaline-Figueroa J.L., Saunders L., Trapnell C., Kannan S. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell Syst*. 2020;10(3):265-274. doi 10.1016/j.cels.2020.02.003

Saelens W., Cannoodt R., Todorov H., Saeys Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol*. 2019;37(5):547-554. doi 10.1038/s41587-019-0071-9

Sanchez-Castillo M., Blanco D., Tienda-Luna I.M., Carrion M.C., Huang Y. A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics*. 2017;34(6):964-970. doi 10.1093/bioinformatics/btx605

Sekula M., Gaskins J., Datta S. A sparse Bayesian factor model for the construction of gene co-expression networks from single-cell RNA sequencing count data. *BMC Bioinformatics*. 2020;21(1):361. doi 10.1186/s12859-020-03707-y

Shalek A.K., Satija R., Shuga J., Trombetta J.J., Gennert D., Lu D., Chen P., Gertner R.S., Gaublomme J.T., Yosef N., Schwartz S., Fowler B., Weaver S., Wang J., Wang X., Ding R., Raychowdhury R., Friedman N., Hacohen N., Park H., May A.P., Regev A. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*. 2014;510(7505):363-369. doi 10.1038/nature13437

Sönmezer C., Kleinendorst R., Imanci D., Barzaghi G., Villacorta L., Schübeler D., Benes V., Molina N., Krebs A.R. Molecular co-occupancy identifies transcription factor binding cooperativity *in vivo*. *Mol. Cell*. 2020;81(2):255-267. doi 10.1016/j.molcel.2020.11.015

Specht A.T., Li J. LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. *Bioinformatics*. 2017;33(5):764-766. doi 10.1093/bioinformatics/btw729

Stock M., Popp N., Fiorentino J., Scialdone A. Topological benchmarking of algorithms to infer gene regulatory networks from single-cell RNA-seq data. *Bioinformatics.* 2024;40(5):btae267. doi 10.1093/bioinformatics/btae267

Tang F., Barbacioru C., Wang Y., Nordman E., Lee C., Xu N., Wang X., Bodeau J., Tuch B.B., Siddiqui A., Lao K., Surani M.A. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*. 2009;6(5):377-382. doi 10.1038/nmeth.1315

Tieri P., Castiglione F. Modeling macrophage differentiation and cellular dynamics. In: Wolkenhauer O. (Ed.). Systems Medicine. Integrative, Qualitative and Computational Approaches. Academic Press, 2021;511-520. doi 10.1016/B978-0-12-801238-3.11644-7

Van de Sande B., Flerin C., Davie K., De Waegeneer M., Hulselmans G., Aibar S., Seurinck R., Saelens W., Cannoodt R., Rouchon Q., Verbeiren T., De Maeyer D., Reumers J., Saeys Y., Aerts S. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc*. 2020;15(7):2247-2276. doi 10.1038/s41596-020-0336-2

Wagner A., Regev A., Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol*. 2016;34(11):1145-1160. doi 10.1038/nbt.3711

Wang P., Qin J., Qin Y., Zhu Y., Wang L.Y., Li M.J., Zhang M.Q., Wang J. ChIP-Array 2: integrating multiple omics data to construct gene regulatory networks. *Nucleic Acids Res*. 2015;43(W1):264-269. doi 10.1093/nar/gkv398

Woodhouse S., Piterman N., Wintersteiger C.M., Göttgens B., Fisher J. SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. *BMC Syst. Biol.* 2018;12(1):59. doi 10.1186/s12918-018-0581-y

Zhang S., Pyne S., Pietrzak S., Halberg S., McCalla S.G., Siahpirani A.F., Sridharan R., Roy S. Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. *Nat. Commun*. 2023;14(1):3064. doi 10.1038/s41467-023-38637-9

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 28.10.2024. После доработки 21.11.2024. Принята к публикации 22.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):982-992

doi 10.18699/vjgb-24-105

Сопоставление показателей мозговой активности у китайских и российских студентов в условиях распознавания информации, отнесенной к себе и другим людям

Ц. Сы[1], Ц. Тянь[1], В.А. Савостьянов [1, 3], Д.А. Лебедкин [1, 4], А.В. Бочаров [1, 3], А.Н. Савостьянов [1, 2, 3]

[1] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[3] Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия
[4] Национальный исследовательский Томский государственный университет, Томск, Россия
a.savostianov@g.nsu.ru

Аннотация. Нейровычислительные технологии – область междисциплинарных исследований и разработок, которая находит широкое применение в современной цифровой медицине. Одна из задач нейровычислительных технологий состоит в создании методик изучения мозговой активности человека в условиях социально-ориентированной деятельности при помощи современных информационных подходов. Цель предлагаемого исследования – разработать методику сбора и обработки психофизиологических данных, позволяющую изучать функциональные состояния головного мозга человека, ассоциированные с отнесением внешней информации к самому субъекту или другим людям. Под самоотнесением (самореференцией) понимается субъективная оценка человеком поступающей из внешней среды информации как имеющей отношение к нему самому. Отнесение информации к другим людям или неодушевленным объектам – это оценка информации как сообщения о ком-то другом или о вещах. В современной нейрофизиологии сложились два подхода к исследованию самореференции: 1 – регистрация мозговой активности в условиях покоя с последующим опросом участника на предмет выявления самоотнесенных мыслей; 2 – регистрация мозговой активности, вызванной самоотнесенными стимулами. В представленной работе была апробирована технология, сочетающая регистрацию и анализ ЭЭГ с просмотром видеозаписей изображений лица самого испытуемого или незнакомого ему человека. Новизна нашего подхода состоит в использовании видеозаписей человеческого лица, полученных на первом этапе обследования, для индукции состояний покоя, ассоциированных с распознаванием информации о разных субъектах, на более поздних этапах обследования. Нами был разработан программно-аппаратный модуль, т. е. комплект связанных друг с другом программ и процедур их применения, состоящий из блоков, позволяющих проводить полный цикл регистрации и обработки психологических и нейрофизиологических данных. При помощи этого модуля показатели электрической активности головного мозга (ЭЭГ), отражающие индивидуальные особенности распознавания информации, отнесенной к самому себе и другим людям, были сопоставлены между группами из 30 китайских (14 мужчин и 16 женщин, средний возраст 23.2 ± 0.4 года) и 32 российских (15 мужчин, 17 женщин, средний возраст 22.1 ± 0.4 года) участников. Мы проверили гипотезу, что различия в мозговой активности в интервалах функционального покоя между китайскими и российскими участниками зависят от их психологических различий в показателях коллективизма. Было выявлено, что функциональная активность мозга зависит от субъектной отнесенности лицевого видео, которое участники просматривали между интервалами покоя. Межнациональные различия наблюдались в активности переднего и заднего хабов дефолт-системы и зависели от субъектной отнесенности информации. У китайских, но не у российских участников выявлены достоверные положительные корреляции между уровнем коллективизма и спектральной плотностью в переднем хабе дефолт-системы во всех экспериментальных условиях для широкого ряда частотных диапазонов. Разработанный программно-аппаратный модуль включен в интегрированную цифровую платформу для проведения исследований в области системной биологии и цифровой медицины.

Ключевые слова: нейровычислительные технологии; программно-аппаратный модуль; методы обработки данных; самоотнесенные процессы; ЭЭГ покоя; дефолт-система мозга; межнациональные различия; коллективизм.

Для цитирования: Сы Ц., Тянь Ц., Савостьянов В.А., Лебедкин Д.А., Бочаров А.В., Савостьянов А.Н. Сопоставление показателей мозговой активности у китайских и российских студентов в условиях распознавания информации, отнесенной к себе и другим людям. *Вавиловский журнал генетики и селекции*. 2024;28(8):982-992. doi 10.18699/vjgb-24-105

Финансирование. Разработка программно-аппаратного модуля проведена в рамках бюджетного проекта ИЦиГ СО РАН № FWNR-2022-0020. Часть исследования, касающаяся подготовки психологических тестов и подбора экспериментальных групп, выполнена при финансовой поддержке Программы развития Томского государственного университета (Приоритет-2030).

© Сы Ц., Тянь Ц., Савостьянов В.А., Лебедкин Д.А., Бочаров А.В., Савостьянов А.Н., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

Comparison of brain activity metrics in Chinese and Russian students while perceiving information referencing self or others

Q. Si[1], J. Tian[1], V.A. Savostyanov [1, 3], D.A. Lebedkin [1, 4], A.V. Bocharov [1, 3], A.N. Savostyanov [1, 2, 3]

[1] Novosibirsk State University, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[3] Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia
[4] Tomsk State University, Tomsk, Russia

a.savostianov@g.nsu.ru

Abstract. Neurocomputing technology is a field of interdisciplinary research and development widely applied in modern digital medicine. One of the problems of neuroimaging technology is the creation of methods for studying human brain activity in socially oriented conditions by using modern information approaches. The aim of this study is to develop a methodology for collecting and processing psychophysiological data, which makes it possible to estimate the functional states of the human brain associated with the attribution of external information to oneself or other people. Self-reference is a person's subjective assessment of information coming from the external environment as related to himself/herself. Assigning information to other people or inanimate objects is evaluating information as a message about someone else or about things. In modern neurophysiology, two approaches to the study of self-referential processing have been developed: (1) recording brain activity at rest, then questioning the participant for self-reported thoughts; (2) recording brain activity induced by self-assigned stimuli. In the presented paper, a technology was tested that combines registration and analysis of EEG with viewing facial video recordings. The novelty of our approach is the use of video recordings obtained in the first stage of the survey to induce resting states associated with recognition of information about different subjects in later stages of the survey. We have developed a software and hardware module, i. e. a set of related programs and procedures for their application consisting of blocks that allow for a full cycle of registration and processing of psychological and neurophysiological data. Using this module, brain electrical activity (EEG) indicators reflecting individual characteristics of recognition of information related to oneself and other people were compared between groups of 30 Chinese (14 men and 16 women, average age 23.2 ± 0.4 years) and 32 Russian (15 men, 17 women, average age 22.1 ± 0.4 years) participants. We tested the hypothesis that differences in brain activity in functional rest intervals between Chinese and Russian participants depend on their psychological differences in collectivism scores. It was revealed that brain functional activity depends on the subject relevance of the facial video that the participants viewed between resting-state intervals. Interethnic differences were observed in the activity of the anterior and parietal hubs of the default-mode network and depended on the subject attribution of information. In Chinese, but not Russian, participants significant positive correlations were revealed between the level of collectivism and spectral density in the anterior hub of the default-mode network in all experimental conditions for a wide range of frequencies. The developed software and hardware module is included in an integrated digital platform for conducting research in the field of systems biology and digital medicine.

Key words: neurocomputing technologies; hardware-software module; data processing methods; self-referential processes; resting-state EEG; default-mode network; interethnic differences; collectivism.

For citation: Si Q., Tian J., Savostyanov V.A., Lebedkin D.A., Bocharov A.V., Savostyanov A.N. Comparison of brain activity metrics in Chinese and Russian students while perceiving information referencing self or others. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):982-992. doi 10.18699/vjgb-24-105

Введение

Нейровычислительные технологии – это техническое направление по разработке методик для сбора и компьютерного анализа данных нейрофизиологических обследований, которое находит широкое применение в цифровой медицине для создания новых подходов к диагностике и терапии заболеваний. Целью нейровычислительных технологий является разработка программ и устройств для получения информации об анатомо-функциональной организации нервной системы в норме и при патологиях.

Теория референции предложена в работах логиков и лингвистов первой половины XX в. (для обзора см. Яковлева, 2011). Информационная референция – это оценка входящей информации как имеющей отношение к какому-то конкретному объекту или субъекту. Под термином «самоотнесение» (или «самореференция») понимается оценка какого-либо события как связанного с самим субъектом, воспринимающим информацию об этом событии (Northoff et al., 2006; Neff, McGehee, 2010). Термин «самоотнесение» принципиально отличается по своему содержанию от терминов «рефлексия» (размышление о себе) и «самоконтроль» (управление своими действиями), так как относится не к управлению поведением или самооценке, а к области анализа входящей из внешней среды информации как имеющей или не имеющей отношение к самому субъекту. К настоящему времени сформировалось два принципиально различающихся подхода к исследованию нейрофизиологических маркеров субъектного отнесения информации. В рамках первого подхода мозговая активность, оцениваемая при помощи электроэнцефалограммы (ЭЭГ), магнитоэнцефалограммы (МЭГ) или функциональной магнитно-резонансной томографии (фМРТ), регистрируется в условиях функционального покоя, т. е. без выполнения экспериментальных заданий (Knyazev et al.,

2012, 2016). После завершения регистрации мозговой активности участники проходят опрос на предмет выяснения их фокусирования на самоотнесенных событиях. Другой подход состоит в представлении участникам нескольких групп стимулов с однозначной отнесенностью к нему самому, знакомым и незнакомым людям или же к неодушевленным объектам (Quevedo et al., 2018; Knyazev et al., 2020, 2024).

Целью нашего исследования является разработка новой экспериментальной модели, сочетающей в себе оба описанных выше подхода к изучению самореферентной активности мозга человека, т. е. тех нейрофизиологических процессов, которые лежат в основе самоотнесения информации. В этой модели участнику предлагается внешняя информация (просмотр видеоизображений) о нем самом или о другом человеке в сравнении с наблюдением неодушевленного объекта. В интервалах между просмотром видеоизображений участник закрывает глаза и какое-то время не получает внешней стимуляции. Предлагаемая технология включает сбор данных, основанный на комбинировании регистрации ЭЭГ с видеозаписью человеческих лиц (Савостьянов и др., 2022), предобработку данных ЭЭГ для очистки целевого сигнала от нерелевантного шума (Delorme, Makeig, 2004), локализацию источников мозговых сигналов на поверхности коры головного мозга и поиск статистических взаимосвязей между нейрофизиологической активностью и психологическими особенностями участников обследования (Pascual-Margui, 2002). Кроме того, наша методика включает проведение психологического тестирования для выявления личностных особенностей участников и симптоматики выраженности у них депрессии. В рамках предлагаемой статьи мы апробируем создаваемую технологию для поиска нейрофизиологических различий, обусловленных разным отношением к самому себе в группах российских и китайских студентов. Мы предполагаем, что россияне более склонны к индивидуалистическому определению собственной личности, тогда как для китайцев более характерны коллективистские способы самоопределения. Разработанные методики и компьютерные программы для сбора и обработки данных, а также сами собранные в этом исследовании данные включены в качестве одного из модулей в интегрированную цифровую платформу «Биоинформатика и системная компьютерная биология», разрабатываемую в ИЦиГ СО РАН.

Материалы и методы

Программный модуль для сбора и обработки данных. Нами был создан программный модуль для сбора и обработки данных, который включен в интегрированную цифровую платформу «Биоинформатика и системная компьютерная биология», разрабатываемую в ИЦиГ СО РАН. Модуль состоит как из программных продуктов, разработанных сотрудниками ИЦиГ СО РАН, так и из блоков, взятых нами из открытых источников. В совокупности все блоки модуля позволяют осуществить полный цикл по сбору и обработке психологических и нейрофизиологических данных, начиная от предварительного опроса участников для получения их согласия на прохождение обследования и заканчивая статистической обработкой полученных результатов. Перечень программ, входящих в состав модуля, представлен в табл. 1.

Испытуемые. Для исследования были приглашены 30 студентов и аспирантов из Китая (14 мужчин и 16 женщин, средний возраст 23.2 ± 0.4 года) и 32 российских студента и аспиранта (15 мужчин, 17 женщин, средний возраст 22.1 ± 0.4 года), обучающихся на базе Новосибирского государственного университета. Перед началом эксперимента все участники заполнили анкету, которая содержала вопросы о наличии у них неврологических или психиатрических заболеваний и употреблении алкоголя или психоактивных веществ. Кроме того, все участники давали информированное согласие на прохождение экспериментального обследования в соответствии с Хельсинкской декларацией по биомедицинской этике. Протокол эксперимента был одобрен этическим комитетом НИИ Нейронаук и медицины.

Психологическое тестирование. Участники заполняли психологические опросники на личностную и ситуативную тревожность (STAI: State-Trait Anxiety Inventory, Spielberger et al., 1970; русскоязычная адаптация Ханин, 1976), опросник для оценки выраженности депрессивной симптоматики (BDI: Beck's depression Inventory, Beck et al., 1996), тест «Коллективная и независимая Я-концепция» (SCS: Self Construal Scale, Singelis, 1994), опросник на аффилиацию со своей семьей (RISC: Relational-interdependent self-construal, Cross et al., 2000). Тестирование проводилось при помощи специального интернет-приложения, разработанного на платформе Yandex. Российские участники заполняли опросники на русском языке, китайские участники – на китайском языке.

Экспериментальный метод и этапы обработки данных представлены в виде блок-схемы на рис. 1. ЭЭГ записывалась в звуко- и светоизолированной камере. В ходе эксперимента выполнялись три условия. При первом условии ЭЭГ регистрировалась в течение 12 минут без функциональной нагрузки (3 интервала по 2 минуты с закрытыми глазами и 3 интервала по 2 минуты с открытыми глазами). В интервалах, когда испытуемый открывал глаза, он видел черный экран монитора компьютера. В этот период у испытуемого совместно с ЭЭГ записывалось видеоизображение лица в течение всех 12 минут. Второе и третье условие отличались от первого тем, что при втором условии при открытых глазах испытуемый видел видеозапись собственного лица, полученную при выполнении первого фонового задания, а при третьем условии ему/ей предъявлялась видеозапись лица незнакомого человека (всегда мужчины – для испытуемого мужчины и женщины – для испытуемой женщины). Порядок второго и третьего задания менялся случайно. Примерно для половины испытуемых вначале шло второе задание, а потом третье, для другой половины – наоборот.

Регистрация ЭЭГ. ЭЭГ регистрировалась при помощи 130-канального усилителя NVX-132, Россия. 128 каналов ЭЭГ были расположены по международной схеме 5–5 % с референтным электродом Cz, заземляющим электродом AFz, добавочно – каналы ЭОГ, ЭКГ, полоса пропускания 0.1–100 Гц, частота дискретизации сигнала 1000 Гц. Регистрация ЭЭГ проводилась с помощью программы NeoReg-recoder.

Таблица 1. Перечень программно-аппаратных блоков, включенных в модуль для регистрации и обработки нейрофизиологических данных

Номер блока в составе модуля	Название программного продукта и его разработчик	Оборудование, необходимое для работы программы	Назначение блока
1	Специальная Интернет-форма, реализованная на платформе Yandex сотрудниками ИЦиГ СО РАН	Цифровое мобильное устройство с подключением к интернету	Проведение дистанционного психологического тестирования участников для оценки их личностных особенностей и выраженности депрессивной симптоматики
2	Программа NeoRec, Медицинские компьютерные системы, https://mks.ru/	Усилитель биоэлектрических сигналов NVX-132	Регистрация биоэлектрической активности мозга в условиях функционального покоя
3	Open Broadcaster Software, OBS Studio, https://obsproject.com/	Видеокамера, подключенная к регистрирующему компьютеру	Регистрация видеозаписей человеческого лица
4	Программа для разметки записей ЭЭГ на основе лицевого видео. Реализована сотрудниками ИЦиГ СО РАН на платформе Inquisit, https://www.millisecond.com/	Усилитель биоэлектрических сигналов NVX-132, Steam Tracker для синхронизации меток событий	Предъявление участнику обследования записей лицевого видео с одновременной регистрацией и разметкой ЭЭГ на участки покоя и стимуляции
5	EEGlab_toolbox, Swartz Center for Computational Neuroscience, https://sccn.ucsd.edu/eeglab/index.php	Персональный компьютер	Предобработка записей ЭЭГ, включая частотную фильтрацию, ре-референс сигналов, анализ независимых компонент и удаление внемозгового шума из ЭЭГ-сигнала
6	eLoreta: low resolution brain electromagnetic tomography, The KEY Institute for Brain-Mind Research, https://www.uzh.ch/keyinst/loreta.htm	Персональный компьютер	Вычисление спектральной плотности в разных частотных диапазонах
7			Локализация источников мозговой активности на поверхности коры головного мозга
8			Проведение регрессионного и корреляционного анализов для поиска ассоциаций между психологическими особенностями участников и их нейрофизиологическими характеристиками
9	IBM SPSS software, IBM, https://www.ibm.com/spss	Персональный компьютер	Статистическая обработка полученных результатов

Предобработка ЭЭГ. Процедура re-reference (т. е. вычисление значений усредненного референта и вычитание этих значений из записи ЭЭГ) была проведена для удаления артефактов тонического напряжения мышц скальпа. Глазодвигательные и иные артефакты были удалены из ЭЭГ при помощи анализа независимых компонент, для чего использовался программный пакет EEGlab v 14_1_2b, предназначенный для среды MatLab (ICA, Delorme, Makeig, 2004). ICA (independent component analysis) – это широко применяемая технология анализа сигналов, позволяющая, в частности, отделить целевой сигнал от нерелевантного шума. Затем записи ЭЭГ разделялись на периоды, когда участник сидел с закрытыми глазами, и периоды, когда глаза были открыты. Дальнейший анализ проводился только для тех интервалов ЭЭГ, которые записывались при закрытых глазах, но были окружены периодами наблюдения соответствующих стимулов. После извлечения этих фрагментов ЭЭГ они разделялись на двухсекундные временные интервалы.

Локализация источников активности на поверхности коры головного мозга. Дальнейший анализ проводился при помощи программного пакета eLoreta (Pascual-Margui, 2002). eLoreta – математическая модель и основанный на этой модели программный продукт, направленный на решение обратной задачи ЭЭГ, т. е. на реконструкцию функциональных процессов в головном мозге на основе компьютерного анализа распределения электрических сигналов на поверхности головы. eLoreta позволяет локализовать источники мозговой активности на основе интерполяции данных по многим ЭЭГ-электродам.

Для каждого двухсекундного интервала вычислялись значения спектральной плотности в частотных диапазонах дельта (2–4 Гц), тета (4–8 Гц), альфа-1 (8–10 Гц), альфа-2 (10–12 Гц), бета-1 (12–16 Гц), бета-2 (16–20 Гц), бета-3 (20–25 Гц) и гамма (25–35 Гц) ритмов. Затем для каждого участника вычислялся показатель суммарного спектра на всем интервале ЭЭГ-тестирования отдельно для каждого из трех экспериментальных условий (у каждого участника

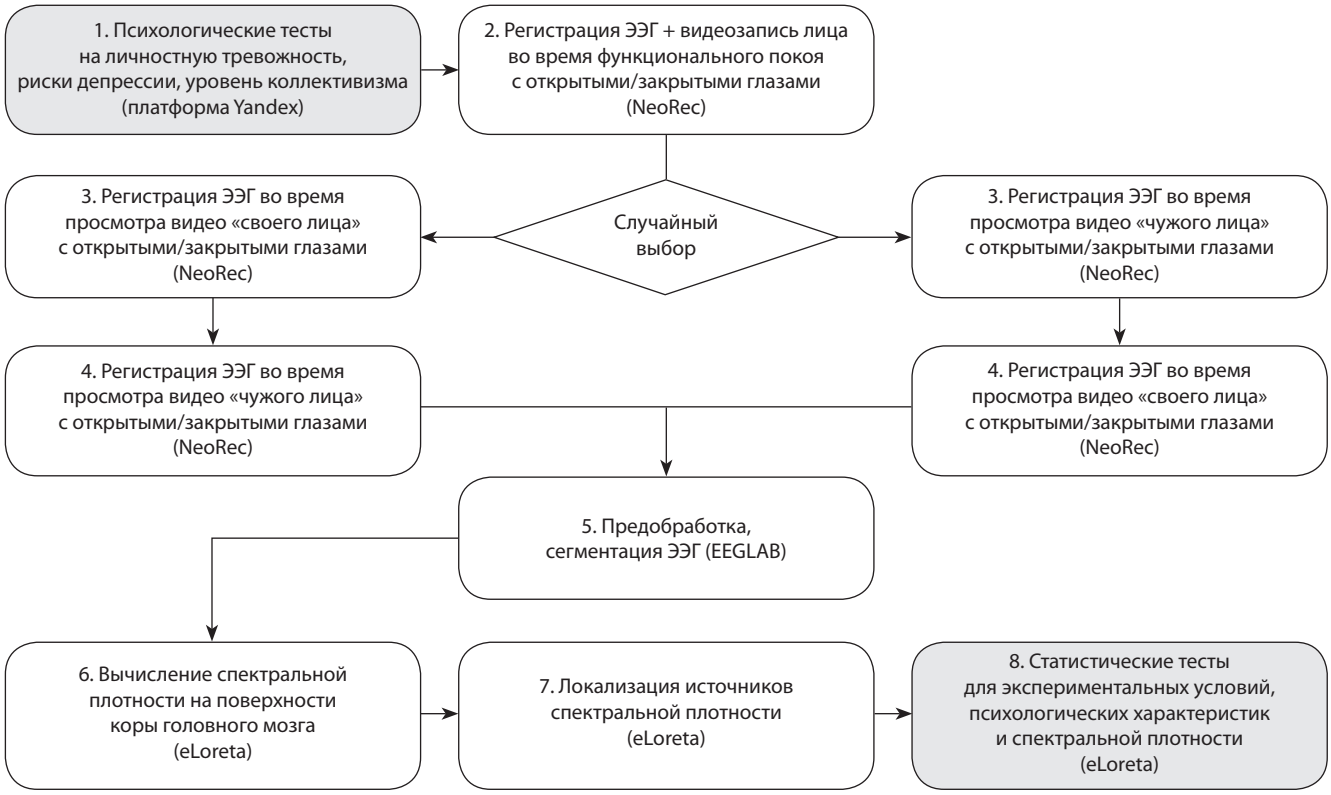


Рис. 1. Блок-схема этапов сбора и обработки данных с указанием компьютерных программ, использованных в нашем исследовании.

для этого использовалось от 150 до 170 двухсекундных интервалов). Спектры считались независимо для каждого из 128 каналов ЭЭГ, включенных в обработку данных. Анализ сравнений спектральной плотности между разными условиями («пустой экран», «свое лицо» и «чужое лицо») на уровне источников был выполнен в программе eLoreta. Отрезок записи ЭЭГ с частотой дискретизации 1000 Гц длительностью 3000 мс после появления блока был использован для расчета спектральной плотности источников в программе eLoreta (Pascual-Margui, 2002).

Статистический анализ результатов. Статистическая обработка результатов психологического тестирования выполнена в программе IBM SPSS. Сравнения проводились с помощью однонаправленного ANOVA с психологическими характеристиками в качестве независимой переменной и межгрупповыми факторами «группа» (российская или китайская), «пол» (мужской или женский) и возраст в качестве разделяющих переменных.

Зависимости между экспериментальными состояниями и показателями ЭЭГ, а также между психологическими и национальными особенностями и показателями ЭЭГ оценивались в пакете eLoreta. Статистическая значимость результатов сравнения разных условий была оценена с использованием *t*-статистики для связанных измерений (paired groups) с применением метода рандомизации статистического непараметрического картирования (statistical nonparametric mapping, SnPM), который включает в себя поправку на множественные сравнения. Метод рандомизации SnPM в eLoreta основан на подходе бутстреппинга и производится путем множественных непараметрических пермутационных сравнений. Для коррекции множественных сравнений использовали 5000 рандомизаций. Корреляционный анализ был выполнен для поиска зависимости спектральной плотности от показателей личностных черт и выраженности депрессивной симптоматики.

Результаты

Статистически достоверные результаты исследований и методы их получения представлены в табл. 2.

Результаты психологического тестирования

Для показателя личностной тревожности по тесту STAI основной эффект фактора «национальность» не был достоверным ($p > 0.3$). Был выявлен достоверный эффект фактора «пол», $F(1; 62) = 6.47$; $p = 0.014$; $\eta^2 = 0.1$, средняя тревожность у женщин (30.6 ± 1.6) была выше, чем у мужчин (24.8 ± 1.7). Для показателя выраженности депрессивной симптоматики по тесту BDI выявлено статистически достоверное значение фактора «национальность», $F(1; 62) = 18.62$; $p < 0.0001$; $\eta^2 = 0.243$. Средний показатель выраженности депрессивной симптоматики в китайской группе был выше (9.2 ± 1.1), чем в российской группе (2.8 ± 1.0).

Тест RISC выявил статистически достоверные различия между национальными группами, $F(1; 62) = 7.27$; $p = 0.009$; $\eta^2 = 0.111$, по значимости семейных ценностей. Ценность семьи для китайских участников была выше (5.1 ± 0.2), чем у российских участников (4.3 ± 0.2). Тест SCS также выявил высокодостоверное значение фактора «национальность», $F(1; 62) = 23.41$; $p < 0.0001$;

Таблица 2. Перечень основных статистических результатов исследования, методов и программных продуктов для их получения

Результат	Уровень достоверности	Статистический метод	Программный продукт для статистической оценки
Психологические различия в уровне коллективизма между российскими и китайскими испытуемыми	$p < 0.0001$	Односторонний дисперсионный анализ (ANOVA)	IBM SPSS
Психологические различия в выраженности депрессивной симптоматики между российскими и китайскими испытуемыми	$p < 0.0001$	Односторонний дисперсионный анализ (ANOVA)	IBM SPSS
Различия в спектральной плотности для разных экспериментальных условий в обеих группах	$p < 0.01$	*t*-статистика для зависимых выборок	eLoreta
Различия в спектральной плотности между российской и китайской группами	$p < 0.05$	*t*-статистика для независимых выборок	eLoreta
Корреляции между спектральной плотностью и показателями личностных черт, включая коллективизм	$p < 0.05$	Регрессионный анализ с независимой переменной	eLoreta

$\eta^2 = 0.288$ для показателя коллективизма. Для участников из китайской группы показатель коллективизма был выше (5.0±0.1), чем для участников из российской группы (4.5±0.1) (рис. 2). Для этого показателя было выявлено достоверное взаимодействие факторов «пол» и «национальность», $F(1; 62) = 5.87$; $p = 0.019$; $\eta^2 = 0.092$. Российские (4.6±0.1) и китайские (4.9±0.1) женщины незначительно различались по этому показателю, тогда как для российских (4.3±0.1) и китайских (5.2±0.1) мужчин различия были выражены более существенно.

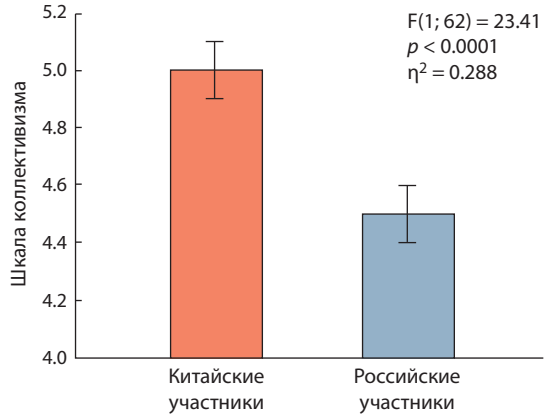


Рис. 2. Различия между китайскими и российскими участниками по показателю коллективизма в тесте SCS.

Результаты eLoreta при сравнении разных экспериментальных условий для обобщенной группы, состоящей из 62 китайских и российских участников

При помощи программного пакета eLoreta сравнивались показатели спектральной плотности на интервалах ЭЭГ с закрытыми глазами, которые следовали за интервалами наблюдения своего лица, чужого лица или пустого экрана. Было выявлено, что спектральная плотность в частотных диапазонах дельта (2–4 Гц), альфа-2 (10–12 Гц) и гамма (25–35 Гц) ритмов при закрытых глазах после наблюдения собственного лица выше, чем при закрытых глазах после наблюдения пустого экрана. Необходимо специально отметить, что мышечные артефакты были удалены из записей ЭЭГ при помощи анализа независимых компонент. Согласно данным Delorme и Makeig (2004), этот метод дает возможность удалить больше 80 % всех мышечных шумов. Это позволяет утверждать, что амплитуда электрических потенциалов в дельта- и гамма-диапазонах не была обусловлена просто поверхностной тонической электромиограммой (ЭМГ). Статистически наиболее достоверные различия ($p = 0.0036$) зафиксированы для участков предлобной коры обоих полушарий (медиальная фронтальная область, 11 поле по Бродману и орбитофронтальная кора, 47 поле по Бродману) в диапазоне альфа-2-ритма (рис. 3, *а*). Похожие результаты выявлены при сравнении условий «чужое лицо» и «пустой экран» (см. рис. 3, *б*). Показано, что так же как и при первом сравнении, спектральная плотность в предлобной коре в диапазоне альфа-2-ритма выше для условия «чужое лицо» в сравнении с условием «пустой экран» ($p = 0.002$). При сравнении ЭЭГ-интервалов, записанных после наблюдения лица незнакомого человека, было выявлено, что спектральная плотность в частотных диапазонах альфа-1 (8–10 Гц) и альфа-2 (10–12 Гц) ритмов на ЭЭГ-интервалах при закрытых глазах после наблюдения видеозаписи своего лица выше, чем на интервалах с закрытыми глазами после наблюдения лица незнакомого человека. Достоверные различия в спектральной плотности для этих условий ($p = 0.0104$) выявлены для теменной коры (верхняя теменная доля, 7 поле по Бродману, см. рис. 3, *в*).

Результаты eLoreta при сравнении разных экспериментальных условий для китайских и российских участников

Сопоставление показателей спектральной плотности между группой китайских и российских испытуемых в интер-

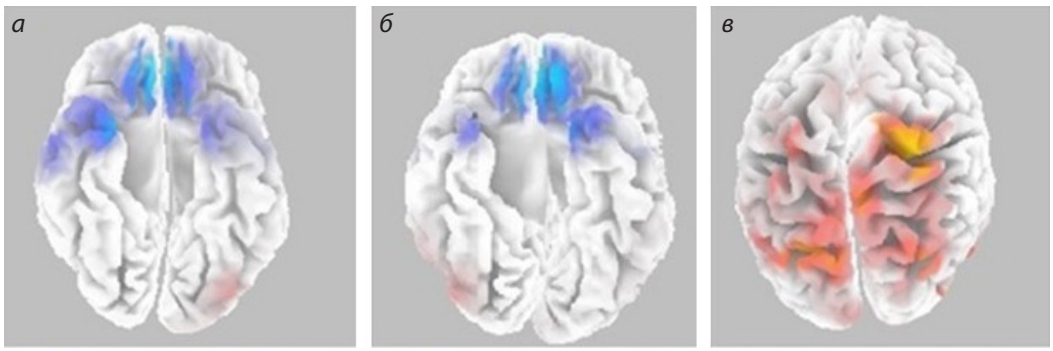


Рис. 3. Сопоставление спектральной плотности в диапазоне альфа-2 (10–12 Гц) ритма при сравнении интервалов с закрытыми глазами между условиями: (*а*) пустой экран – свое лицо; (*б*) пустой экран – чужое лицо; (*в*) свое лицо – чужое лицо.

Синим цветом отмечены области коры, в которых спектральная плотность достоверно выше для лицевых условий, чем для условия «пустой экран». Красным цветом отмечены области коры, в которых спектральная плотность достоверно выше для условия «свое лицо» в сравнении с условием «чужое лицо».

валах с закрытыми глазами, следующих за наблюдением пустого экрана, не выявило каких-либо статистически значимых межгрупповых различий. Обе группы показали одинаковое распределение спектральной плотности во всех отделах коры и всех частотных диапазонах. При межнациональных сравнениях в условиях с закрытыми глазами, следующих за наблюдением видеозаписи своего лица, выявлены достоверные различия в диапазонах альфа-2- и гамма-ритмов ($p = 0.044$) (рис. 4). У китайских участников в сравнении с российскими наблюдалась повышенная спектральная плотность в диапазоне альфа-2-ритма в теменной и височной коре (38 поле по Бродману), тогда как у российских участников в сравнении с китайскими наблюдалась повышенная спектральная плотность в медиальных лобных отделах коры в диапазоне гамма-ритма (3, 4 и 6 поля по Бродману).

При межнациональных сравнениях в условиях с закрытыми глазами между наблюдением видеозаписи лица незнакомого человека также выявлены достоверные различия в диапазонах альфа-2- и гамма-ритмов ($p = 0.0002$), но они существенно отличались от тех результатов, которые получены для условия «свое лицо» как по топографии эффекта, так и по направленности межнациональных различий. У китайских участников в отличие от российских была выявлена достоверная большая спектральная плотность в правой нижней височной области коры (38 поле по Бродману) в диапазоне гамма-ритма, тогда как российские участники в сравнении с китайскими показали большие значения спектральной плотности в обоих диапазонах (альфа-2- и гамма-ритмов) в предлобных отделах коры (медиальная фронтальная область, 11 поле по Бродману и орбитофронтальная кора, 47 поле по Бродману) (рис. 5).

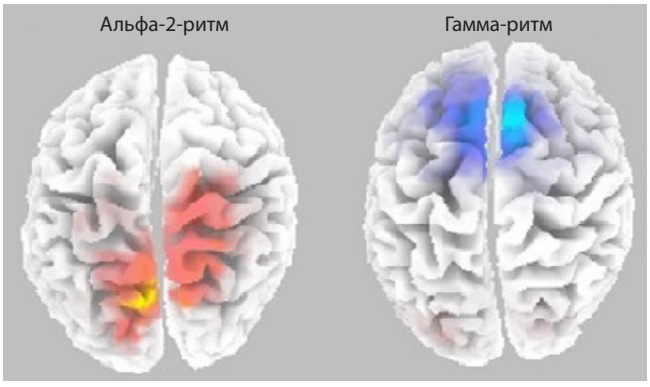


Рис. 4. Сопоставление спектральной плотности в диапазоне альфа-2 (10–12 Гц) и гамма (25–35 Гц) ритмов при сравнении китайской и российской групп для ЭЭГ-интервалов с закрытыми глазами между наблюдением участниками своего лица. Представлена конвекситальная поверхность мозга.

Для китайских участников характерна большая, в сравнении с российскими, спектральная плотность альфа-2-ритма в задних (теменных и височных) отделах коры (области отмечены красным цветом), тогда как для российских участников выявлены достоверно большие значения спектральной плотности гамма-ритма в медиальных лобных отделах коры (отмечено синим цветом).

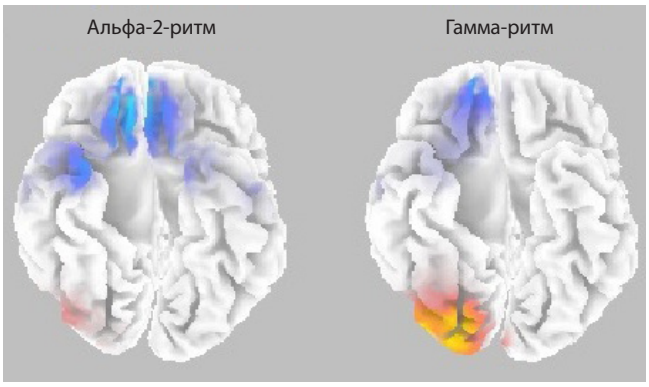


Рис. 5. Сопоставление спектральной плотности в диапазоне альфа-2 (10–12 Гц) и гамма (25–35 Гц) ритмов при сравнении китайской и российской групп для ЭЭГ-интервалов с закрытыми глазами между наблюдением участниками чужого лица. Представлена базальная поверхность мозга.

Для китайской группы характерна большая, в сравнении с российской группой, спектральная плотность гамма-ритма в правой нижней височной коре (области отмечены красным цветом), тогда как для российской группы выявлены достоверно большие значения спектральной плотности как альфа-2-, так и гамма-ритмов в предлобной коре (отмечено синим цветом).

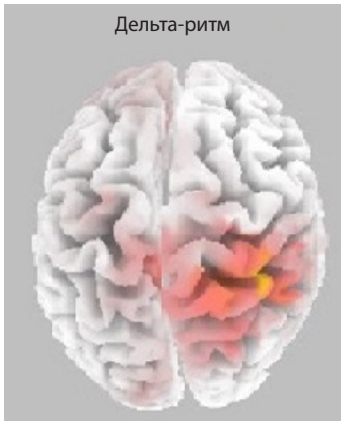


Рис. 6. Корреляции между показателем коллективизма и спектральной плотностью дельта-ритма в группе российских участников в интервалах с закрытыми глазами, следующих за наблюдением лица незнакомого человека. Представлена конвекситальная поверхность мозга.

Области коры, в которых выявлены достоверные позитивные корреляции уровня коллективизма с показателями спектральной плотности ЭЭГ, отмечены красным цветом.

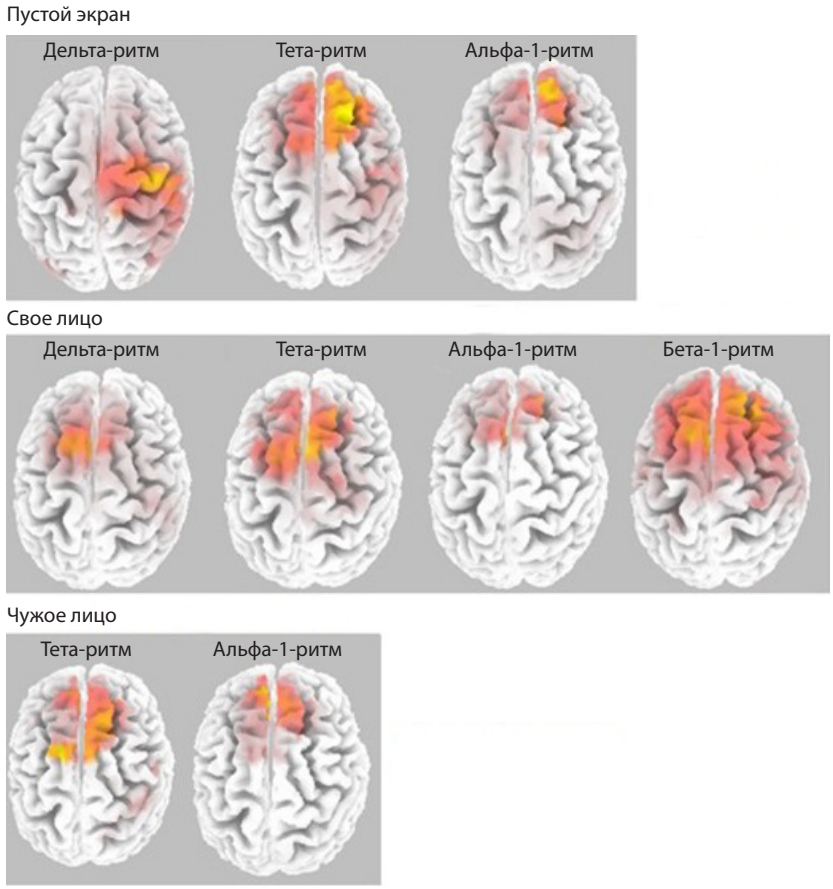


Рис. 7. Корреляции между показателем коллективизма и спектральной плотностью для условия «пустой экран» (первый ряд), «свое лицо» (второй ряд) и «чужое лицо» (третий ряд) в разных частотных диапазонах в группе китайских участников. Представлена конвекситальная поверхность мозга.

Области коры, в которых выявлены позитивные корреляции между уровнем коллективизма и спектральной плотности на ЭЭГ, отмечены красным цветом.

Результаты eLoreta при выявлении эффектов психологических показателей, зависящих от национальности и пола участников

Корреляции между показателем коллективизма по тесту SCS для смешанной группы российских и китайских испытуемых были статистически низкодостоверными. Для условия «пустой экран» достоверность отсутствовала (p = 0.1954). Для условия «свое лицо» (p = 0.0968) и «чужое лицо» (p = 0.0664) для обеих групп p-уровень был близок к достоверному значению, но не достигал его.

В российской выборке для коллективизма не выявлено значимых корреляций для условий «пустой экран» или «свое лицо». Достоверные корреляции выявлены только для спектральной плотности в диапазоне дельта-ритма для условия «чужое лицо» (p = 0.043) в правой височной коре (22 поле по Бродману) (рис. 6).

В противоположность российской выборке, для китайских участников статистически достоверные корреляции с показателем коллективизма выявлены для всех трех условий («пустой экран» p = 0.001; «свое лицо» p = 0.0032; «чужое лицо» p = 0.0334). Также можно заметить, что позитивные корреляции с показателем коллективизма в китайской группе выявились для широкого спектра частот дельта-, тета-, альфа- и бета-ритмов. Эти коррелятивные зависимости в основном выявлены в пределах переднего кластера дефолт-системы (медиальные отделы лобной и предлобной коры) и частично в правой теменно-височной коре (рис. 7).

Обсуждение

Разработка программно-аппаратного модуля для сбора и анализа данных

Целью данной работы было создание нейровычислительной технологии и разработка программно-аппаратного модуля для сбора и анализа данных, позволяющих изучать мозговые процессы, лежащие в основе личностной самореференции. Ранее мы предложили подход, совмещающий анализ ЭЭГ покоя с анализом активности лицевой мимической мускулатуры, записанной в тех же условиях (Савостьянов и др., 2022). Основной результат новой работы состоит в демонстрации возможности применения лицевых видеозаписей, полученных на начальном этапе эксперимента, для инициации у участников процессов референции информации к себе или другим людям. Такая модель сбора данных совмещается с уже хорошо известными подходами к очистке ЭЭГ-сигнала от шума (Delorme, Makeig, 2004) и локализации источников мозговой активности на поверхности коры мозга (Pascual-Margui, 2002).

Одним из результатов исследования является разработка программно-аппаратного модуля, включающего несколько последовательно подклю-

чаемых блоков для планирования эксперимента, сбора, предобработки и анализа данных, а также для проведения межгрупповых статистических сравнений. Этот модуль может в дальнейшем быть использован для проведения широкого спектра нейрофизиологических исследований, в том числе для установления маркеров аффективных заболеваний, таких как депрессия, тревожное расстройство или расстройства аутистического спектра.

Нейрофизиологические корреляты обработки самоотнесенной информации

Интерес исследователей к изучению нейрофизиологических механизмов обработки самоотнесенной информации определяется, во-первых, фундаментальной ролью самореференции в формировании человеческой личности, а во-вторых, наличием широкого ряда психиатрических заболеваний, симптомами которых являются различные нарушения в личностной самооценке (Bradley et al., 2016; Quevedo et al., 2018). В современной нейрофизиологии проходит дискуссия о наличии или отсутствии особого анатомического субстрата для самореферентных процессов в головном мозге (Northoff, Bermpohl, 2004; Northoff et al., 2005; Hu et al., 2016). В качестве основной самореферентной структуры мозга часто рассматривается дефолт-система – несколько связанных между собой участков коры, которые демонстрируют снижение уровня физиологической активности при переходе человека от состояния покоя к выполнению когнитивных заданий (Raichle, 2015; Knyazev et al., 2020, 2024).

Построение модели собственной личности в существенной степени определяется социокультурной спецификой субъекта. Было показано, что представители западной (американской) и ориентальной (японской) культур принципиально различаются по критериям так называемой Я-концепции, т. е. способу самоопределения (Markus, Kitayama, 1991). Большинство американцев демонстрировали индивидуалистические личностные установки, тогда как для японцев был свойственен коллективизм. Сравнение ЭЭГ-коррелятов, отражающих активность дефолт-системы в покое у представителей российской и китайской (Тайвань) культур, показало, что для большинства участников из Тайваня было характерно доминирование переднего (медиальная префронтальная кора) хаба дефолт-системы мозга, тогда как у российских участников наблюдалось доминирование заднего (предклинье) отдела этой системы (Knyazev et al., 2012). Была предложена гипотеза, что межнациональные различия в электрофизиологических процессах в дефолт-системе могут быть вызваны различиями в Я-концепции по критерию индивидуализм-коллективизм, характерными для представителей российской (преимущественно индивидуалистической) и китайской (коллективистской) культур. В нашем случае мы экспериментально проверяли гипотезу Князева с коллегами (2012) при помощи данных психологических опросников SCS и RISC.

Результаты межнациональных сравнений

В рамках представленного исследования было выполнено сравнение двух выборок неклинических испытуемых, проживающих на момент обследования в России, – российской и китайской. Обследование включало заполнение психологических тестов для выявления личностных особенностей участников и выраженности у них симптоматики депрессии. Нейрофизиологическая часть обследования состояла в регистрации ЭЭГ в трех экспериментальных условиях: 1 – в интервалах между наблюдением пустого экрана, 2 – в интервалах между просмотром участником видеозаписи собственного лица, 3 – в интервалах между просмотром видеозаписи лица незнакомого для участника человека.

Психологические сравнения показали, что российские и китайские испытуемые не различались по показателю личностной тревожности (тест STAI). Для выраженности депрессивной симптоматики (тест BDI) выявлено, что симптомы депрессии у китайских испытуемых выражены сильнее, чем у российских участников. Такое различие можно объяснить тем, что участники из КНР длительное время находились далеко от своего дома, тогда как российские участники находились в более привычных для них условиях. По показателям коллективизма для обоих использованных нами тестов (RISC и SCS) были выявлены высокодостоверные различия между китайскими и российскими участниками. Как и ожидалось, для китайских участников выявлены более высокие показатели коллективизма, чем для российских участников.

Сопоставления спектральной плотности между парами условий «пустой экран» *vs* «свое лицо», «пустой экран» *vs* «чужое лицо», «свое лицо» *vs* «чужое лицо» для обобщенной группы из всех участников без учета их национальности и пола выявило статистически значимые различия преимущественно в диапазоне альфа-2-ритма. Различия между нейтральным (пустой экран) и обоими социальными (как «свое», так и «чужое лицо») условиями были локализованы в пределах переднего хаба дефолт-системы (медиальная предлобная кора). В обоих случаях спектральная плотность альфа-ритма была выше для социального, чем для нейтрального условия. Различия между своим и чужим лицом были локализованы в пределах заднего хаба дефолт-системы (медиальная теменная кора) и выражались в большей спектральной плотности для своего, чем для чужого лица.

Межнациональные различия, без учета фактора пола и психологических различий, не были выявлены на ЭЭГ, записанной в интервалах между наблюдением пустого экрана, но были выявлены для интервалов между наблюдением как своего, так и чужого лица. Для случая своего лица различия выявлены в диапазоне альфа-2-ритма в заднем хабе дефолт-системы (у китайских участников спектральная плотность выше, чем у российских), но в диапазоне гамма-ритма в переднем хабе дефолт-системы у российских участников спектральная плотность выше, чем у китайских. Для условия «чужое лицо» в переднем хабе дефолт-системы у российских участников была выявлена большая плотность источников как альфа-, так и гамма-ритмов, тогда как для китайских участников большая спектральная плотность выявлена в височной коре. Таким образом, наш результат в целом подтверждает вывод Г.Г. Князева с коллегами (2012) о наличии межнациональных различий в работе переднего и заднего хабов дефолт-системы.

В группе российских испытуемых оценки коллективизма коррелировали с показателями мозговой активности только для условия «чужое лицо». Эти корреляции затрагивали задний хаб дефолт-системы. Наоборот, в группе китайских испытуемых коллективизм оказался психологической метрикой, для которой были найдены множественные достоверные корреляции для всех трех экспериментальных условий и одновременно нескольких частотных диапазонов. Большинство значимых корреляций в китайской группе найдено для мозговых структур из переднего (медиальная лобная, медиальная предлобная кора) хаба дефолт-системы. Таким образом, мы подтверждаем гипотезу, что различия в активности дефолт-системы между российскими и китайскими испытуемыми обусловлены в основном их различиями по показателю коллективизма.

В целом, благодаря предложенной в этом исследовании новой экспериментальной модели, нам удалось подтвердить гипотезу Г.Г. Князева с коллегами (2012) о том, что межкультурные различия в активности дефолт-системы у китайских и российских участников ассоциированы с их различиями в показателях коллективизма.

В результате исследования мы осуществили начальный этап разработки комплексной нейровычислительной технологии сбора и анализа психологических и физиологических данных, позволяющей исследовать динамику обработки самореферентной информации в зависимости от культурологической специфики участников обследования. Разработанный нами программно-аппаратный модуль включен в интегрированную цифровую платформу «Биоинформатика и системная компьютерная биология», разрабатываемую в ИЦиГ СО РАН в рамках бюджетного проекта № FWNR-2022-0020. Можно ожидать, что полученный подход будет в дальнейшем совмещен с результатами нейрокомпьютерных исследований на основе обработки фМРТ (Haxby et al., 2001) или с данными психогенетических исследований. Например, для части наших испытуемых были собраны данные, касающиеся их однонуклеотидных полиморфизмов в локусах нейротрансмиттерных систем мозга (Ivanov et al., 2022). Поэтому результаты психологических и нейрофизиологических исследований могут быть сопоставлены с генетическими особенностями участников. Кроме того, для классификации участников на подгруппы, ассоциированные с разным уровнем стресса, могут быть использованы сверточные нейронные сети, использующие метрики ЭЭГ в качестве входных параметров (Fu et al., 2023).

Заключение

1. Электрическая активность мозга, записанная в интервалах функционального покоя, следующих за стимуляцией, различается для условий после предъявления участникам нейтральной, самоотнесенной или отнесенной к другим людям информации. Такая зависимость проявляется в показателях спектральной плотности альфа-2-ритма в корковых областях, входящих в состав дефолт-системы мозга.
2. Функциональная активность дефолт-системы у китайских и российских испытуемых различается в интервалах покоя, следующих за наблюдением субъект-отнесенных стимулов, но не различается для интервалов, следующих за наблюдением пустого экрана. Функциональная активность в переднем и заднем хабе дефолт-системы существенно зависит от национальной принадлежности участников.
3. Функциональная активность переднего хаба дефолт-системы ассоциирована с показателями коллективизма у китайских, но не у российских участников.

Ограничения

1. При записи ЭЭГ не регистрировалась ЭМГ скальпа, определяющая психоэмоциональное напряжение. Хотя нами была выполнена процедура вычисления и удаления усредненного референта, можно допустить, что эффекты личностных черт и национальности в диапазонах гамма- и бета- ритмов связаны не только с мозговой, но и с мышечной активностью.
2. Нами выбраны стандартные, а не индивидуальные, границы частотных диапазонов, что может снизить точность выявления персонализированных ЭЭГ-коррелятов когнитивных процессов, особенно для альфа-ритма. К сожалению, выбранный нами программный пакет не позволяет анализировать спектральную плотность в индивидуальных диапазонах.
3. Хотя все участницы женского пола были опрошены перед обследованием для установления у них недели месячного цикла, мы не учитывали психоэндокринологический фактор флуктуации гормонального состояния у женщин при анализе результатов ЭЭГ, что могло снизить точность сделанных выводов.

Мы принимаем все перечисленные ограничения и постараемся учесть их в дальнейших исследованиях.

Список литературы / References

Савостьянов А.Н., Вергунов Е.Г., Сапрыгин А.Е., Лебедкин Д.А. Апробация технологии оценки мимики лиц для изучения динамики функциональных состояний человека в ЭЭГ-парадигме покоя. *Вавиловский журнал генетики и селекции.* 2022;26(8): 765-772. doi 10.18699/VJGB-22-92
[Savostyanov A.N., Vergunov E.G., Saprygin A.E., Lebedkin D.A. Validation of a face image assessment technology to study the dynamics of human functional states in the EEG resting-state paradigm. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2022;26(8):765-772. doi 10.18699/VJGB-22-92]

Ханин Ю.Л. Краткое руководство к шкале реактивной и личностной тревожности Ч.Д. Спилбергера. Ленинград, 1976
[Khanin Y.L. A Brief Guide to the C.D. Spielberger State and Trait Anxiety Scale. Leningrad, 1976 (in Russian)]

Яковлева Е. Теория референции и теория психосистематики. *Известия Российского государственного педагогического университета им. А.И. Герцена*. 2011;131:226-233
[Yakovleva E.V. Theory of reference and theory of psyhosystematics. *Izvestiia Rossiiskogo Gosudarstvennogo Pedagogicheskogo Universiteta im. A.I. Gertsena = Izvestia: Herzen University Journal of Humanities and Sciences*. 2011;131:226-233 (in Russian)]

Beck A.T., Steer R.A., Brown G.K. Manual for the Beck Depression Inventory II. San Antonio, TX: Psychological Corporation, 1996

Bradley K.A., Colcombe S., Henderson S.E., Alonso C.M., Milham M.P., Gabbay V. Neural correlates of self-perceptions in adolescents with major depressive disorder. *Dev. Cogn. Neurosci*. 2016; 19:87-97. doi 10.1016/j.dcn.2016.02.007

Cross S.E., Bacon P.L., Morris M.L. The relational-interdependent self-construal and relationships. *J. Pers. Soc. Psychol*. 2000;78(4): 791-808

Delorme A., Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci*. *Methods*. 2004;134(1):9-21. doi 10.1016/j.jneumeth.2003.10.009

Fu X., Tamozhnikov S.S., Saprygin A.E., Istomina N.A., Klemeshova D.I., Savostyanov A.N. Convolutional neural networks for classifying healthy individuals practicing or not practicing meditation according to the EEG data. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):851-858. doi 10.18699/VJGB-23-98

Haxby J.V., Gobbini M.I., Furey M.L. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 2001;293(5539):2425-2430. doi 10.1126/science.1063736

Hu C., Di X., Eickhoff S.B., Zhang M., Peng K., Guo H., Sui J. Distinct and common aspects of physical and psychological self-representation in the brain: a meta-analysis of self-bias in facial and self-referential judgements. *Neurosci. Biobehav*. *Rev*. 2016;61:197-207. doi 10.1016/j.neubiorev.2015.12.003

Ivanov R., Kazantsev F., Zavarzin E., Klimenko A., Milakhina N., Matushkin Y.G., Savostyanov A., Lashin S. ICBrainDB: an integrated database for finding associations between genetic factors and EEG markers of depressive disorders. *J. Pers. Med.* 2022;12(1):53. doi 10.3390/jpm12010053

Knyazev G.G., Savostyanov A.N., Volf N.V., Liou M., Bocharov A.V. EEG correlates of spontaneous self-referential thoughts: a cross-cultural study. *Int. J. Psychophysiol*. 2012;86(2):173-181. doi 10.1016/j.ijpsycho.2012.09.002

Knyazev G.G., Savostyanov A.N., Bocharov A.V., Tamozhnikov S.S., Saprigyn A.E. Task-positive and task-negative networks and their relation to depression: EEG beamformer analysis. *Behav. Brain Res*. 2016;306:160-169. doi 10.1016/j.bbr.2016.03.033

Knyazev G.G., Savostyanov A.N., Bocharov A.V., Levin E.A., Rudych P.D. Intrinsic connectivity networks in the self- and other-referential processing. *Front. Hum. Neurosci*. 2020;14:579703. doi 10.3389/fnhum.2020.579703

Knyazev G.G., Savostyanov A.N., Bocharov A.V., Saprigyn A.E. Representational similarity analysis of self-versus other-processing: effect of trait aggressiveness. *Aggress. Behav*. 2024;50(1):e22125. doi 10.1002/ab.22125

Markus H.R., Kitayama S. Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev*. 1991;98(2):224-253. doi 10.1037/0033-295X.98.2.224

Neff K.D., McGehee P. Self-compassion and psychological resilience among adolescents and young adults. *Self Identity*. 2010;9(3):225-240. doi 10.1080/15298860902979307

Northoff G., Bermpohl F. Cortical midline structures and the self. *Trends Cogn. Sci*. 2004;8(3):102-107. doi 10.1016/j.tics.2004.01.004

Northoff G., Heinzel A., De Greck M., Bermpohl F., Dobrowolny H., Panksepp J. Self-referential processing in our brain – a meta-analysis of imaging studies on the self. *NeuroImage*. 2005;31(1):440-457. doi 10.1016/j.neuroimage.2005.12.002

Pascual-Margui R.D. Standardized low-resolution brain electromagnetic tomography (sLORETA). Technical details. *Methods Find. Exp. Clin. Pharmacol*. 2002;24(Suppl.D):5-12

Quevedo K., Harms M., Sauder M., Scott H., Mohamed S., Thomas K.M., Schallmo M.-P., Smyda G. The neurobiology of self face recognition among depressed adolescents. *J. Affect. Disord*. 2018; 229:22-31. doi 10.1016/j.jad.2017.12.023

Raichle M.E. The brain's default mode network. *Annu. Rev. Neurosci*. 2015;38:433-447. doi 10.1146/annurev-neuro-071013-014030

Singelis T.M. The measurement of independent and interdependent self-construals. *Personality Social Psychol. Bull*. 1994;20(5):580-591. doi 10.1177/0146167294205014

Spielberger C.D., Gorsuch R.L., Lushene R.E. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press, 1970

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 05.10.2024. После доработки 12.11.2024. Принята к публикации 13.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):993-1007

doi 10.18699/vjgb-24-106

Концепция природной реконструкции генома.
Часть 2. Влияние фрагментов экстраклеточной двуцепочечной ДНК на гемопоэтические стволовые клетки

В.С. Рузанова [1&], С.Г. Ошихмина [1, 2&], А.С. Проскурина [1], Г.С. Риттер [1], С.С. Кирикович [1], Е.В. Левитес [1], Я.Р. Ефремов [1], Т.В. Карамышева[1], М.И. Мещанинова [3], А.Л. Мамаев[4], О.С. Таранов [5], А.С. Богачев[2], С.В. Сидоров[6], С.Д. Никонов[7], О.Ю. Леплина [8], А.А. Останин [8], Е.Р. Черных [8], Н.А. Колчанов [1], Е.В. Долгова [1#], С.С. Богачев [1#]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
[4] ООО «Лаборатория Ангиофарм», Новосибирск, Россия
[5] Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора, р. п. Кольцово, Новосибирская область, Россия
[6] Городская клиническая больница № 1, Новосибирск, Россия
[7] Новосибирский научно-исследовательский институт туберкулеза, Новосибирск, Россия
[8] Научно-исследовательский институт фундаментальной и клинической иммунологии, Новосибирск, Россия

labmolbiol@mail.ru

Аннотация. В настоящей части исследования доказывается первая составляющая концепции «природной реконструкции генома». На модельных организмах мышь и человек показано, что CD34+ гемопоэтические предшественники костного мозга захватывают фрагменты экстраклеточной двуцепочечной ДНК естественным природным механизмом. Известно, что в процессе интернализации экстраклеточных фрагментов ДНК принимают участие структуры гликокаликса, в состав которых входят гликопротеины/протеогликаны, гликозилфосфатидилинозитол-заякоренные белки и скавенджер-рецепторы. Проведенный биоинформационный анализ свидетельствует, что основные поверхностные маркерные белки гемопоэтических стволовых клеток относятся к указанным группам факторов и содержат специфические сайты связывания ДНК, включающие гепарин-связывающий домен и кластеры положительно заряженных аминокислотных остатков. С использованием системы Electrophoretic mobility shift assay показано прямое взаимодействие CD34 и CD84 (SLAMF5) гликопротеинов, маркеров гемопоэтических стволовых клеток, с фрагментами двуцепочечной ДНК. В клетках, негативных по CD34, также интернализующих фрагменты, происходит конкатемеризация доставленных внутрь клетки фрагментов. При этом в одну структуру сшивается до пяти мономеров олигонуклеотидов, содержащих девять теломерных повторов TTAGGG. Доставленные в гемопоэтические стволовые клетки экстраклеточные фрагменты инициируют деление исходной гемопоэтической стволовой клетки таким образом, что одна из дочерних клеток уходит в терминальную дифференцировку, а вторая сохраняет свой низкодифференцированный статус. В составе колоний после обработки клеток костного мозга препаратом hDNA^{gr} количество CD34+ клеток возрастает до 3 % (модельный организм – человек). Одновременно обработка препаратом hDNA^{gr} индуцирует пролиферацию стволовых клеток крови и их ближайших потомков и стимулирует колониеобразование (модельные организмы – мышь, крыса, человек). Наиболее часто в результате обработки экстраклеточной двуцепочечной ДНК активируется гранулоцитарно-макрофагальный росток кроветворения. Процесс коммитирования манифестируется появлением и репарацией пангеномных одноцепочечных разрывов. Время перехода в направлении дифференцировки (время появления и репарации пангеномных одноцепочечных разрывов) составляет около 7 суток. Предполагается, что в момент инициации пангеномных одноцепочечных разрывов в клетке создается «рекомбиногенная ситуация» и активируются молекулярные репаративно-рекомбинационные механизмы. Во всех проведенных экспериментах по анализу индивидуальных молекул в качестве фактора сравнения использовался ангиогенин рекомбинантный человеческий. Во всех других экспериментах одной из сравниваемых групп являлись гемопоэтические стволовые клетки, обработанные ангиогенином.

Ключевые слова: гемопоэтические стволовые клетки; экстраклеточная ДНК; интернализация; терминальная дифференцировка; одноцепочечные разрывы.

Для цитирования: Рузанова В.С., Ошихмина С.Г., Проскурина А.С., Риттер Г.С., Кирикович С.С., Левитес Е.В., Ефремов Я.Р., Карамышева Т.В., Мещанинова М.И., Мамаев А.Л., Таранов О.С., Богачев А.С., Сидоров С.В., Никонов С.Д., Леплина О.Ю., Останин А.А., Черных Е.Р., Колчанов Н.А., Долгова Е.В., Богачев С.С. Концепция природной реконструкции генома. Часть 2. Влияние фрагментов экстраклеточной двуцепочечной ДНК на гемопоэтические стволовые клетки. *Вавиловский журнал генетики и селекции*. 2024;28(8):993-1007. doi 10.18699/vjgb-24-106

Финансирование. Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации в Институте цитологии и генетики СО РАН (государственный бюджетный проект № FWNR-2022-0016) и при поддержке И.Н. Зайцевой, А.А. Пуртова и ООО «ЭС.ЛАБ ДИАГНОСТИК».

© Рузанова В.С., Ошихмина С.Г., Проскурина А.С., Риттер Г.С., Кирикович С.С., Левитес Е.В., Ефремов Я.Р., Карамышева Т.В., Мещанинова М.И., Мамаев А.Л., Таранов О.С., Богачев А.С., Сидоров С.В., Никонов С.Д., Леплина О.Ю., Останин А.А., Черных Е.Р., Колчанов Н.А., Долгова Е.В., Богачев С.С., 2024

[&, #] Авторы внесли равный вклад.

Контент доступен под лицензией Creative Commons Attribution 4.0

A concept of natural genome reconstruction.
Part 2. Effect of extracellular double-stranded DNA fragments on hematopoietic stem cells

V.S. Ruzanova [1&], S.G. Oshikhmina [1, 2&], A.S. Proskurina [1], G.S. Ritter [1], S.S. Kirikovich [1], E.V. Levites [1], Y.R. Efremov [1], T.V. Karamysheva[1], M.I. Meschaninova [3], A.L. Mamaev[4], O.S. Taranov [5], A.S. Bogachev[2], S.V. Sidorov[6], S.D. Nikonov[7], O.Y. Leplina [8], A.A. Ostanin [8], E.R. Chernykh [8], N.A. Kolchanov [1], E.V. Dolgova [1#], S.S. Bogachev [1#] ✉

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[4] Laboratory Angiopharm LLC, Novosibirsk, Russia
[5] State Scientific Center of Virology and Biotechnology "Vector" of Rospotrebnadzor, Koltsovo, Novosibirsk region, Russia
[6] City Clinical Hospital No. 1, Novosibirsk, Russia
[7] Novosibirsk Tuberculosis Research Institute, Novosibirsk, Russia
[8] Research Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia
✉ labmolbiol@mail.ru

Abstract. In this part of the study, the first component of the concept of "natural genome reconstruction" is being proven. It was shown with mouse and human model organisms that CD34+ hematopoietic bone marrow progenitors take up fragments of extracellular double-stranded DNA through a natural mechanism. It is known that the process of internalization of extracellular DNA fragments involves glycocalyx structures, which include glycoproteins/protein glycans, glycosylphosphatidylinositol-anchored proteins and scavenger receptors. The bioinformatic analysis conducted indicates that the main surface marker proteins of hematopoietic stem cells belong to the indicated groups of factors and contain specific DNA binding sites, including a heparin-binding domain and clusters of positively charged amino acid residues. A direct interaction of CD34 and CD84 (SLAMF5) glycoproteins, markers of hematopoietic stem cells, with double-stranded DNA fragments was demonstrated using an electrophoretic mobility shift assay system. In cells negative for CD34, which also internalize fragments, concatemerization of the fragments delivered into the cell occurs. In this case, up to five oligonucleotide monomers containing 9 telomeric TTAGGG repeats are stitched together into one structure. Extracellular fragments delivered to hematopoietic stem cells initiate division of the original hematopoietic stem cell in such a way that one of the daughter cells becomes committed to terminal differentiation, and the second retains its low-differentiated status. After treatment of bone marrow cells with hDNA^{gr}, the number of CD34+ cells in the colonies increases to 3 % (humans as the model organism). At the same time, treatment with hDNA^{gr} induces proliferation of blood stem cells and their immediate descendants and stimulates colony formation (mouse, rat and humans as the model organisms). Most often, the granulocyte-macrophage lineage of hematopoiesis is activated as a result of processing extracellular double-stranded DNA. The commitment process is manifested by the appearance and repair of pangenomic single-strand breaks. The transition time in the direction of differentiation (the time it takes for pangenomic single-strand breaks to appear and to be repaired) is about 7 days. It is assumed that at the moment of initiation of pangenomic single-strand breaks, a "recombinogenic situation" ensues in the cell and molecular repair and recombination mechanisms are activated. In all experiments with individual molecules, recombinant human angiogenin was used as a comparison factor. In all other experiments, one of the experimental groups consisted of hematopoietic stem cells treated with angiogenin.

Key words: hematopoietic stem cells; extracellular DNA; internalization; terminal differentiation; single-strand breaks.

For citation: Ruzanova V.S., Oshikhmina S.G., Proskurina A.S., Ritter G.S., Kirikovich S.S., Levites E.V., Efremov Y.R., Karamysheva T.V., Meschaninova M.I., Mamaev A.L., Taranov O.S., Bogachev A.S., Sidorov S.V., Nikonov S.D., Leplina O.Y., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Dolgova E.V., Bogachev S.S. A concept of natural genome reconstruction. Part 2. Effect of extracellular double-stranded DNA fragments on hematopoietic stem cells. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):993-1007. doi 10.18699/vjgb-24-106

Введение

Гемопоэтическая стволовая клетка (ГСК) и ее костномозговая стволовая ниша образуют уникальную клеточную систему, которая формирует и поддерживает на протяжении жизни организма баланс клеточных элементов крови и отвечает за репарацию тканей и органов. Понятие ГСК – сложно-составляющее, характеризует несколько клеточных состояний и несколько типов клеток различной анатомической локализации, развивающихся в различные клеточные ростки. Считается, что существует три класса ГСК: миелоидно-ориентированные, лимфоидно-ориентированные и «сбалансированные», которые имеют различный потенциал дифференцировки, закрепленный эпигенетически. Клональный анализ свидетельствует, что эти классы клеток состоят из двух популяций: короткоживущих ГСК и долгоживущих прогениторов. Первые уходят в дифференцировку и пролиферацию в течение нескольких недель, тогда как долгоживущие прогениторы находятся в покоящейся стадии G0 на протяжении длительного времени (Muller-Sieburg, Sieburg, 2008).

Общепринято, что долгоживущие покоящиеся ГСК мыши имеют фенотип Lin– Kit+ Sca-1+ CD150+ CD34– Flk2– CD48–. На одну ГСК приходится около 30 тысяч мононуклеаров костного мозга, и около 80 % ГСК оста-

ются в состоянии покоя, сохраняя «стволовость», до конца жизни (Morita et al., 2010; Zhang, Sadek, 2014; Wilkinson et al., 2020).

Гемопоэтическая стволовая клетка окружена клетками, относящимися к нескольким типам, образующими нишу, в которой ГСК осуществляет свои функции. В составе стволовой ниши определяют эндотелиальные клетки, множественные мезенхимальные типы клеток (адипоциты, CXCL12+ адвентициальные ретикулярные [CAR] клетки, клетки остеоростка, лептинR+ и нестин+ клетки и NG2+ клетки оболочки артериол), а также немиелинизированные шванновские клетки и гемопоэтические клетки (макрофаги и мегакариоциты) (Lévesque et al., 2010; Mendelson, Frenette, 2014; Kumar, Geiger, 2017; Szade et al., 2018; Lucas, 2019).

В настоящее время предполагают существование двух типов ниш кроветворных стволовых клеток в костном мозге взрослого человека. Остеобластическая ниша определяет покоящееся состояние ранних примитивных прогениторов, длительно сохраняющих «стволовость». При активации ГСК дифференцируются в предшественники крови, располагающиеся в пределах сосудистой ниши, прилегая к синусоидным эндотелиальным клеткам (Redondo et al., 2017).

Краеугольным, фундаментальным свойством примитивной ГСК является ее имманентный выбор: сохранить покоящееся состояние, поделиться симметрично с образованием двух одинаковых ГСК или поделиться несимметрично и дать коммитированную клетку и далее – развитие определенного клеточного ростка.

Функциональность ГСК напрямую связана с балансом между покоем и активацией. Снижение способности ГСК выходить из состояния покоя редуцирует потенциал воспроизводства форменных элементов крови. С другой стороны, если неоправданно большое количество ГСК выходит из состояния покоя или не возвращается в состояние покоя после активации, то происходит истощение пула ГСК, что приводит к функциональной несостоятельности костного мозга (Scharf et al., 2020). Известно, что ГСК молодого организма чаще выходят в симметричное деление или пролиферируют, тогда как прогениторы взрослого и стареющего организма в основном находятся в состоянии покоя (Desterke et al., 2021).

В процессе выбора состояния ГСК принимают участие многие факторы. Первоочередными являются анатомическая локализация ГСК в составе стволовой ниши и связанный с этим уровень гипоксии в области нахождения указанной системы клеток. Гипоксия – один из ключевых факторов состояния ГСК, и большинство покоящихся и примитивных ГСК находятся в гипоксических областях костного мозга с пониженной перфузией крови (Forristal, Levesque, 2014; Zhang, Sadek, 2014). Секретируемые факторы стволовых ниш и самой ГСК, мембран-связанные факторы (Winkler et al., 2012; Forristal, Levesque, 2014; Goncalves et al., 2016; Silberstein et al., 2016; Redondo et al., 2017; Chen T.L. et al., 2018; Scharf et al., 2020; Desterke et al., 2021) являются важнейшими участниками процессов, формирующих биологическое состояние ГСК. При этом один и тот же фактор может выступать как индуктором покоя для одного типа ГСК, так и индуктором перехода в цикл и коммитирование для другого типа, как это показано для ангиогенина (Goncalves et al., 2016). Мигрирующие периферические лейкоциты и секретируемые ими гистамин и ФНО-α, а также другие клетки костного мозга и периферической крови приводят к активации покоящихся прогениторов (Lucas, 2019; Pinho, Frenette, 2019). Фармакологические препараты различного происхождения, воспаление, голод, ксенобиотики окружающей среды, радиация также определяют судьбу ГСК (Chen T.L. et al., 2018; Scharf et al., 2020; Kiang et al., 2021; Wang et al., 2021).

Несимметричное деление и следующие за этим коммитирование и пролиферация составляют базовый механизм восполнения популяций клеток крови. Этот процесс представляет собой тонко регулируемую цепь событий с разнообразной и многочисленной палитрой возможных индукторов. Как уже было сказано, процесс терминальной дифференцировки, пролиферации и мобилизации ГСК может быть активирован факторами окружающей среды и физиологическими системами организма как системами интегральных раздражителей, формирующих общий вектор ответа ГСК и ее окружения (стволовой ниши). В результате такого воздействия происходит активация молекулярных сигнальных каскадов и генных платформ, определяющих судьбу ГСК и коммитированного потомка (Kulkarni, Kale, 2020). Один из инициирующих факторов этого процесса – воспаление. В результате воспалительной реакции в кровяное русло и лимфатическую систему выбрасывается огромное количество разнообразных активных молекул, включая палитру провоспалительных цитокинов, глюкокортикоидов (Pierce et al., 2017), гранулоцитарно-макрофагальный колониестимулирующий фактор (GM-CSF) и др., которые выполняют роль триггера выхода покоящейся ГСК в цикл. Также в процессе воспаления, независимо, стерильного или индуцированного патогеном, в кровяном русле появляется значительное количество молекул ДНК апоптотирующих клеток организма (селф-ДНК) или патоген-ассоциированной двуцепочечной ДНК (дцДНК) и РНК (Jiang, Pisetsky, 2005; Saitoh et al., 2010; Lauková et al., 2019; Korabecna et al., 2020; Kananen et al., 2023). Участие воспалительного процесса в терминальной дифференцировке ГСК означает, что все факторы, выделяемые в кровь при воспалении, включая фрагменты экстраклеточной селф/патоген-консумент-ассоциированной ДНК, транзитно, конкурентно или ограниченно влияют на принятие решения примитивными предшественниками (Seita, Weissman, 2010). Считается, что воспаление смещает направление дифференцировки гемопоэтических предшественников в миелоидном направлении (Kovtonyuk et al., 2016).

В наших недавних работах показано, что стволовые клетки различного генеза, стволовые опухолевые клетки (Ritter et al., 2022), ГСК (Potter et al., 2024) интернализуют фрагменты экстраклеточной дцДНК естественным природным механизмом. Мы полагаем, что это новое свойство низкодифференцированных клеток, в том числе ГСК, является переходным звеном в понимании процессов, протекающих в гемопоэтических предшественниках, включая выход в терминальную дифференцировку и пролиферацию при их взаимодействии с фрагментами экстраклеточной дцДНК, циркулирующей в кровяном русле.

Существует еще одно событие, которое представляет собой основу концепции, предложенной в первой части работы. Это появление одноцепочечных разрывов (ников) в геноме стволовых клеток и связь этого процесса с терминальной дифференцировкой прогениторов.

Впервые одноцепочечные разрывы хроматина, возникающие при коммитировании, были описаны в работах конца прошлого века на нескольких эукариотических моделях. Для анализа событий, происходящих с хроматином ядра при коммитировании, использовались индукторы DMSO, бутират натрия, бутирилхолин, ретиноевая кислота. Детекцию одноцепочечных разрывов проводили методами седиментационного анализа (Jacobson et al., 1975; Scher, Friend, 1978), хроматографии на гидроксиапатите (Pulito et al., 1983), щелочной фильтрации (McMahon et al., 1984; Boerrigter et al., 1989; Kaminskas, Li, 1989), *in situ* ник-трансляции (Iseki, 1986; Patkin et al., 1995), щелочного электрофореза (McMahon et al., 1984; Vatolin et al., 1997). Оказалось, что появление и репарация одноцепочечных разрывов – дозозависимый и времязависимый процесс, который не коррелирует с направлением дифференцировки (Scher, Friend, 1978; Farzaneh et al., 1982).

Установлено, что никирование хроматина связано с активностью кальций/магний-зависимых ДНКаз, т. е. является энзиматическим, а одноцепочечные разрывы рандомны (McMahon et al., 1984; Kaminskas, Li, 1989). Репарация одноцепочечных разрывов сопряжена с активностью ADP-рибозил трансферазы, которая, как считается, через стимуляцию лигазной активности также вовлечена в контроль дифференцировки (Farzaneh et al., 1982; Johnstone, Williams, 1982). Интересные и сложно интерпретируемые результаты получены в работе (Patkin et al., 1995), где методом *in situ* ник-трансляции показано, что метафазные хромосомы стволовых клеток в переходе в коммитированное состояние содержат многочисленные ники.

Таким образом, установлено, что появление одноцепочечных разрывов тесно коррелирует с терминальной дифференцировкой стволовых клеток. Это событие считается наиболее ранней манифестацией начавшегося процесса коммитирования. Одноцепочечные разрывы не связаны с апоптозом, не заканчиваются гибелью клетки, и через определенное время целостность хроматина восстанавливается. Одним из объяснений феномена является предположение, что в момент релаксации генома, связанной с появлением ников, происходит активация генов, необходимых для коммитирования (Jacobson et al., 1975; Farzaneh et al., 1982). Мы полагаем, что обнаруженный феномен является краеугольным камнем всего процесса дифференцировки: биологическим, надмолекулярным, масштабным проявлением смены клеткой своего биологического статуса. Именно пангеномные одноцепочечные разрывы позволяют клетке, по-видимому, с минимальными энергетическим затратами реорганизовать топологию хроматина недифференцированного состояния в новую архитектуру, требующуюся для специализации клетки, что, естественно, связано с принципиальным изменением платформы экспрессирующихся генов, как следует из рассуждений в (Jacobson et al., 1975; Farzaneh et al., 1982). Именно это явление мы попытались охарактеризовать в настоящей части работы в новых экспериментальных реалиях, причем обозначив индуктором процесса экстраклеточные фрагменты дцДНК.

К сожалению, в доступной литературе за последние двадцать лет нам не удалось обнаружить публикации, связанные с появлением и репарацией одноцепочечных разрывов в гемопоэтических стволовых клетках. И абсолютно непонятно, почему это направление, характеризующее терминальный переход низкодифференцированных стволовых клеток различного генеза, не получило дальнейшего развития.

Таким образом, во второй части цикла работ были проведены исследования, характеризующие интернализацию фрагментов дцДНК в ГСК, индукцию этими фрагментами терминальной дифференцировки прогениторов, манифестацией которой являются возникновение и репарация пангеномных одноцепочечных разрывов.

Материалы и методы

Экспериментальные животные. В работе использованы самцы мышей CBA/Lac в возрасте от 2 до 5 месяцев, старые самцы мышей CBA/Lac в возрасте от 9 до 12 месяцев, самцы крыс линии Wistar в возрасте от 2 до 6 месяцев и старые самцы крыс линии Wistar в возрасте от 18 до 22 месяцев, выведенные в ЦКП «Виварий конвенциональных животных» Института цитологии и генетики СО РАН (Новосибирск, Россия). Животных содержали группами по 6–10 мышей и по 3–4 крысы в клетке со свободным доступом к пище и воде. Все эксперименты с животными были одобрены Комитетом по уходу и использованию животных Института цитологии и генетики СО РАН. Мышей выводили из эксперимента методом цервикальной дислокации шейных позвонков, крыс – методом эвтаназии с использованием CO_2 или декапитацией.

Клетки костного мозга человека. В работе были использованы клетки криоконсервированного сепарата костного мозга больных лимфомой Ходжкина, предоставленного криобанком НИИФКИ.

Препарат hDNA^{gr} (реконструктор ДНК генома человека) выделен из плацент здоровых женщин. Тотальная геномная ДНК была фрагментирована до 1–20 нуклеосомных мономеров (200–2000 п. н.) ультразвуковой дезинтеграцией, депротеинизирована с использованием протеиназы К и экстрагирована обработкой фенол-хлороформом.

Ангиогенин рекомбинантный человеческий предоставлен ООО «Лаборатория Ангиофарм» (Новосибирск, Россия). Мечение ангиогенина Cy5 проводили согласно протоколу производителя компании Lumiprobe (Германия).

TAMRA-меченый ДНК зонд. Флуоресцентное мечение красителем TAMRA ДНК *AluI* повтора человека проводили с использованием ПЦР на основе включения TAMRA-5′-dUTP (дезоксиуридинтрифосфата), как описано в работе (Dolgova et al., 2014).

Оценка изменения подвижности в геле комплекса белков CD34 и SLAMF5 и ДНК зондов. Для анализа взаимодействия белков CD34 и SLAMF5 и TAMRA-ДНК зонда и P^{32}-меченого двуцепочечного $(TTAGGG)_9$ теломерного повтора образцы белков и ДНК инкубировали в различных соотношениях и различное время при 37 °C в 10 мМ PBS (см. подпись к рис. 1). Включение γP^{32}-ATP

и нативный электрофорез в полиакриамидном геле проводили согласно стандартным процедурам (Маниатис и др., 1984; DNA Cloning…, 1985).

Выделение клеток костного мозга. Для выделения костного мозга животных выводили из эксперимента, выделяли бедренные и большеберцовые кости, отсекали эпифизы и промывали костномозговую полость IMDM + 2 % FBS. Клеточную суспензию пропускали несколько раз через иглу 21-го калибра, чтобы избавиться от костномозговых розеток, фильтровали через 40 мкм. Клетки переосаждали 10 мин при 400g и ресуспендировали в буфере со 130 мМ хлоридом аммония для лизиса эритроцитов на 3–5 мин. Затем буфер разбавляли в 10 раз PBS, клетки вновь переосаждали, ресуспендировали в среде IMDM и подсчитывали в камере Горяева.

Интернализация ДНК и ангиогенина в ГСК человека и мыши. Для окрашивания колоний ГСК к клеткам в 100 мкл среды IMDM добавляли антитела против мышиных Sca-1 и c-Kit согласно протоколу производителя и 0.1 мкг TAMRA-меченой ДНК. Полученная смесь была внесена на чашки Петри (диаметр 35 мм) с колониями ГСК, не касаясь метилцеллюлозы и колоний, и рассредоточена по небольшой площади поверхности. Сбор данных и визуализация выполнялись с использованием лазерного сканирующего конфокального микроскопа LSM 780 NLO (Zeiss) и программного обеспечения ZenLight.

Для количественного определения TAMRA-положительных клеток в клетках костного мозга или в суспензии клеток колоний 1 млн клеток инкубировали в 400 мкл IMDM с 0.1 мкг TAMRA-ДНК при комнатной температуре в течение 30 мин в темноте. После этого клетки осаждали при 400g в течение 5 мин при 25 °C, промывали небольшим объемом среды и ресуспендировали в конечном объеме среды. В случае клеток c-Kit+/Sca-1+/TAMRA+ оценку окрашивания проводили одновременно.

Для флуоресцентного микроскопического конфокального анализа 5 мкг ангиогенина, меченного Cy5, в форме монопрепарата или в смеси с антителами добавляли к 3 млн клеток костного мозга или колоний, ресуспендированных в 1 мл среды для культивирования клеток в 12-луночном планшете. Через 30–60 мин инкубации клетки анализировали микроскопически с использованием лазерного сканирующего конфокального микроскопа LSM 780 NLO (Zeiss) и программного обеспечения ZenLight. FACS-анализ клеток проводили на проточном цитометре BD FACSAria III в Центре коллективного пользования проточной цитометрии Института цитологии и генетики СО РАН.

Оценка количества ДНК в ГСК. Для инкубации клеток колоний ГСК с Alu повтором человека на 10-й день колонии, полученные после индукции клеток костного мозга препаратом $hDNA^{gr}$, были собраны с двух чашек Петри (диаметр 35 мм) добавлением 8 мл среды IMDM, осаждены центрифугированием при 400g в течение 8 мин, после чего клетки снова были промыты 2 мл среды и осаждены. К полученным клеткам был добавлен фрагмент Alu повтора человека до концентрации 0.23 мкг на 1 млн клеток на 30 мин. Далее клетки были отмыты, осаждены центрифугированием при 400g 5 мин и ресуспендированы в 1 мл PBS.

ПЦР в реальном времени проводили с использованием набора BioMaster RT-qPCR (краситель SYBR Green) (#RM03-200, «Биолабмикс», Россия). Для создания калибровочной кривой были выбраны стандартные праймеры M13 (M13 for: 5′-GTAAA-ACGAC-GGCCA-G-3′, M13 rev: 5′-CAGGA-AACAG-CTATG-AC-3′) и различное количество ДНК Alu-повтора (0–5000 пг). Каждую концентрацию испытывали в трех экземплярах. Линейную зависимость содержания Ct от содержания ДНК Alu строили с помощью программного обеспечения Bio-Rad CFX Manager v3.1.

Обработка клеток костного мозга выбранными индукторами. Клетки костного мозга, выделенные из старых животных и сепаратов костного мозга больных лимфомой Ходжкина, инкубировали с индукторами ($hDNA^{gr}$, ангиогенин или оба индуктора совместно) в течение 1 ч в атмосфере 5 % CO_2, влажности 95 %, 37 °C из расчета: на 3 млн клеток – 500 мкг $hDNA^{gr}$, либо 500 нг ангиогенина, либо 500 мкг $hDNA^{gr}$ и 500 нг ангиогенина совместно в 1 мл среды IMDM без сыворотки. Контрольные клетки костного мозга (без обработок) инкубировали в среде IMDM без сыворотки с добавлением PBS в объеме, равном объему добавленного индуктора к активированным клеткам костного мозга. Термин «индукторы», которым обозначены в исследовании и ДНК, и ангиогенин, используется нами для характеристики любого предполагаемого или ожидаемого ответа ГСК, индуцируемого обработкой этими веществами.

Культивирование клеток костного мозга в метилцеллюлозной среде. Клетки костного мозга после/без активации индукторами переосаждали в течение 10 мин при 400g и ресуспендировали в среде IMDM + 2 % FBS. Для количественного определения и анализа миелоидных предшественников клетки костного мозга мыши помещали в метилцеллюлозную среду MethoCult M3434, а клетки костного мозга крысы и человека – в метилцеллюлозную среду MethoCult H4034 (Stem Cell Technologies). Метилцеллюлозный анализ, подсчет количества колоний и выделение клеток из метилцеллюлозной среды после культивации осуществляли согласно инструкции производителя. Анализ проводили в чашках Петри диаметром 35 мм, хранящихся в процессе развития колоний в чашке Петри большего диаметра с дополнительным увлажнением внутренней атмосферы.

Анализ количества одноцепочечных и двуцепочечных разрывов методом «кометных хвостов». Клетки костного мозга, выделенные из старых мышей и сепаратов костного мозга больных лимфомой Ходжкина, после инкубации с индукторами (ангиогенин, $hDNA^{gr}$, ангиогенин+$hDNA^{gr}$) и без нее культивировали в течение 10–12 дней в метилцеллюлозной среде. При выделении клеток из метилцеллюлозной среды образовавшиеся колонии объединяли и отмывали от среды согласно инструкции производителя. Затем клетки образовавшихся колоний считали в камере Горяева и вновь инкубировали с индукторами. Клетки переосаждали 10 мин при 400g, ресуспендировали в среде IMDM + 2 % FBS, помещали в метилцеллюлозную среду и рассаживали по лункам в 24-луночном планшете. Каждый день в одно и то же время (спустя 24, 48, 72, 96, 120, 144 ч после начала обработки

индукторами) отбирали образец клеток, отмывая их от метилцеллюлозной среды. Клетки колоний заливали в блоки 1 % легкоплавкой агарозы из расчета 5 тыс. клеток на один блок. До начала анализа блоки хранили в 0.5 М ЭДТА при 4 °С. Нулевая точка представляла собой клетки колоний до повторной обработки индукторами.

Перед электрофорезом блоки споласкивали в ТЕ-буфере, инкубировали с лизирующим буфером (50 мМ ЭДТА, 1 % саркозил (Serva, Гейдельберг, Германия), 1 мг/мл протеиназы К (Thermo Fisher Scientific, Уолтем, США)) в течение 20 мин при 50 °С.

Перед нативным электрофорезом блоки окрашивали в течение 10 мин в буфере TAE, содержащем 0.5 мкг/мл бромистого этидия («Медиген», Новосибирск, Россия). Блоки фиксировали на агарозной подложке и фракционировали электрофоретически в 1xTAE буфере при 36 В, 299 мА (камера Model H4 Horizontal Gel Electrophoresis System (BRL, США)) в течение 30 мин.

Щелочной электрофорез проводили в буфере, содержащем 300 мМ NaOH и 1 мМ ЭДТА (pH > 13). Перед щелочным электрофорезом блоки споласкивали в буфере для электрофореза и фиксировали на агарозной подложке. Подложку с блоками помещали в буфер для электрофореза на 30 мин, затем образцы фракционировали при 36 В, 299 мА (камера Model H4 Horizontal Gel Electrophoresis System (BRL, США)) в течение 30 мин. После электрофореза подложку с блоками переносили в нейтральный буфер, содержащий 0.4 М Трис (pH 7.5), на 15 мин. Затем нейтральный буфер меняли на новый, добавляли 1 мкг/мл бромистого этидия и окрашивали подложку с блоками в течение 30 мин.

Подложку с блоками споласкивали дистиллированной водой, делали препараты и высушивали их при 37 °С в течение суток. После сушки препараты 0.5–1 ч отмывали в дистиллированной воде. Микроскопический анализ выполняли в ЦКП микроскопического анализа биологических объектов ИЦиГ СО РАН с использованием Zeiss Axio Imager M2 (Carl Zeiss Microscopy, Оберкохен, Германия). Параметры «кометных хвостов» оценивали в программе CASP (CASP, Вроцлав, Польша) и ImageJ.

Статистический анализ проводили с использованием программного обеспечения Statistica 8 (StatSoft, США). Достоверность различий оценивали с помощью U-критерия Манна–Уитни. Статистическая значимость ($p < 0.05$ или $p < 0.01$) указана в подписях к рисункам.

Результаты

Способность ГСК интернализовать фрагменты дцДНК

В наших недавних работах (Dolgova et al., 2014; Petrova et al., 2022; Ritter et al., 2022) открыто и описано новое общебиологическое свойство стволовых клеток различного генеза. Экспериментально показано, что ГСК мыши, а также все проанализированные низкодифференцированные клетки высших эукариот, включая стволовые опухолевые клетки, способны захватывать фрагменты дцДНК окружающей среды естественным природным механизмом. Взаимодействие экстраклеточных дцДНК с клеткой и их интернализация происходят за счет элементов гликокаликса гликопротеинов/протеогликанов, гликозилфосфатидилинозитол-заякоренных белков и системы скавенджер-рецепторов и осуществляются кавеоло/клатрин-зависимым механизмом. Важнейшей и характерной особенностью является уникальность паттерна гликопротеинов/протеогликанов, гликозилфосфатидилинозитол-заякоренных белков и скавенджер-рецепторов, расположенных на поверхности клетки одного типа. Эта уникальность ограничивается тремя функциональными доменами, составленными различными представителями указанных трех функциональных доменов, а именно молекулами, относящимися к гликопротеинам/протеогликанам, гликозилфосфатидилинозитол-заякоренным белкам и скавенджер-рецепторам. То есть каждая стволовая клетка может иметь как минимум три функциональных домена, которые определяют контакт экстраклеточных двуцепочечных нуклеиновых кислот с клеткой и их интернализацию. Для молекул дцДНК основным сайтом связывания является гепарин-связывающий домен, представленный на различных белках клеточной поверхности или C1q доменом, или гепарин-связывающим доменом, или доменом положительно заряженных аминокислот (Petrova et al., 2022; Ritter et al., 2022).

В настоящей работе дополнительно была проведена FACS и иммунофлуоресцентная оценка способности ГСК человека в сравнении с ГСК мыши интернализовать фрагменты экстраклеточной дцДНК. В качестве фактора сравнения был выбран ангиогенин рекомбинантный человеческий, для которого известны многие молекулярные подробности его влияния на клетку. Также было оценено количество экстраклеточной дцДНК, доставляемое в CD34+ ГСК человека.

Как сказано выше, в процессе интернализации в стволовые клетки основное участие принимают факторы гликокаликса: гликопротеины/протеогликаны, гликозилфосфатидилинозитол-заякоренные белки и система скавенджер-рецепторов. Мы проанализировали последние результаты, содержащие атлас поверхностных маркеров ГСК человека, на предмет присутствия указанных типов белковых молекул (Rix et al., 2022). Показано, что в последовательностях выбранных белков локализованы специфические домены, определяющие интернализацию экстраклеточных фрагментов дцДНК (кластеры позитивно заряженных аминокислотных остатков, гепарин-связывающий домен). Результаты проведенного анализа представлены в таблице. Установлено, что несколько специфических для ГСК поверхностных гликопротеинов, и в первую очередь CD34, содержат требуемые для интернализации домены.

Характеристика прямого молекулярного взаимодействия дцДНК *Alu*-TAMRA/теломерный повтор ($n = 9$) и белков маркеров ГСК CD34 и CD84 (SLAMF5). В наших работах (Petrova et al., 2022; Ritter et al., 2022) провозглашается и доказывается идея, что интернализация дцДНК в стволовые клетки различного генеза связана с развитой на мембранах этих клеток структурой гликокаликса, в состав которого входят протеингликаны-гликопротеины, гликозилфосфатидилинозитол-заякоренные белки и белки системы скавенджер-рецепторов. Предполагается, что взаимодействие с этими белками имеет сложную физическую и молекулярную иерархию и что

Специфические поверхностные белки ГСК человека,
содержащие домены позитивно заряженных аминокислот и гепарин-связывающий домен

Поверхностные маркеры ГСК	Наименование	Положительно заряженные аминокислоты	Гепарин/ДНК связывающие сайты
CD90	Thy-1 membrane glycoprotein	-FSLTRETKKHVLFGTVG-	–
CD34	CD34 molecule	-LVRRGARAGPRMPRGW- -ISSKLQLMKKHQSD-	-EVRPQCLLLVLANRTE-
KIT	KIT proto-oncogene, receptor tyrosine kinase	-FLRRKRDS- -ADKRRSVRIG-	–
VNN2 (GPI-80)	Vanin 2	-EGKLVARYHKVC-	–
SPN (CD43)	Sialophorin	-LLLWRRRQKRRTGA- -FGRRKSRQGS-	-RQKRRTGALVLSRGGKRN-
CD44	CD44 molecule	-ILAVCIAVNSRRRCGQKKKLV-	–
CD9	CD9 molecule	-AIRRNREM-	–
CD48	CD48 molecule	-FESKFKGRVRLD- -GDKRPLPKEL-	–
CD84	CD84 molecule	-TTKRYNLQIYRRLGKPKITQ-	-LFKRRQGRIF- (a-helix)
ITGA6 (CD49f)	Integrin subunit alpha 6	-ESHNSRKKREI-	-TLKRQKQK- -FFKRSRYD-
GPRC5C	G-protein coupled receptor class C group 5 member C	-CGRYKRWRKHGV-	–
PROCR (EPCR)	Protein C receptor	–	–
RET	Ret proto-oncogene	-VSRRARRFA-	-ALRRPKCA-
PROM1 (CD133)	Prominin 1	-QVRTRIKRSRKLA-	-DCKKNRGT
CD59	CD59 molecule	–	–
PTPRC	Protein tyrosine phosphatase receptor type C	-DLHKKRSC- -ELRHSKRKDS-	-LRRQRCL- (a-helix)

Примечание. Зеленым цветом обозначены кластеры положительно заряженных аминокислот, красным – ДНК связывающие сайты, синим – гепарин-связывающие сайты.

именно физический контакт дцДНК и указанных факторов является основой «протаскивания» дцДНК внутрь клетки.

В настоящей серии экспериментов мы попытались оценить возможность прямого физического взаимодействия двух типов молекул: дцДНК и белков-маркеров ГСК. В качестве дцДНК субстрата были выбраны: меченный TAMRA *AluI* зонд, постоянно используемый в лаборатории, и теломерный повтор ($n = 9$) в форме двуцепочечного олигонуклеотида суммарной длиной 54 п.н., меченный P^{32}. В качестве отвечающих белковых факторов были выбраны CD34 и CD84 (SLAMF5). Основные характеристики, имеющие отношение к контакту этих белков и дцДНК, указаны в таблице. Результаты проведенных экспериментов представлены на рис. 1 и подробно описаны в подрисуночной подписи. В данной части исследования впервые в прямом эксперименте показана возможность химического/молекулярного/физического взаимодействия молекулы дцДНК и специфических поверхностных маркеров ГСК CD34 и SLAMF5.

Демонстрация интернализации экстраклеточных дцДНК фрагментов в ГСК (Sca1+ для мыши и CD34+ для человека). Используя флуоресцентный микроскопический и FACS анализ, мы показали присутствие меченого дцДНК зонда в CD34+ клетках костного мозга человека и Sca1 клетках костного мозга мыши. Фактор сравнения – ангиогенин рекомбинантный человеческий – также интернализуют примитивные Sca1 гемопоэтические клетки мыши и CD34+ стволовые клетки человека (Приложение 1)[1]. Проведенная оценка количества доставляемой в CD34+ ГСК человека дцДНК зонда свидетельствует, что ~0.02 % экстраклеточных фрагментов (в пересчете на гаплоидный геном) присутствуют во внутреннем пространстве клеток этого типа в данном эксперименте. Полученные расчеты находятся в согласии с нашими многочисленными количественными оценками, в которых показано, что стволовые клетки различного генеза в зависимости от происхождения и состояния захватывают от 0.01 до 3.0 % фрагментов экстраклеточной дцДНК (в пересчете на гаплоидный геном) (Dolgova et al., 2013, 2016, 2019; Potter et al., 2018, 2024).

Была проведена серия экспериментов, прямо демонстрирующих интернализацию экстраклеточных ДНК фрагментов в ГСК (Sca1+ для мыши и CD34+ для человека), полученных из клеток костного мозга (рис. 2, *A*, *B*). Известно, что процесс интернализации молекул в клетку состоит из фазы мобилизации на цитоплазматической мембране, фазы интернализации и фазы присутствия и процессинга. В этой связи, чтобы избежать спекуляций по поводу детекции в эксперименте не интернализованных, а мобилизованных на цитоплазматической мембране

[1] Приложения 1–6 см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx35.pdf

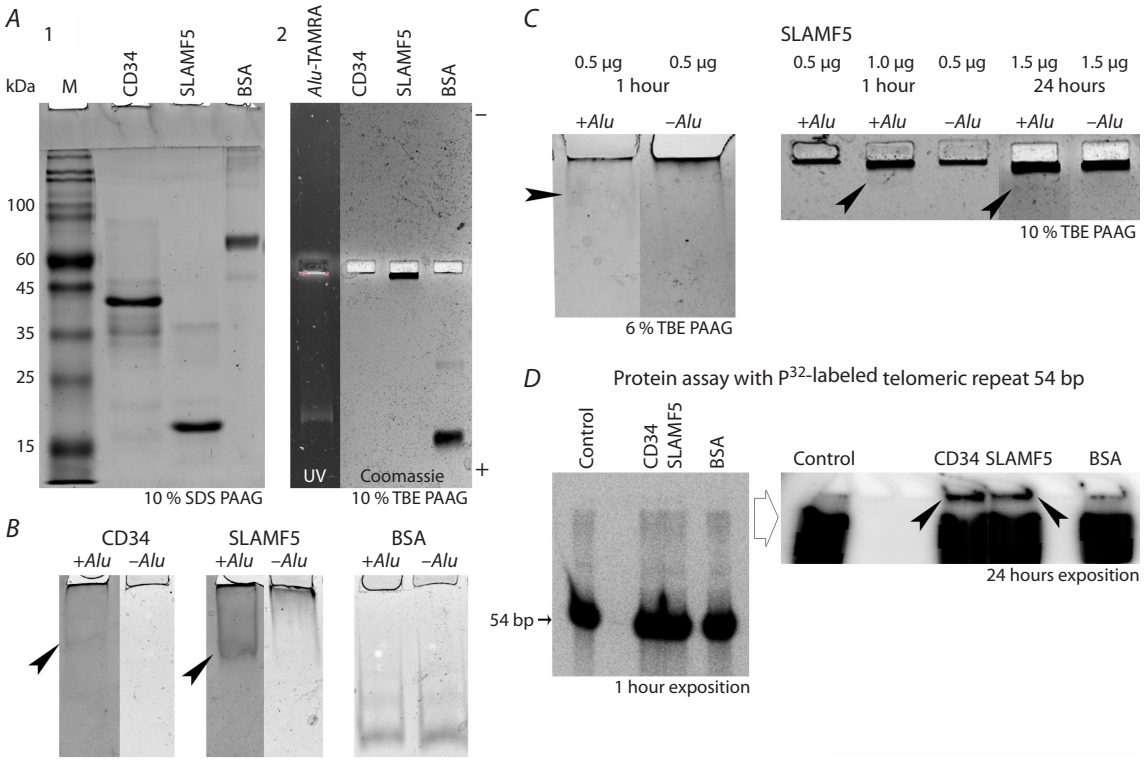


Рис. 1. Анализ прямого молекулярного взаимодействия дцДНК *Alu*-TAMRA/теломерный повтор (n = 9) и белков маркеров ГСК CD34 и CD84 (SLAMF5).

A – электрофоретический анализ миграции анализируемых факторов в 10 % SDS (1) и 10 % нативном трис-боратном горизонтальном (2) полиакриламидном геле. Маркеры ГСК в нативных условиях не имеют электрофоретической подвижности и не входят в гель. На правой панели часть геля с темным полем демонстрирует миграцию дцДНК *AluI* ДНК зонда. *B* – оценка изменения электрофоретической подвижности в геле образцов факторов после формирования комплексов с TAMRA *AluI* ДНК зондом. Детектируется мигрирующая фракция белков (CD34, SLAMF5), что свидетельствует о появлении заряда у белковых молекул, который, по-видимому, определяется молекулой ДНК, с которой молекулы белка образовали физическую связь (отмечено стрелками). С BSA изменение миграции белка не детектируется. *C* – оценка некоторых параметров формирования комплекса между TAMRA *AluI* ДНК зондом и SLAMF5. На левой панели приведена электрофореграмма комплекса ДНК зонда и SLAMF5 в нативном 6 % полиакриламидном геле. Количество белка, нанесенного на гель, идентично в контрольном и экспериментальном образцах. Отчетливо видно формирование мигрирующей фракции белка и одновременное снижение его количества не старте. На правой панели (10 % нативный трис-боратный гель) приведены результаты нескольких режимов формирования комплекса между ДНК зондом и SLAMF5 (отмечено стрелками). Оказалось, что связь белка и ДНК не определяется временем и соотношением молярного количества факторов. Этот факт свидетельствует об отсутствии стехиометрии между TAMRA *AluI* ДНК зондом и SLAMF5. *D* – оценка ДНК-белкового взаимодействия между CD34, SLAMF5 и BSA с использованием P^{32}-меченого двуцепочечного олигонуклеотида, содержащего 9 теломерных повторов (54 п.н.). Специфическое взаимодействие между ДНК и белками четко детектируется в образцах CD34 и SLAMF5 (отмечено стрелками).

молекул ДНК, была разработана и применена процедура подготовки образцов клеток, описанная в Приложении 2.

Можно видеть, что исходные молекулы дцДНК зонда сформировали конкатемерные формы (возможно, циклизованные), содержащие до 6–7 повторов (300–350 п.н.) исходного фрагмента (54 п.н.) (см. рис. 2, *C*, указаны черными стрелками) в клетках, негативных по маркерам ГСК, как для мыши, так и для человека. Этот факт хорошо согласуется с нашими ранними результатами (Dolgova et al., 2013; Potter et al., 2018, 2024). Также четко видно присутствие меченого материала в геномной фракции ДНК мышиной модели.

Настоящее исследование не претендовало на глубокий анализ популяций клеток, обладающих способностью захватывать экстраклеточную ДНК. В исследовании акцент был сделан исключительно на факт интернализации, где показано, аналогично с предыдущими результатами, что CD34+ клетки захватывают внеклеточную ДНК и что в костном мозге присутствует популяция CD34– клеток, также обладающая способностью интернализовать экстраклеточные дцДНК фрагменты, которая может включать любые варианты как мультипотентных прогениторов, так и коммитированных потомков.

Терминальная дифференцировка, пролиферация ГСК и колониеобразование, индуцированные ангиогенином, hDNA^{gr} и совместно обоими индукторами

В работе использовалась геномная фрагментированная до 1–10 нуклеосомных мономеров (физиологический размер ДНК апоптотирующих клеток организма – селф-ДНК, всегда присутствующей в кровяном периферическом русле) депротеинизированная дцДНК человека, носящая название hDNA^{gr}, или геномный реконструктор. В качестве индуктора для сравнения использовался ангиогенин рекомбинантный человеческий.

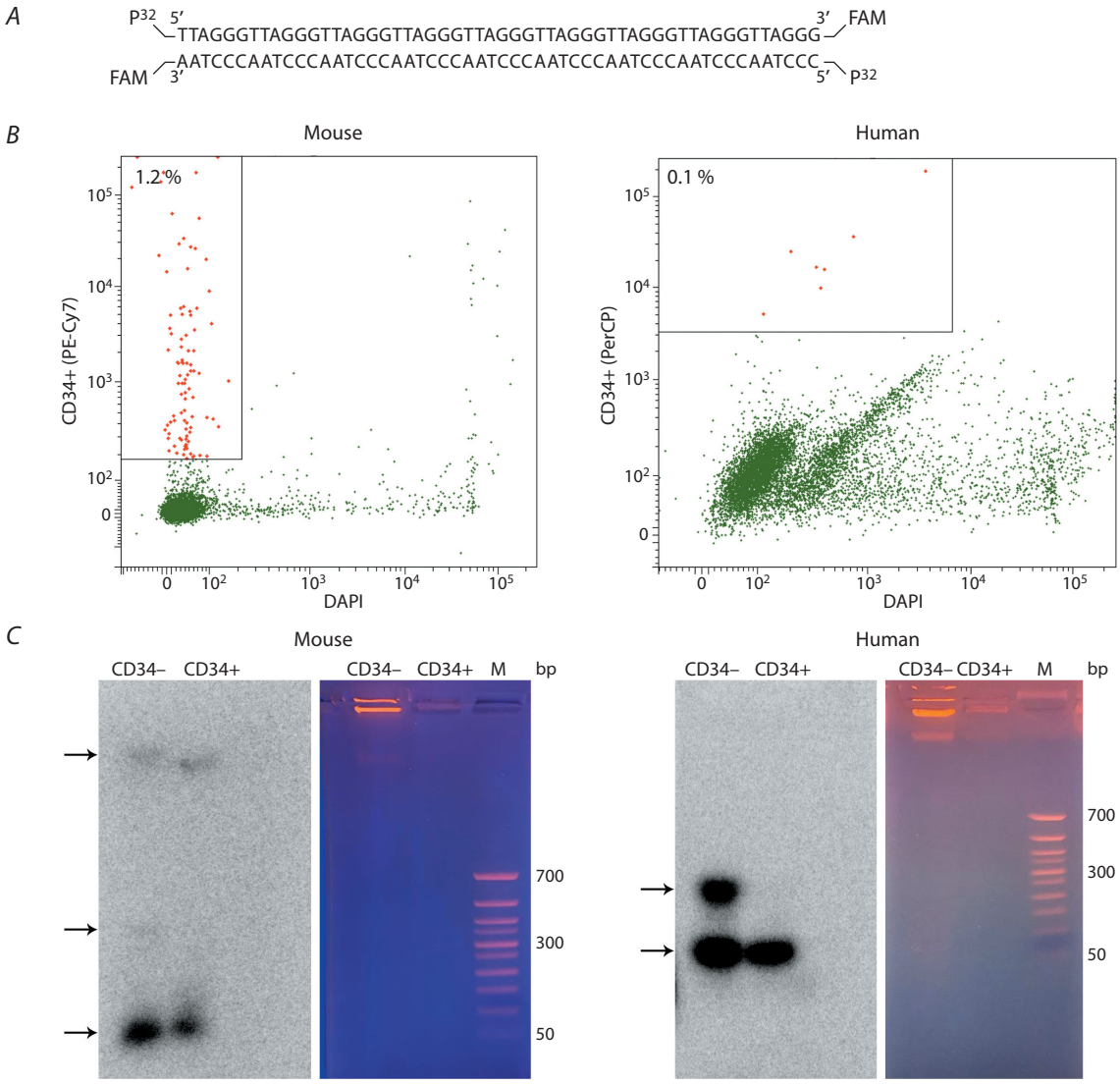


Рис. 2. Оценка интернализации фрагментов экстраклеточной дцДНК в прямом эксперименте по интернализации.

A – структура ДНК зонда; *B* – FACS анализ образцов клеток костного мозга мыши и человека по маркеру CD34; *C* – гель-электрофорез и радиоавтограф материала ДНК, присутствующего во внутренних компартментах сортированных ГСК мыши и человека. Стрелками отмечены бэнды, соответствующие ДНК зонду, конкатемерная (возможно, циклизованная) форма, метка в геномной ДНК.

Была проведена серия экспериментов по анализу стимуляции образования колоний и пролиферативной активности ГСК костного мозга после обработки выбранными индукторами на трех моделях: клетки костного мозга мыши, клетки костного мозга крысы и криоконсервированные клетки костного мозга человека. Установлено, что обработка клеток ангиогенином, hDNA^{gr} и ангиогенином+hDNA^{gr} может стимулировать образование колоний (суммарное количество) у выбранных моделей. Для hDNA^{gr} в экспериментах для мышиной и человеческой моделей количество новых колоний в некоторых случаях увеличивалось на 20–30 % (Приложение 3, рис. 1, *A*, *C*). Достоверное увеличение количества колоний наблюдалось на мышиной модели при обработке ангиогенином и ангиогенином+hDNA^{gr}.

Ангиогенин на мышиной модели достоверно стимулирует пролиферацию клеток в растущих колониях. Основной отвечающий росток – CFU-GM, что достоверно проявляется на человеческой модели. Обработка активаторами клеток сепарата костного мозга не индуцирует апоптоз, не стимулирует выживание CD34+ клеток. Добавление к клеткам сепарата после его размораживания hDNA^{gr} и ангиогенина+hDNA^{gr} усиливает пролиферацию CD34+ клеток; ангиогенин не обладает стимулирующим эффектом и не влияет на способность hDNA^{gr} усиливать пролиферацию CD34+ клеток (см. Приложение 3, рис. 1).

Проведено также сравнение пролиферативной активности CD34+ ГСК по синтезу пролиферативного фактора Ki-67 после обработки индукторами перед высевом на метилцеллюлозу и пролиферативной активности этих клеток, выраженной в количестве клеток на одну колонию после инкубирования на метилцеллюлозе в течение 11–15 сут. Корреляции двух признаков не обнаружено (см. Приложение 3, рис. 2).

Оценка возможности интернализации TAMRA-меченого ПЦР фрагмента 500 п. н. клетками колоний, отобранных на 7-е, 15-е сутки культивирования на метилцеллюлозе

Во всех наших работах лейтмотивом является идея, что экстраклеточные фрагменты ДНК поглощают примитивные стволовые предшественники. Для человека такими клетками являются CD34+ прогениторы. В нашей работе (Potter et al., 2024) на мышиной модели было показано, что в сформированных колониях после индукции терминальной дифференцировки фрагментами экстраклеточной дцДНК количество примитивных прогениторов гемопоэза существенно возрастает. Это позволяет использовать их для анализа различных событий, происходящих с ГСК, что невозможно сделать с ГСК костного мозга.

Аналогичная работа была проведена на модели клеток криоконсервированного костного мозга человека. Оценено процентное содержание CD34+ стволовых клеток в колониях, образованных ГСК после их индукции в составе клеток костного мозга индукторами hDNA^{gr}, ангиогенином и совместно. Оказалось, что при обработке ГСК в составе клеток костного мозга hDNA^{gr} на 15-е сутки культивирования количество таких клеток в колониях увеличивается до 2.7 % *vs* 1.56 % в конкретном эксперименте (клетки костного мозга после стимуляции GM-CSF). То есть в колониях есть достаточное количество клеток, потентных к интернализации экстраклеточного генетического материала, требующихся для уверенной детекции присутствия в клетке экстраклеточного материала ДНК. Ни ангиогенин, ни совместное применение препаратов не увеличивало количество гемопоэтических предшественников в колониях (Приложение 4).

Анализ формирования пангеномных одноцепочечных разрывов в клетках колоний потомков примитивных прогениторов, обработанных индуктором hDNA^{gr} в составе клеток костного мозга

В ранних работах, проанализированных в разделе Введение, было показано, что в процессе коммитирования при индукции терминальной дифференцировки геном эмбриональных стволовых клеток вскрывается пангеномными одноцепочечными разрывами, которые репарируются, не приводя к гибели клетки. Мы полагаем, что этот процесс необходим для смены архитектуры хроматина, характеризующей недифференцированные стволовые клетки крови, на пространственную организацию экспрессирующихся генов, характерную уже для коммитированного потомка (Jacobson et al., 1975; Scher, Friend, 1978; Farzaneh et al., 1982; McMahon et al., 1984; Boerrigter et al., 1989; Kaminskas, Li, 1989; Vatolin et al., 1997).

Было сделано предположение, что это общий процесс для любого типа примитивных прогениторов, в том числе для ГСК. Анализ, проведенный в первой части исследования и в работе (Potter et al., 2024), свидетельствовал, что выбранные индукторы индуцируют колониеобразование и терминальную дифференцировку активированных ими ГСК костного мозга мышей, крыс, человека. Значит, ситуация появления пангеномных одноцепочечных разрывов может тоже быть неотъемлемой частью биологии ГСК. Для мыши доля ГСК колоний составляла 12–15 % (Potter et al., 2024), для человека в колониях количество ГСК увеличивается до ~3 % (см. Приложение 4). Следовательно, при анализе клеток колонии, образованной ГСК костного мозга после ее однократной индукции в составе костного мозга, и после повторной индукции клеток этой колонии на 15-е сутки будет достаточно клеток, сохранивших недифференцированное состояние и претерпевающих терминальную дифференцировку, в которых можно будет определить появление одноцепочечных разрывов.

Работа проводилась на мышиной и человеческой моделях с использованием индукторов hDNA^{gr}, ангиогенина и совместно обоих препаратов. Также было оценено количество одноцепочечных разрывов в ДНК клеток в колониях на 15-е сутки после всех манипуляций, описанных выше.

Проведенный анализ выявил существенные и достоверные отличия в анализируемых параметрах для разных точек образцов и контроля (рис. 3). Идет увеличение количества клеток с максимальным количеством ДНК в хвосте для человеческой модели в 72–96 ч инкубации и для мышиной модели в 96 ч инкубации для клеток, обработанных препаратом hDNA^{gr}. Ангиогенин, как монопрепарат, практически не влияет на индукцию одноцепочечных разрывов и увеличение массы ДНК в хвосте. По-видимому, полное «залечивание» одноцепочечных разрывов для человеческой модели происходит на 7–9-е сутки инкубации (Приложение 5).

Полученные результаты по анализу динамики изменения длины хвостов комет в указанные временные точки позволили оценить ориентировочное количество индуцированных пангеномных одноцепочечных разрывов (рис. 4).

Для оценки количества одноцепочечных разрывов было сделано несколько допущений. Принято, что одна нить ДНК одной хромосомы рвется ником, образуя две равные части. Любые другие варианты предполагали привлечение мощного математического аппарата, что не соответствовало поставленным в работе задачам. Наименьший размер хромосомы составляет приблизительно 50 млн п. н., в этой связи расчет количества разрывов велся на длину наименьшей хромосомы размером ~50 млн п. н. В этом наиболее простом варианте предполагалось, что если нить ДНК рвется на две равные части одним ником (щелочные условия), то длина хвоста, образуемого одной нитью, уменьшается в два раза. Если двумя никами, то каждая из предыдущих частей уменьшается еще в два раза, и т. д. То есть если в первой точке длина хвоста оценивалась как 10 в единицах шкалы, что, как предполагалось, означает отсутствие разрывов или их нативное количество, то длина хвоста, в два раза превышающая длину хвоста первого деления (20), будет соответствовать появлению одного разрыва на исходную длину молекулы (хромосомы). Такой расчет предполагает, что для перехода в следующий интервал необходимо, чтобы все фрагменты ДНК, полученные на предыдущем этапе, получили еще один разрыв. Таким образом, количество разрывов оценивалось по формуле $2n + 1$, где n – количество разрывов предыдущего интервала. На графике толщина кубышки отражает количество клеток в интервале. Количество разрывов, посчитанное для интервала, умножалось на долю клеток в этом интервале, и для указанных моментов времени считалось

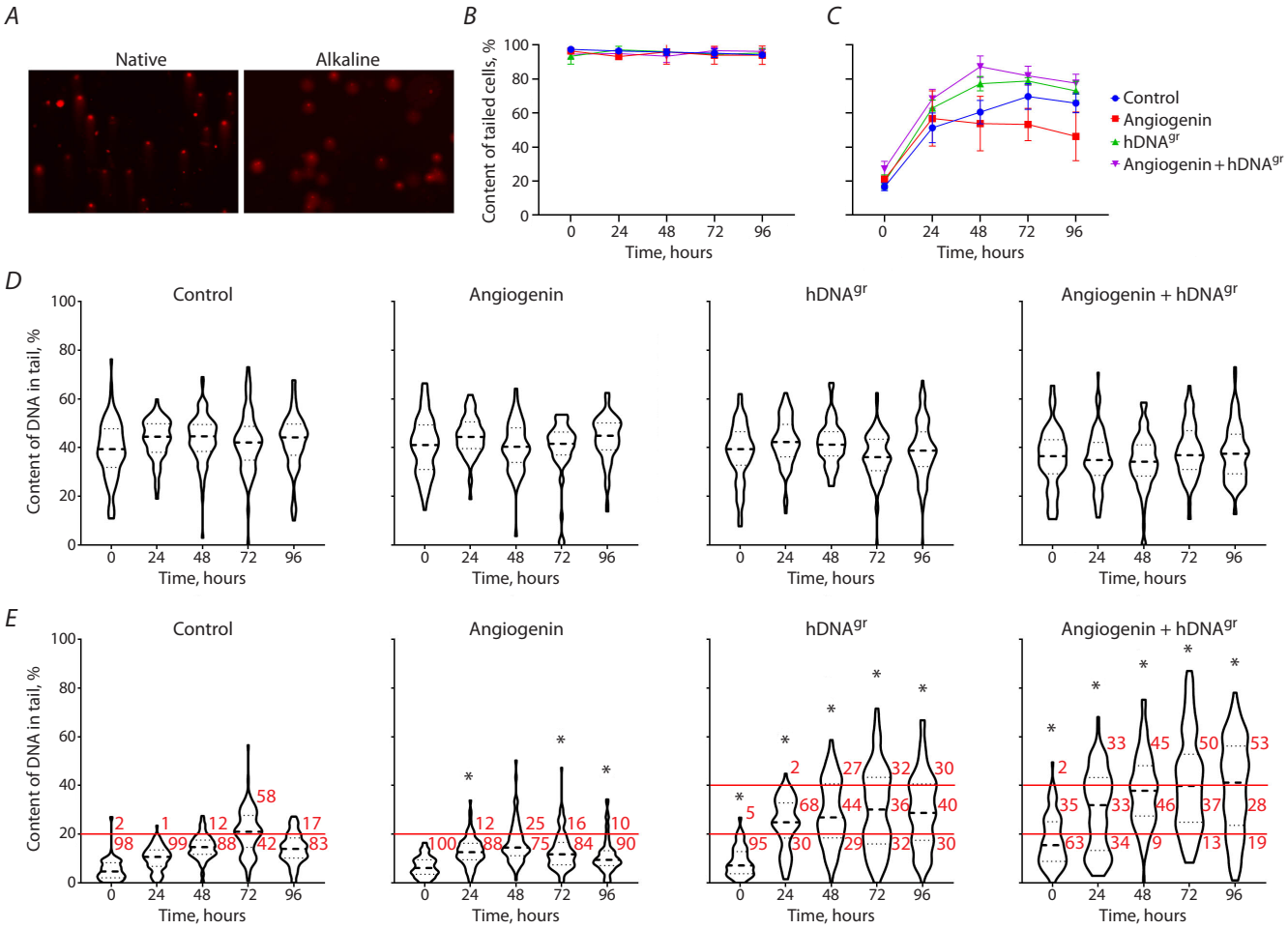


Рис. 3. Человеческая модель. *A* – клетки и кометные хвосты при нативном и щелочном электрофорезе. *B*, *C* – содержание клеток с хвостом при нативном (*B*) и щелочном (*C*) электрофорезе. *D*, *E* – диаграммы, отражающие количество клеток с различным содержанием ДНК в хвосте при нативном (*D*) и щелочном (*E*) электрофорезе.

Жирной штриховой линией обозначена медиана значений, тонким пунктиром – межквартильный диапазон. Красным цветом указано процентное содержание клеток, в которых содержание ДНК в хвосте составляет от 0 до 20 %, от 20 до 40 % и выше 40 % (соответствующие диапазоны выделены красными линиями). * Отличия достоверны по сравнению с контрольной группой, $p < 0.01$, критерий Манна–Уитни.

среднее количество разрывов на клетку, на основе чего был построен график изменения количества разрывов в зависимости от времени.

В результате проведенной оценки установлено, что при указанном варианте подсчета максимальное количество одноцепочечных разрывов приходится на 72–96 ч (для двух независимых экспериментов) и составляет ~2.5–3.5 ника на 50 млн п. н. хроматина. В контрольном образце количество ников находится в интервале 1.0–1.5 на 50 млн п. н. хроматина (см. рис. 4).

Для образца, обработанного ангиогенином, в точке максимальных пертурбаций хроматина, в клетке, может индуцироваться количество ников, незначительно превышающее контрольные значения, что не противоречит результатам по колониестимуляции, где установлено положительное влияние ангиогенина на развитие нескольких типов колоний.

Обсуждение

Обнаруженный факт интернализации фрагментов дцДНК в ГСК и индукции ими терминальной дифференцировки и колониеобразования предполагал, что по аналогии с эмбриональными стволовыми клетками (Vatolin et al., 1997) в гемопоэтических стволовых клетках в состоянии терминальной дифференцировки также будут индуцированы одноцепочечные разрывы. Проведенный на двух выбранных моделях анализ свидетельствовал об аналогичном биологическом явлении, происходящем в ГСК. В состоянии терминальной дифференцировки в ГСК появляются, развиваются и репарируются пангеномные одноцепочечные разрывы. Полученный результат вместе с совокупностью экспериментальных данных, представленных в литературных источниках, означал, что мы имеем дело с процессом общебиологического уровня. Пангеномные одноцепочечные разрывы являются необходимым условием переориентации активности генных платформ, определяющих недифференцированное состояние на генные платформы, характерные для коммитированного состояния клетки.

На протяжении последних двух десятилетий основное внимание исследователей уделялось двуцепочечным разрывам и множеству процессов, связанных с их появлением

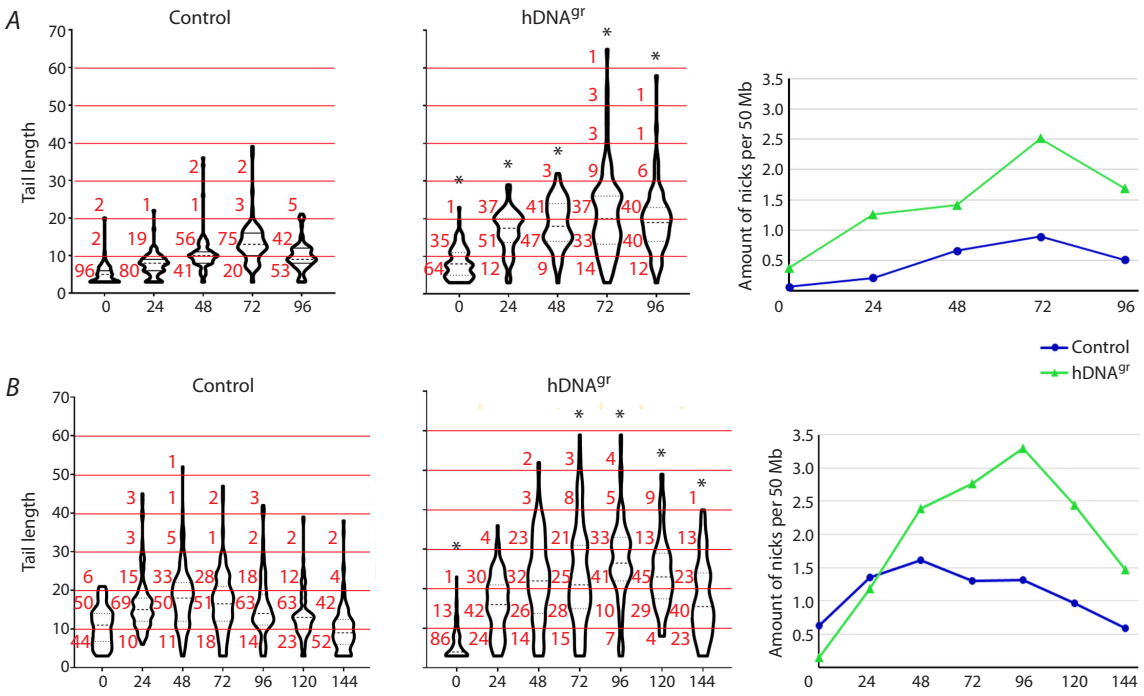


Рис. 4. Результаты двух независимых экспериментов на человеческой модели (*A*, *B*).

Приведены диаграммы для контрольной группы клеток и для клеток, обработанных hDNA^{gr}, отображающие длину кометных хвостов в условных единицах (ось *Y*), временные интервалы с шагом 24 ч (ось *X*). Красным цветом обозначено количество клеток в процентах, длина хвоста которых находится в соответствующем интервале. * Отличия достоверны по сравнению с контрольной группой, $p < 0.01$, критерий Манна–Уитни. Справа приведены графики зависимости рассчитанного количества ников на 50 млн п. н. (ось *Y*) от временного интервала (ось *X*).

и индукцией ими репарационно-рекомбинационных событий в клетке (So et al., 2017). Тем не менее за последние несколько лет в научном сообществе вновь возник интерес к никам или одноцепочечным разрывам хромосом, что отражено в том числе в нескольких обзорных работах (Xu, 2015; Vriend, Krawczyk, 2017; Maizels, Davis, 2018; Zilio, Ulrich, 2021). Лейтмотивом нового внимания к никированной ДНК хроматина стала забытая идея ник-инициированной гомологичной рекомбинации. Проведенный анализ говорит о том, что ники представляют собой не менее значимые, чем двуцепочечные разрывы, интермедиаты метаболизма молекулы ДНК хроматина, индуцирующие репарационно-рекомбинационные процессы в клетке. Однако в отличие от двуцепочечных разрывов репарация одноцепочечных разрывов (ников) значительно реже приводит к фатальным изменениям структуры генома. Основной механизм репарации одноцепочечных разрывов – гомологичная рекомбинация.

Сказанное означает, что одноцепочечные разрывы индуцируют развитие в клетке рекомбиногенной ситуации. Идея рекомбиногенной ситуации максимально полно изложена в нашем пионерном обзоре (Лихачева и др., 2008). Термин «рекомбиногенная ситуация» характеризует активность молекулярной машины клетки, запущенной изменением архитектуры хроматина высшего порядка, одним из индукторов которого являются одноцепочечные разрывы.

Основной посыл высказанных в упомянутом обзоре мыслей заключается в том, что если в клетке активирована «рекомбиногенная ситуация» и если в клетке в этот момент времени присутствуют интернализованные из внеклеточного пространства фрагменты экстраклеточной дцДНК, то указанные фрагменты становятся естественными участниками репарационно-рекомбинационного процесса, запущенного активированными молекулярными механизмами. Это означает, что экстраклеточные фрагменты могут участвовать в рекомбинационном процессе в качестве естественного субстрата для рекомбинации. Таким образом, можно полагать, что найден общебиологический механизм, объясняющий возможность появления в реципиентном геноме внехромосомной генетической информации в результате или прямой гомологичной интеграции экстраклеточных фрагментов дцДНК в реципиентный геном, или формирования стабильных, «генетически» активных экстрахромосомных ассоциатов.

В наших работах охарактеризовано два механизма с участием фрагментов дцДНК в репаративно-рекомбинационных событиях. Так, в статье (Likhacheva et al., 2007) показано участие экзогенной ДНК человека в спасении костномозговых прогениторов мыши от летальной дозы гамма-радиации, что сопровождалось выживанием экспериментальных животных. Сохранение жизнеспособности ГСК связано с интернализацией фрагментов дцДНК в стволовую клетку крови и репаративно-рекомбинационной коррекцией двуцепочечных разрывов хроматина, индуцированных жестким облучением. В серии других наших исследований продемонстрировано участие экстраклеточной дцДНК в процессах блокирования репарации межцепочечных сшивок в стволовых опухолевых клетках. В результате такого участия стволовая опухолевая клет-

ка не в состоянии завершить репарацию повреждений хроматина, индуцированных цитостатиком, и уходит в апоптотическую гибель (Ruzanova et al., 2022). Другие многочисленные исследования свидетельствуют, что одноцепочечные разрывы индуцируют гомологичную рекомбинацию генетического материала ядра клетки (Vriend, Krawczyk, 2017; Maizels, Davis, 2018).

Заключение

Таким образом, экстраклеточные дцДНК фрагменты интернализуются в ГСК естественным природным механизмом, индуцируют терминальную дифференцировку стволовых клеток крови и стимулируют колониеобразование. Молекулярной манифестацией этих процессов являются пангеномные одноцепочечные разрывы. Возникновение пангеномных одноцепочечных разрывов индуцирует в стволовой клетке крови рекомбиногенную ситуацию. В ходе этого процесса возможна интеграция фрагментов экстраклеточной дцДНК в реципиентный геном ГСК. Теоретически вероятны несколько вариантов интеграции, а именно: по механизму ends in/ends out, гомологичного реципрокного обмена, в результате генной конверсии или однонитевого отжига, негомологичной интеграции (Rubnitz, Subramani, 1984; Hastings et al., 1993; Li et al., 2001; Langston, Symington, 2004; Chen J.M. et al., 2007; Rass et al., 2012).

В следующих частях цикла исследования мы приводим экспериментальные доказательства как интеграции экстраклеточных фрагментов дцДНК в геном ГСК, так и формирования кольцевых структур, которые образуют комплексы с ДНК хромосом, сохраняющиеся при жестких условиях фракционирования. В Приложении 6 высказываются соображения по поводу событий, связанных с терминальной дифференцировкой ГСК после интернализации экстраклеточной дцДНК, и очевидного противоречия с данными проточной цитометрии, где в колониях CD34+ ГСК не только не исчезают, но их количество увеличивается по сравнению с исходным образцом клеток костного мозга (см. Приложение 4).

Список литературы / References

Лихачева А.С., Рогачев В.А., Николин В.П., Попова Н.А., Шилов А.Г., Себелева Т.Е., Струнки Д.Н., Черных Е.Р., Гельфгат Е.Л., Богачев С.С., Шурдов М.А. Участие экзогенной ДНК в молекулярных процессах, протекающих в соматической клетке. *Информ. вестн. ВОГиС*. 2008;12(3):426-473
[Likhacheva A.S., Rogachev V.A., Nikolin V.P., Popova N.A., Shilov A.G., Sebeleva T.E., Strunkin D.N., Chernykh E.R., Gel'fgat E.L., Bogachev S.S., Shurdov M.A. Involvement of exogenous DNA in the molecular processes in somatic cell. *Informatsionnyy Vestnik VOGiS = The Herald of Vavilov Society for Geneticists and Breeders.* 2008;12(3):426-473 (in Russian)]

Маниатис Т., Фрич Э., Сэмбрук Д. Методы генетической инженерии. Молекулярное клонирование. М.: Мир, 1984
[Maniatis T., Fritch E., Sambrook D. Methods of Genetic Engineering. Molecular Cloning. Moscow: Mir Publ., 1984 (in Russian)]

Boerrigter M.E.T.I., Mullaart E., Van Der Schans G.P., Vijg J. Quiescent human peripheral blood lymphocytes do not contain a sizable amount of preexistent DNA single-strand breaks. *Exp. Cell Res*. 1989;180(2):569-573. doi 10.1016/0014-4827(89)90085-2

Chen J.M., Cooper D.N., Chuzhanova N., Férec C., Patrinos G.P. Gene conversion: mechanisms, evolution and human disease. *Nat. Rev. Genet*. 2007;8(10):762-775. doi 10.1038/NRG2193

Chen T.L., Chiang Y.W., Lin G.L., Chang H.H., Lien T.S., Sheh M.H., Sun D.S. Different effects of granulocyte colony-stimulating factor and erythropoietin on erythropoiesis. *Stem. Cell Res. Ther*. 2018; 9(1):119. doi 10.1186/S13287-018-0877-2

Desterke C., Bennaceur-Griscelli A., Turhan A.G. EGR1 dysregulation defines an inflammatory and leukemic program in cell trajectory of human-aged hematopoietic stem cells (HSC). *Stem. Cell Res. Ther.* 2021;12(1):419. doi 10.1186/S13287-021-02498-0

DNA Cloning. A practical approach. Ed. D.M. Glover. IRL Press, 1985

Dolgova E.V., Efremov Y.R., Orishchenko K.E., Andrushkevich O.M., Alyamkina E.A., Proskurina A.S., Bayborodin S.I., Nikolin V.P., Popova N.A., Chernykh E.R., Ostanin A.A., Taranov O.S., Omigov V.V., Minkevich A.M., Rogachev V.A., Bogachev S.S., Shurdov M.A. Delivery and processing of exogenous double-stranded DNA in mouse CD34+ hematopoietic progenitor cells and their cell cycle changes upon combined treatment with cyclophosphamide and double-stranded DNA. *Gene*. 2013;528(2):74-83. doi 10.1016/j.gene.2013.06.058

Dolgova E.V., Alyamkina E.A., Efremov Y.R., Nikolin V.P., Popova N.A., Tyrinova T.V., Kozel A.V., Minkevich A.M., Andrushkevich O.M., Zavyalov E.L., Romaschenko A.V., Bayborodin S.I., Taranov O.S., Omigov V.V., Shevela E.Y., Stupak V.V., Mishinov S.V., Rogachev V.A., Proskurina A.S., Mayorov V.I., Shurdov M.A., Ostanin A.A., Chernykh E.R., Bogachev S.S. Identification of cancer stem cells and a strategy for their elimination. *Cancer Biol. Ther*. 2014;15(10):1378-1394. doi 10.4161/cbt.29854

Dolgova E.V., Potter E.A., Proskurina A.S., Minkevich A.M., Chernych E.R., Ostanin A.A., Efremov Y.R., Bayborodin S.I., Nikolin V.P., Popova N.A., Kolchanov N.A., Bogachev S.S. Properties of internalization factors contributing to the uptake of extracellular DNA into tumor-initiating stem cells of mouse Krebs-2 cell line. *Stem. Cell Res. Ther*. 2016;7(1):76. doi 10.1186/s13287-016-0338-8

Dolgova E.V., Petrova D.D., Proskurina A.S., Ritter G.S., Kisaretova P.E., Potter E.A., Efremov Y.R., Bayborodin S.I., Karamysheva T.V., Romanenko M.V., Netesov S.V., Taranov O.S., Ostanin A.A., Chernykh E.R., Bogachev S.S. Identification of the xenograft and its ascendant sphere-forming cell line as belonging to EBV-induced lymphoma, and characterization of the status of sphere-forming cells. *Cancer Cell Int*. 2019;19:120. doi 10.1186/S12935-019-0842-X

Farzaneh F., Zalin R., Brill D., Shall S. DNA strand breaks and ADP-ribosyl transferase activation during cell differentiation. *Nature*. 1982;300(5890):362-366. doi 10.1038/300362A0

Forristal C.E., Levesque J.-P. Targeting the hypoxia-sensing pathway in clinical hematology. *Stem Cells Transl. Med.* 2014;3(2):135-140. doi 10.5966/SCTM.2013-0134

Goncalves K.A., Silberstein L., Li S., Severe N., Hu M.G., Yang H., Scadden D.T., Hu G.F. Angiogenin promotes hematopoietic regeneration by dichotomously regulating quiescence of stem and progenitor cells. *Cell*. 2016;166(4):894-906. doi 10.1016/J.CELL.2016.06.042

Hastings P.J., McGill C., Shafer B., Strathern J.N. Ends-in *vs*. ends-out recombination in yeast. *Genetics*. 1993;135(4):973-980. doi 10.1093/GENETICS/135.4.973

Iseki S. DNA strand breaks in rat tissues as detected by in situ nick translation. *Exp. Cell Res*. 1986;167(2):311-326. doi 10.1016/0014-4827(86)90172-2

Jacobson G.K., Pinon R., Esposito R.E., Esposito M.S. Single-strand scissions of chromosomal DNA during commitment to recombination at meiosis. *Proc. Natl. Acad. Sci. USA*. 1975;72(5):1887-1891. doi 10.1073/PNAS.72.5.1887

Jiang N., Pisetsky D.S. The effect of inflammation on the generation of plasma DNA from dead and dying cells in the peritoneum. *J. Leukoc. Biol*. 2005;77(3):296-302. doi 10.1189/JLB.0704411

Johnstone A.P., Williams G.T. Role of DNA breaks and ADP-ribosyl transferase activity in eukaryotic differentiation demonstrated in human lymphocytes. *Nature*. 1982;300(5890):368-370. doi 10.1038/300368A0

Kaminskas E., Li J.C. DNA fragmentation in permeabilized cells and nuclei. The role of (Ca^{2+} + Mg^{2+})-dependent endodeoxyribonuclease. *Biochem. J*. 1989;261(1):17-21. doi 10.1042/BJ2610017

Kananen L., Hurme M., Bürkle A., Moreno-Villanueva M., Bernhardt J., Debacq-Chainiaux F., Grubeck-Loebenstein B., Malavolta M., Basso A., Piacenza F., Collino S., Gonos E.S., Sikora E., Gradinaru D., Jansen E.H.J.M., Dollé M.E.T., Salmon M., Stuetz W., Weber D., Grune T., Breusing N., Simm A., Capri M., Franceschi C., Slagboom E., Talbot D., Libert C., Raitanen J., Koskinen S., Härkänen T., Stenholm S., Ala-Korpela M., Lehtimäki T., Raitakari O.T., Ukkola O., Kähönen M., Jylhä M., Jylhävä J. Circulating cell-free DNA in health and disease – the relationship to health behaviours, ageing phenotypes and metabolomics. *GeroScience*. 2023;45(1): 85-103. doi 10.1007/S11357-022-00590-8

Kiang J.G., Zhai M., Lin B., Smith J.T., Anderson M.N., Jiang S. Co-therapy of pegylated G-CSF and ghrelin for enhancing survival after exposure to lethal radiation. *Front. Pharmacol*. 2021;12: 628018. doi 10.3389/FPHAR.2021.628018

Korabecna M., Zinkova A., Brynychova I., Chylikova B., Prikryl P., Sedova L., Neuzil P., Seda O. Cell-free DNA in plasma as an essential immune system regulator. *Sci. Rep*. 2020;10(1):17478. doi 10.1038/S41598-020-74288-2

Kovtonyuk L.V., Fritsch K., Feng X., Manz M.G., Takizawa H. Inflamm-aging of hematopoiesis, hematopoietic stem cells, and the bone marrow microenvironment. *Front. Immunol*. 2016;7:502. doi 10.3389/FIMMU.2016.00502

Kulkarni R., Kale V. Physiological cues involved in the regulation of adhesion mechanisms in hematopoietic stem cell fate decision. *Front. Cell Dev. Biol.* 2020;8:611. doi 10.3389/FCELL.2020.00611

Kumar S., Geiger H. HSC niche biology and HSC expansion *ex vivo*. *Trends Mol. Med*. 2017;23(9):799. doi 10.1016/J.MOLMED.2017.07.003

Langston L.D., Symington L.S. Gene targeting in yeast is initiated by two independent strand invasions. *Proc. Natl. Acad. Sci. USA*. 2004; 101(43):15392-15397. doi 10.1073/PNAS.0403748101

Lauková L., Bertolo E.M.J., Zelinková M., Borbélyová V., Čonka J., Gaál Kovalčíková A., Domonkos E., Vlková B., Celec P. Early dynamics of plasma DNA in a mouse model of sepsis. *Shock*. 2019; 52(2):257-263. doi 10.1097/SHK.0000000000001215

Lévesque J.P., Helwani F.M., Winkler I.G. The endosteal 'osteoblastic' niche and its role in hematopoietic stem cell homing and mobilization. *Leukemia.* 2010;24(12):1979-1992. doi 10.1038/leu.2010.214

Li J., Read L.R., Baker M.D. The mechanism of mammalian gene replacement is consistent with the formation of long regions of heteroduplex DNA associated with two crossing-over events. *Mol. Cell. Biol*. 2001;21(2):501-510. doi 10.1128/MCB.21.2.501-510.2001

Likhacheva A.S., Nikolin V.P., Popova N.A., Rogachev V.A., Prokhorovich M.A., Sebeleva T.E., Bogachev S.S., Shurdov M.A. Exogenous DNA can be captured by stem cells and be involved in their rescue from death after lethal-dose γ-radiation. *Gene Therapy Mol. Biol*. 2007;11:305-314

Lucas D. Leukocyte trafficking and regulation of murine hematopoietic stem cells and their niches. *Front. Immunol*. 2019;10:387. doi 10.3389/FIMMU.2019.00387/BIBTEX

Maizels N., Davis L. Initiation of homologous recombination at DNA nicks. *Nucleic Acids Res*. 2018;46:6962-6973. doi 10.1093/NAR/GKY588

McMahon G., Alsina J.L., Levy S.B. Induction of a Ca^{2+}, Mg^{2+}-dependent endonuclease activity during the early stages of murine erythroleukemic cell differentiation. *Proc. Natl. Acad. Sci. USA*. 1984; 81(23):7461-7465. doi 10.1073/PNAS.81.23.7461

Mendelson A., Frenette P.S. Hematopoietic stem cell niche maintenance during homeostasis and regeneration. *Nat. Med*. 2014;20(8): 833-846. doi 10.1038/NM.3647

Morita Y., Ema H., Nakauchi H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. *J. Exp. Med*. 2010;207(6):1173-1182. doi 10.1084/JEM.20091318

Muller-Sieburg C., Sieburg H.B. Stem cell aging: survival of the laziest? *Cell Cycle*. 2008;7(24):3798-3804. doi 10.4161/CC.7.24.7214

Patkin E.L., Kustova M.E., Noniashvili E.M. DNA-strand breaks in chromosomes of early mouse embryos as detected by *in situ* nick translation and gap filling. *Genome*. 1995;38:381-384. doi 10.1139/G95-049

Petrova D.D., Dolgova E.V., Proskurina A.S., Ritter G.S., Ruzanova V.S., Efremov Y.R., Potter E.A., Kirikovich S.S., Levites E.V., Taranov O.S., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. The new general biological property of stem-like tumor cells (Part II: Surface molecules, which belongs to distinctive groups with particular functions, form a unique pattern characteristic of a certain type of tumor stem-like cells). *Int. J. Mol. Sci.* 2022; 23(24):15800. doi 10.3390/ijms232415800

Pierce H., Zhang D., Magnon C., Lucas D., Christin J.R., Huggins M., Schwartz G.J., Frenette P.S. Cholinergic signals from the CNS regulate G-CSF-mediated HSC mobilization from bone marrow via a glucocorticoid signaling relay. *Cell Stem. Cell*. 2017;20:648-658.e4. doi 10.1016/J.STEM.2017.01.002

Pinho S., Frenette P.S. Haematopoietic stem cell activity and interactions with the niche. *Nat. Rev. Mol. Cell Biol*. 2019;20(5):303-320. doi 10.1038/S41580-019-0103-9

Potter E.A., Proskurina A.S., Ritter G.S., Dolgova E.V., Nikolin V.P., Popova N.A., Taranov O.S., Efremov Y.R., Bayborodin S.I., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. Efficacy of a new cancer treatment strategy based on eradication of tumor-initiating stem cells in a mouse model of Krebs-2 solid adenocarcinoma. *Oncotarget*. 2018;9(47):28486-28499. doi 10.18632/oncotarget.25503

Potter E.A., Dolgova E.V., Proskurina A.S., Ruzanova V.S., Efremov Y.R., Kirikovich S.S., Oshikhmina S.G., Mamaev A.L., Taranov O.S., Bryukhovetskiy A.S., Grivtsova L.U., Kolchanov N.A., Ostanin A.A., Chernykh E.R., Bogachev S.S. Stimulation of mouse hematopoietic stem cells by angiogenin and DNA preparations. *Braz. J. Med. Biol. Res*. 2024;57:e13072. doi 10.1590/1414-431X2024E13072

Pulito V.L., Miller D.L., Sassa S., Yamane T. DNA fragments in Friend erythroleukemia cells induced by dimethyl sulfoxide. *Proc. Natl. Acad. Sci. USA*. 1983;80(19):5912-5915. doi 10.1073/PNAS.80.19.5912

Rass E., Grabarz A., Bertrand P., Lopez B.S. Double strand break repair, one mechanism can hide another: alternative non-homologous end joining. *Cancer Radiother.* 2012;16:1-10. doi 10.1016/J.CANRAD.2011.05.004

Redondo P.A., Pavlou M., Loizidou M., Cheema U. Elements of the niche for adult stem cell expansion. *J. Tissue Eng*. 2017;8: 2041731417725464. doi 10.1177/2041731417725464

Ritter G.S., Dolgova E.V., Petrova D.D., Efremov Y.R., Proskurina A.S., Potter E.A., Ruzanova V.S., Kirikovich S.S., Levites E.V., Taranov O.S., Ostanin A.A., Chernykh E.R., Kolchanov N.A., Bogachev S.S. The new general biological property of stem-like tumor cells. Part I. Peculiarities of the process of the double-stranded DNA fragments internalization into stem-like tumor cells. *Front. Genetics*. 2022;13:954395. doi 10.3389/fgene.2022.954395

Rix B., Maduro A.H., Bridge K.S., Grey W. Markers for human haematopoietic stem cells: the disconnect between an identification marker and its function. *Front. Physiol*. 2022;13. doi 10.3389/FPHYS.2022.1009160

Rubnitz J., Subramani S. The minimum amount of homology required for homologous recombination in mammalian cells. *Mol. Cell. Biol*. 1984;4(11):2253-2258. doi 10.1128/MCB.4.11.2253-2258.1984

Ruzanova V., Proskurina A., Efremov Y., Kirikovich S., Ritter G., Levites E., Dolgova E., Potter E., Babaeva O., Sidorov S., Taranov O., Ostanin A., Chernykh E., Bogachev S. Chronometric administration of cyclophosphamide and a double-stranded DNA-Mix at interstrand crosslinks repair timing, called "Karanahan" therapy, is

highly efficient in a weakly immunogenic Lewis carcinoma model. *Pathol. Oncol. Res*. 2022;28. doi 10.3389/PORE.2022.1610180

Saitoh T., Fujita N., Yoshimori T., Akira S. Regulation of dsDNA-induced innate immune responses by membrane trafficking. *Autophagy*. 2010;6:430-432. doi 10.4161/AUTO.6.3.11611

Scharf P., Broering M.F., da Rocha G.H.O., Farsky S.H.P. Cellular and molecular mechanisms of environmental pollutants on hematopoiesis. *Int. J. Mol. Sci*. 2020;21(19):6996. doi 10.3390/IJMS21196996

Scher W., Friend C. Breakage of DNA and alterations in folded genomes by inducers of differentiation in Friend erythroleukemic cells. *Cancer Res*. 1978;38:841-849

Seita J., Weissman I.L. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med*. 2010;2(6): 640-653. doi 10.1002/WSBM.86

Silberstein L., Goncalves K.A., Kharchenko P.V., Turcotte R., Kfoury Y., Mercier F., Baryawno N., Severe N., Bachand J., Spencer J.A., Papazian A., Lee D., Chitteti B.R., Srour E.F., Hoggatt J., Tate T., Lo Celso C., Ono N., Nutt S., Heino J., Sipilä K., Shioda T., Osawa M., Lin C.P., Hu G.-fu, Scadden D.T. Proximity-based differential single-cell analysis of the niche to identify stem/progenitor cell regulators. *Cell Stem Cell*. 2016;19(4):530-543. doi 10.1016/J.STEM.2016.07.004

So A., Le Guen T., Lopez B.S., Guirouilh-Barbat J. Genomic rearrangements induced by unscheduled DNA double strand breaks in somatic mammalian cells. *FEBS J*. 2017;284(15):2324-2344. doi 10.1111/FEBS.14053

Szade K., Gulati G.S., Chan C.K.F., Kao K.S., Miyanishi M., Marjon K.D., Sinha R., George B.M., Chen J.Y., Weissman I.L. Where hematopoietic stem cells live: the bone marrow niche. *Antioxid. Redox Signal.* 2018;29:191. doi 10.1089/ARS.2017.7419

Vatolin S.Y., Okhapkina E.V., Matveeva N.M., Shilov A.G., Baiborodin S.I., Philimonenko V.V., Zhdanova N.S., Serov O.L. Scheduled perturbation in DNA during in vitro differentiation of mouse embryo-derived cells. *Mol. Reprod. Dev.* 1997;47(1):1-10. doi 10.1002/(SICI)1098-2795(199705)47:1<1::AID-MRD1>3.0.CO;2-R

Vriend L.E.M., Krawczyk P.M. Nick-initiated homologous recombination: protecting the genome, one strand at a time. *DNA Repair*. 2017; 50:1-13. doi 10.1016/J.DNAREP.2016.12.005

Wang S., Zhang Y., Meng W., Dong Y., Zhang S., Teng L., Liu Y., Li L., Wang D. The involvement of macrophage colony stimulating factor on protein hydrolysate injection mediated hematopoietic function improvement. *Cells*. 2021;10(10):2776. doi 10.3390/CELLS 10102776

Wilkinson A.C., Igarashi K.J., Nakauchi H. Haematopoietic stem cell self-renewal in vivo and ex vivo. *Nat. Rev. Genet.* 2020;21(9):541-554. doi 10.1038/s41576-020-0241-0

Winkler I.G., Barbier V., Nowlan B., Jacobsen R.N., Forristal C.E., Patton J.T., Magnani J.L., Lévesque J.P. Vascular niche E-selectin regulates hematopoietic stem cell dormancy, self renewal and chemoresistance. *Nat. Med*. 2012;18(11):1651-1657. doi 10.1038/NM.2969

Xu S.Y. Sequence-specific DNA nicking endonucleases. *Biomol. Concepts.* 2015;6(4):253-267. doi 10.1515/BMC-2015-0016

Zhang C.C., Sadek H.A. Hypoxia and metabolic properties of hematopoietic stem cells. *Antioxid. Redox. Signal.* 2014;20(12):1891-1901. doi 10.1089/ARS.2012.5019

Zilio N., Ulrich H.D. Exploring the SSBreakome: genome-wide mapping of DNA single-strand breaks by next-generation sequencing. *FEBS J*. 2021;288(13):3948-3961. doi 10.1111/FEBS.15568

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 26.06.2024. После доработки 19.09.2024. Принята к публикации 26.09.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):1008-1017

doi 10.18699/vjgb-24-107

База данных о генах и белках, ассоциированных с нарушениями метаболизма глюкозы (GlucoGenes®): описание и возможности применения в биоинформатических исследованиях

В.В. Климонтов [1] ✉, К.С. Шишин [1], Р.А. Иванов [2], М.П. Пономаренко [2], К.А. Золотарева [2], С.А. Лашин [1, 2]

[1] Научно-исследовательский институт клинической и экспериментальной лимфологии – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ klimontov@mail.ru

Аннотация. Данные в области генетики и молекулярной биологии сахарного диабета стремительно накапливаются. Это ставит задачу создания исследовательских инструментов для быстрого поиска, структурирования и анализа информации в этой области. Мы разработали базу данных о генах и белках человека, ассоциированных с высоким уровнем глюкозы (гипергликемией), низким уровнем глюкозы (гипогликемией) и обоими нарушениями. Сведения были собраны с помощью текст-майнинга научных публикаций, проиндексированных в PubMed и PubMed Central, и анализа генных сетей гипергликемии, гипогликемии и вариабельности гликемии, выполненного с помощью биоинформатической системы ANDSystems. Созданный ресурс (GlucoGenes®) доступен по адресу: https://glucogenes.sysbio.ru/genes/main. Ресурс предоставляет информацию о генах и белках, связанных с риском развития гипергликемии и гипогликемии; регуляторных молекулах с гипергликемической и антигипергликемической активностью; генах, экспрессия которых повышается при высоком и/или низком уровне глюкозы; генах, экспрессия которых снижается при высоком и/или низком уровне глюкозы, а также о молекулах, связанных с нарушениями метаболизма глюкозы иным образом. На основе ресурса проведен эволюционный анализ генов, ассоциированных с нарушениями метаболизма глюкозы. Результаты анализа выявили значительное увеличение (до 40 %) доли генов, имеющих филостратиграфический индекс (phylostratigraphy age index, PAI), соответствующий времени происхождения многоклеточных организмов. Анализ консервативности последовательностей белков по индексу дивергенции (divergency index, DI) показал, что большинство соответствующих генов высококонсервативны (DI < 0.6) или консервативны (DI < 1). При анализе однонуклеотидного полиморфизма (SNP) в проксимальных районах промоторов, влияющих на сродство TATA-связывающего белка, в базе данных GlucoGenes® найден 181 SNP-маркер, который может снижать (45 SNP-маркеров) или повышать (136 SNP-маркеров) экспрессию 52 генов. Мы полагаем, что разработанный ресурс станет полезным инструментом для дальнейших исследований в области молекулярной биологии диабета.

Ключевые слова: ген; белок; сахарный диабет; гипергликемия; гипогликемия; вариабельность глюкозы; база данных; филостратиграфический индекс; однонуклеотидный полиморфизм.

Для цитирования: Климонтов В.В., Шишин К.С., Иванов Р.А., Пономаренко М.П., Золотарева К.А., Лашин С.А. База данных о генах и белках, ассоциированных с нарушениями метаболизма глюкозы (GlucoGenes®): описание и возможности применения в биоинформатических исследованиях. *Вавиловский журнал генетики и селекции*. 2024;28(8):1008-1017. doi 10.18699/vjgb-24-107

Финансирование. Создание базы данных и веб-ресурса GlucoGenes® выполнено за счет гранта Российского научного фонда (№ 20-15-00057-П).

Благодарности. Фрагменты по эволюционному анализу генов и анализу SNP реализованы за счет государственного задания ИЦиГ СО РАН (бюджетный проект № FWNR-2022-0006). Авторы выражают искреннюю признательность О.В. Сайк (НИИКЭЛ – филиал ИЦиГ СО РАН) за огромный вклад в сбор данных и ценные советы по формированию базы данных. Мы также благодарим А.М. Мухина (ИЦиГ СО РАН) за техническую помощь в создании веб-ресурса.

GlucoGenes®, a database of genes and proteins associated with glucose metabolism disorders, its description and applications in bioinformatics research

V.V. Klimontov [1] ✉, K.S. Shishin [1], R.A. Ivanov [2], M.P. Ponomarenko [2], K.A. Zolotareva [2], S.A. Lashin [1, 2]

[1] Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ klimontov@mail.ru

© Климонтов В.В., Шишин К.С., Иванов Р.А., Пономаренко М.П., Золотарева К.А., Лашин С.А., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

Abstract. Data on the genetics and molecular biology of diabetes are accumulating rapidly. This poses the challenge of creating research tools for a rapid search for, structuring and analysis of information in this field. We have developed a web resource, GlucoGenes®, which includes a database and an Internet portal of genes and proteins associated with high glucose (hyperglycemia), low glucose (hypoglycemia), and both metabolic disorders. The data were collected using text mining of the publications indexed in PubMed and PubMed Central and analysis of gene networks associated with hyperglycemia, hypoglycemia and glucose variability performed with ANDSystems, a bioinformatics tool. GlucoGenes® is freely available at: https://glucogenes.sysbio.ru/genes/main. GlucoGenes® enables users to access and download information about genes and proteins associated with the risk of hyperglycemia and hypoglycemia, molecular regulators with hyperglycemic and antihyperglycemic activity, genes up-regulated by high glucose and/or low glucose, genes down-regulated by high glucose and/or low glucose, and molecules otherwise associated with the glucose metabolism disorders. With GlucoGenes®, an evolutionary analysis of genes associated with glucose metabolism disorders was performed. The results of the analysis revealed a significant increase (up to 40 %) in the proportion of genes with phylostratigraphic age index (PAI) values corresponding to the time of origin of multicellular organisms. Analysis of sequence conservation using the divergence index (DI) showed that most of the corresponding genes are highly conserved (DI < 0.6) or conservative (DI < 1). When analyzing single nucleotide polymorphism (SNP) in the proximal regions of promoters affecting the affinity of the TATA-binding protein, 181 SNP markers were found in the GlucoGenes® database, which can reduce (45 SNP markers) or increase (136 SNP markers) the expression of 52 genes. We believe that this resource will be a useful tool for further research in the field of molecular biology of diabetes.

Key words: gene; protein; diabetes mellitus; hyperglycemia; hypoglycemia; glucose variability; database; phylostratigraphic index; single nucleotide polymorphism.

For citation: Klimontov V.V., Shishin K.S., Ivanov R.A., Ponomarenko M.P., Zolotareva K.A., Lashin S.A. GlucoGenes®, a database of genes and proteins associated with glucose metabolism disorders, its description and applications in bioinformatics research. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8): 1008-1017. doi 10.18699/vjgb-24-107

Введение

Сахарный диабет (СД) – одно из самых распространенных и социально значимых заболеваний человека. Согласно оценкам экспертов Международной федерации диабета, в 2021 г. число людей, живущих с СД, в мире достигло 537 млн, к 2045 г. ожидается увеличение этого показателя до 783 млн. Кроме того, более 540 млн человек имеют нарушенную толерантность к глюкозе (International Diabetes Federation, 2021).

В последние годы достигнут значительный прогресс в понимании молекулярных механизмов развития СД и его осложнений. Полногеномные исследования выявили ряд новых генетических локусов, модулирующих риск развития СД и диабетических осложнений в европейских и азиатских популяциях. Исследования в области протеомики, метаболомики и мультиомики пролили свет на молекулярные основы патогенеза заболевания (Langenberg, Lotta, 2018; Lyssenko, Vaag, 2023; Shojima et al., 2023).

В то же время обнаружено влияние воздействия высокой концентрации глюкозы на регуляцию экспрессии генов в различных тканях (Vaulont et al., 2000; Hall et al., 2018; Vega et al., 2020; Zhang S. et al., 2021). Установлено, что влияние высокого уровня глюкозы на экспрессию генов может быть пролонгировано и усугублено эпигенетическими модификациями. Этот механизм считается важным для феномена метаболической памяти и развития диабетических осложнений (Dhawan et al., 2022). Аномально низкий уровень глюкозы также сопровождается рядом биохимических сдвигов. Эти сдвиги связаны прежде всего с реакцией сердечно-сосудистой и нервной систем на гипогликемию (Hanefeld et al., 2016; Rehni, Dave, 2018).

Все большее внимание привлекают молекулярные эффекты повторяющихся эпизодов высокого и низкого уровня глюкозы, которые характерны для феномена высокой вариабельности гликемии (ВГ). Установлено, что повышенная ВГ увеличивает риск развития микро- и макрососудистых осложнений СД и ассоциирована с повышенной смертностью больных от всех причин и от сердечно-сосудистых заболеваний (Ceriello et al., 2019; Wilmot et al., 2019). На молекулярном уровне патофизиологические изменения, ассоциированные с высокой ВГ, реализуются через повышение или понижение экспрессии большого количества генов и изменение активности сигнальных путей, таких как PI3K/Akt, NF-κB, MAPK (ERK), JNK и TGF-β/Smad (Klimontov et al., 2021).

С учетом огромного количества и разнообразия молекулярных изменений при СД для формирования всестороннего и целостного представления о патогенезе заболевания необходимы передовые аналитические инструменты. Искусственный интеллект, биоинформатика и интегративная системная биология предоставляют новые возможности для изучения сложных заболеваний, таких как СД (Nielsen, 2017; Климонтов и др., 2021; Orlov et al., 2021; Putra et al., 2024).

Перспективным подходом в этой области является анализ генных сетей, т. е. групп координированно функционирующих генов, взаимодействующих друг с другом и определяющих тот или иной фенотипический признак организма (Колчанов и др., 2013). Ранее, используя текстмайнинг научных публикаций, индексируемых в базах данных Pubmed и Pubmed Central (PMC), а также биоинформационный анализ, мы реконструировали генные сети, ассоциированные с нарушениями метаболизма глюкозы (НМГ): гипергликемией, гипогликемией и ВГ, и собрали большое количество материалов о молекулах и белках, так или иначе связанных с этими метаболическими нарушениями (Saik, Klimontov, 2020–2022).

На основе полученных результатов была создана база данных, содержащая информацию о генах и белках, свя-

занных с НМГ: гипергликемией, гипогликемией и обоими нарушениями. В этой статье мы представляем описание возможностей базы данных, а также приводим результаты двух биоинформатических исследований, проведенных с ее использованием. В первом исследовании был выполнен эволюционный анализ генов НМГ, во втором – анализ однонуклеотидного полиморфизма (SNP) проксимальных районов длиной 90 п. о. промоторов генов человека, связанных с нарушениями метаболизма глюкозы.

Материалы и методы

Формирование веб-ресурса. Материал для ресурса GlucoGenes® был получен ранее в ходе реконструкции и анализа генных сетей гипергликемии, гипогликемии и ВГ, проведенных с использованием биоинформатической системы ANDSystem (разработана в ИЦиГ СО РАН; доступ: https://anddigest.sysbio.ru/index.php). ANDSystem строит ассоциативные генные сети на основе автоматического анализа текстов (текст-майнинга) научных публикаций, проиндексированных в базах данных PubMed и PMC (Ivanisenko et al., 2015, 2019). Детали анализа полученных с помощью ANDSystem генных сетей НМГ описаны ранее (Saik, Klimontov, 2020–2022).

Для программной реализации базы данных были выбраны реляционная модель данных и система управления базой данных PostgreSQL (https://www.postgresql.org/).

При проектировании веб-ресурса была выбрана клиент-серверная архитектура, состоящая из трех основных компонентов: клиент, сервер и база данных. Для разработки использованы фреймворки Vue.js и Flask; управление доступом к базе данных осуществляется через программный доступ на основе технологии REST.

Филостратиграфический анализ и анализ дивергенции генов, ассоциированных с НМГ. Филостратиграфический анализ – это метод, направленный на определение эволюционного происхождения генов путем анализа присутствия их ортологов – генов, кодирующих гомологичные белки, разошедшиеся в результате видообразования в геномах различных видов. Такой подход выявляет ключевые моменты в эволюции генома, когда наблюдался резкий рост числа новых генов, а также помогает идентифицировать гены, уникальные для отдельных таксонов (Domazet-Lošo, Tautz, 2010).

Нами проведен эволюционный анализ генов, включенных в базу данных GlucoGenes®, с применением филостратиграфического возрастного индекса (phylostratigraphy age index, PAI) и индекса дивергенции (divergency index, DI). Значение PAI показывает, насколько таксон, отражающий возраст гена, удален от корня филогенетического дерева (Мустафин и др., 2021). Под таксоном, отражающим возраст гена, понимается таксон, в котором произошло расхождение изучаемого вида с наиболее удаленным родственным таксоном, в котором был найден ортолог рассматриваемого гена. Чем больше значение PAI исследуемого гена, тем моложе этот ген. Для филостратиграфического анализа использовали программный комплекс Orthoweb (https://orthoweb.sysbio.cytogen.ru/run.html). При расчете PAI применяли метод расчета по группам ортологии KEGG (Kanehisa et al., 2016).

Индекс DI – это показатель эволюционной изменчивости гена. Он рассчитывается на основе соотношения *dN/dS*, где *dN* – доля несинонимичных замен в последовательностях ДНК исследуемого гена и его ортолога; *dS* – доля синонимичных замен. Значение этого показателя было рассчитано на основе сравнения генов человека с генами близкородственных организмов из семейства Hominidae, а именно ортологов, найденных у западной равнинной гориллы *Gorilla gorilla gorilla*, суматранского орангутанга *Pongo abelii* и шимпанзе обыкновенного *Pan troglodytes*. Для расчета значений *dN/dS* применяли модель LPB93 (Yang, 2007). Значение DI в диапазоне от 0 до 1 показывает, что ген подвергается стабилизирующему отбору, 1 – нейтральной эволюции, а более 1 – движущему отбору.

Анализ SNP в проксимальных районах длиной 90 п. о. промоторов генов человека, связанных с НМГ. В базе знаний Human_SNP_TATAdb (Филонов и др., 2023) осуществили поиск вариантов SNP в проксимальных районах длиной 90 п. о. промоторов генов человека, связанных с НМГ, которые способны статистически значимо уменьшать или увеличивать сродство TATA-связывающего белка (TBP) к этим промоторам и, соответственно, уровни экспрессии генов. Для дальнейшей работы оставили SNP, для которых описаны клинические проявления в свободно доступной базе данных ClinVar (Landrum et al., 2014). Наконец, с использованием базы данных PubMed для всех отобранных таким путем клинических SNP-маркеров заболеваний в промоторах генов человека, ассоциированных с нарушениями метаболизма глюкозы, осуществили курируемую аннотацию влияния полиморфного изменения экспрессии этих генов на уровни глюкозы у пациентов-носителей соответствующих SNP.

Результаты

Ресурс GlucoGenes®

Ресурс GlucoGenes® находится в свободном доступе по адресу: https://glucogenes.sysbio.ru/genes. Интерфейс ресурса показан на рис. 1.

База данных GlucoGenes® включает шесть компонентов (таблиц). Таблица Genes содержит названия генов, их описание и идентификаторы NCBI. Таблица Proteins включает названия белков, их описание и идентификаторы в базе данных UniProt, а также ссылки на соответствующие гены в базе данных GlucoGenes®. Таблица Glycemia_related_conditions представляет информацию о НМГ (гипергликемия, гипогликемия). Таблица Types_of_glycemia_gene_association содержит информацию о типах ассоциаций между молекулами и НМГ. Таблица References включает идентификаторы статей Pubmed или PubMed Central с кратким извлечением данных. Таблица Glycemia_gene_association является сводной, аккумулирует информацию из всех вышеперечисленных таблиц. Структура базы данных представлена на рис. 2. В настоящее время база данных включает 561 ген, связанный с НМГ, и 2115 ссылок на литературные источники.

Веб-ресурс GlucoGenes® включает четыре функциональных раздела.

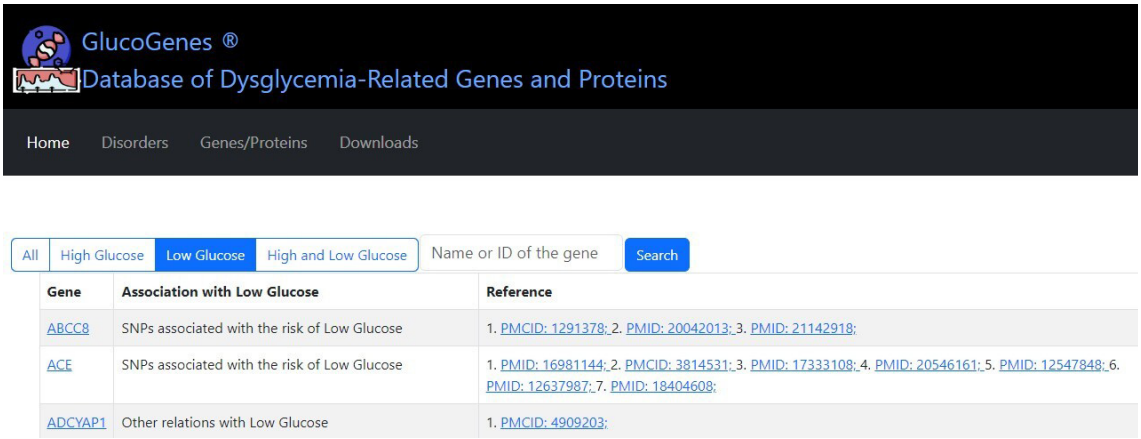


Рис. 1. Фрагмент интерфейса интернет-портала GlucoGenes®.



Рис. 2. Диаграмма «сущность-связь» базы данных GlucoGenes®.

1. На главной странице (Homepage) представлена общая информация о ресурсе и используемых терминах. Мы обозначили гипергликемию, или высокий уровень глюкозы в культуральной среде, как High Glucose (HG), а гипогликемию, или низкий уровень глюкозы в культуральной среде, – как Low Glucose (LG).
2. Страница НМГ (Disorders). Здесь представлены списки генов, ассоциированных с HG, LG, а также с высоким и низким уровнем глюкозы (HLG). Для каждого гена указан тип ассоциации с НМГ. Среди ассоциаций с НМГ мы выделили следующие категории: однонуклеотидные полиморфизмы (SNP), ассоциированные с риском HG, LG или HLG; белки с гипергликемической активностью; белки с антигипергликемическим эффектом; гены, экспрессия которых повышается в условиях HG; гены, экспрессия которых повышается в условиях LG; гены, экспрессия которых понижается в условиях HG; гены, экспрессия которых понижается в условиях LG; другие ассоциации с HG, LG или HLG. Для каждого гена и ассоциации приведены ссылки на соответствующие публикации в PubMed.
3. Каталог генов/белков, связанных с НМГ (Genes/Proteins). Здесь можно найти название гена и NCBI-идентификатор гена, название(я) белка(ов), кодируемого(ых) геном, а также типы ассоциации с НМГ.
4. Страница загрузки (Downloads). С этой страницы можно скачать списки генов, ассоциированных с HG, LG и HLG, а также все ассоциированные гены в формате Excel. Поиск в системе может осуществляться по имени гена, идентификатору гена NCBI или типу НМГ.

Доступ к данным портала возможен и без использования графического интерфейса пользователя, через программный интерфейс (API) REST. Данный интерфейс позволяет получить необходимую информацию при отправке запроса на веб-сервер в виде строки URL. В ответ на такой запрос сервер возвращает результат в виде тек-

Таблица 1. Списки белок-кодирующих генов человека, для которых был проведен филостратиграфический анализ

Группа генов	Описание	Число генов
Все гены в геноме *Homo sapiens*	Все белок-кодирующие гены человека, для которых были посчитаны значения PAI и DI	19 566
Гены, ассоциированные только с гипергликемией	Список генов из GlucoGenes, ассоциированных с высоким уровнем глюкозы	430
Гены, ассоциированные только с гипогликемией	Список генов из GlucoGenes, ассоциированных с низким уровнем глюкозы	140
Гены, ассоциированные с высоким и низким уровнем глюкозы	Список генов из GlucoGenes, ассоциированных с высоким и низким уровнем глюкозы	151

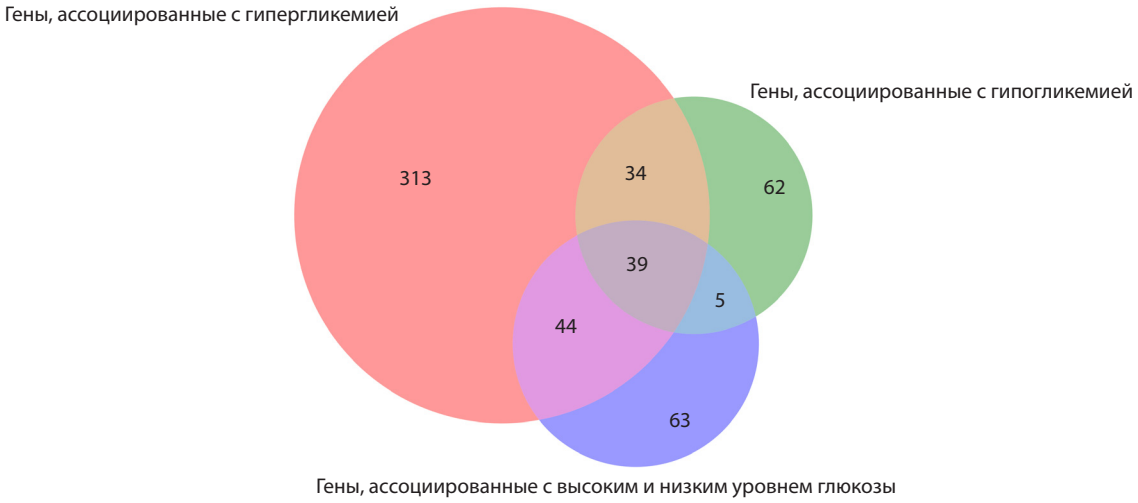


Рис. 3. Диаграмма Венна пересечений групп генов выборок.

стовой страницы или файла, информация в котором структурирована в соответствии с форматом JavaScript Object Notation (JSON) (http://json.org/). Полученный текстовый файл можно открыть любым текстовым редактором. Файл также может быть обработан с помощью различных программных средств, в том числе программ, написанных пользователем в универсальных средах моделирования (например, Matlab, Scilab и др.) или на языках программирования высокого уровня (например, Python, R, C++, Java и др.). Пример REST-запроса приведен ниже (результатом является текстовый файл в структурированном формате JSON):

https://glucogenes.sysbio.ru/api/genes/<geneid> – возвращает карточку с описанием гена <geneid>.

Эволюционные характеристики генов, связанных с НМГ. Мы рассчитали индексы PAI и построили их распределение как для списка белок-кодирующих генов в геноме *Homo sapiens*, так и для генов, представленных в базе данных GlucoGenes®: для генов, ассоциированных с гипергликемией, гипогликемией и колебаниями уровня глюкозы (табл. 1). Примечательно, что часть генов имела ассоциацию более чем с одной патологией (рис. 3).

Распределение значений PAI для всех генов в геноме человека является мультимодальным и имеет два ярко выраженных пика на уровне Cellular Organisms: Metazoa и Vertebrata-Euteleostomi (рис. 4). Первый пик – самый большой; почти 55 % генов в геноме *H. sapiens* имеют PAI от 1 до 3. Второй пик охватывает 32 % генов.

Результаты анализа выявили значительное увеличение доли генов (до 40 %), связанных с регуляцией уровня глюкозы, имеющих индекс PAI = 3, во всех трех категориях (рис. 4). В частности, к этой группе генов относятся *TCF7L2*, *PPARG*, *GCGR*, *IRS1* и *MTNR1B*, продукты которых – важные регуляторы метаболизма глюкозы.

Анализ консервативности последовательностей для тех же списков генов (рис. 5) показал, что большинство из изученных генов высококонсервативны (DI < 0.6) или консервативны (DI < 1). Это свидетельствует о сохранении их функций в процессе эволюции и подчеркивает их критическую роль в биологических процессах, связанных с регуляцией уровня глюкозы. Однако был обнаружен ряд генов с DI выше 1, что указывает на недавнее воздействие на них движущего отбора. К таким генам относятся *SPP1*, *CALCA*, *CD33*, *SULT2A1*, *TNF*, *ECM1*, *CYP3A4* и *EDN1*.

Анализ SNP в проксимальных районах длиной 90 п. о. промоторов генов человека, связанных с НМГ. Всего в базе данных GlucoGenes® нами найден 181 SNP-маркер, который может снижать (45 SNP-маркеров) или повышать (136 SNP-маркеров) экспрессию 52 генов человека, изменяя таким путем уровни глюкозы у пациентов с минорными аллелями этих SNP. В табл. 2 в качестве при-

PAI	Таксон
1	Cellular Organisms
2	Eukaryota
3	Metazoa
4	Chordata
5	Craniata
6	Vertebrata
7	Euteleostomi
8	Mammalia
9	Eutheria
10	Euarchontoglires
11	Primates
12	Haplorrhini
13	Catarrhini
14	Hominidae
15	Homo
16	*Homo sapiens*

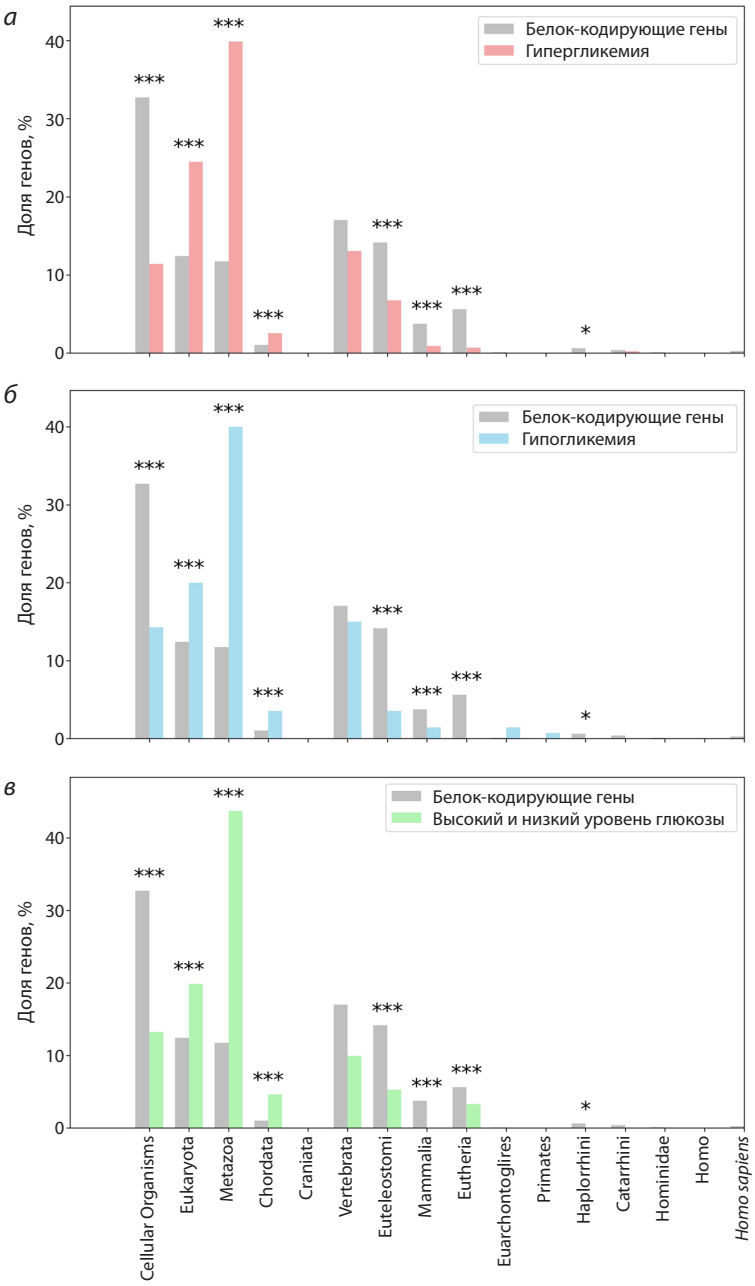


Рис. 4. Распределение белок-кодирующих генов, ассоциированных с НМГ, по значениям PAI.

Здесь и на рис. 5: *а* – все белок-кодирующие гены человека (All_CDS) в качестве контрольного списка и гены, ассоциированные с высоким уровнем глюкозы (Hyperglycemia); *б* – все белок-кодирующие гены человека (All_CDS) в качестве контрольного списка и гены, ассоциированные с низким уровнем глюкозы (Hypoglycemia); *в* – все белок-кодирующие гены человека (All_CDS) в качестве контрольного списка и гены, ассоциированные с высоким и низким уровнем глюкозы (High, and Low Glucose-related genes). Звездочками отмечены столбцы со статистически достоверно различающимися значениями между выборками генов из базы данных и выборкой всех белок-кодирующих генов человека: * *p*-value < 0.05, *** *p*-value < 0.001. Статистическая проверка проводилась при помощи критерия chi-square.

мера представлено 10 SNP в промоторах генов *ABCC8*, *INSR* и *PGM1* человека, доступных в базе данных ClinVar (Landrum et al., 2014), для которых в базе данных GlucoGenes® предлагается механизм ассоциаций с нарушениями метаболизма глюкозы.

Согласно данным, представленным в табл. 2, минорные аллели промоторов генов *ABCC8*, *INSR* и *PGM1* характеризуются измененным сродством к TBP, что может влиять на уровень экспрессии этих генов и объяснять их связь с НМГ. Более полная информация о найденных SNP-маркерах приведена в Приложении[1].

Обсуждение

Достижения в изучении молекулярной биологии СД предоставляют широкие возможности для внедрения технологий высокоточной медицины в лечение заболевания.

[1] Приложение см. по адресу:
https://vavilovj-icg.ru/download/pict-2024-28/appx36.pdf

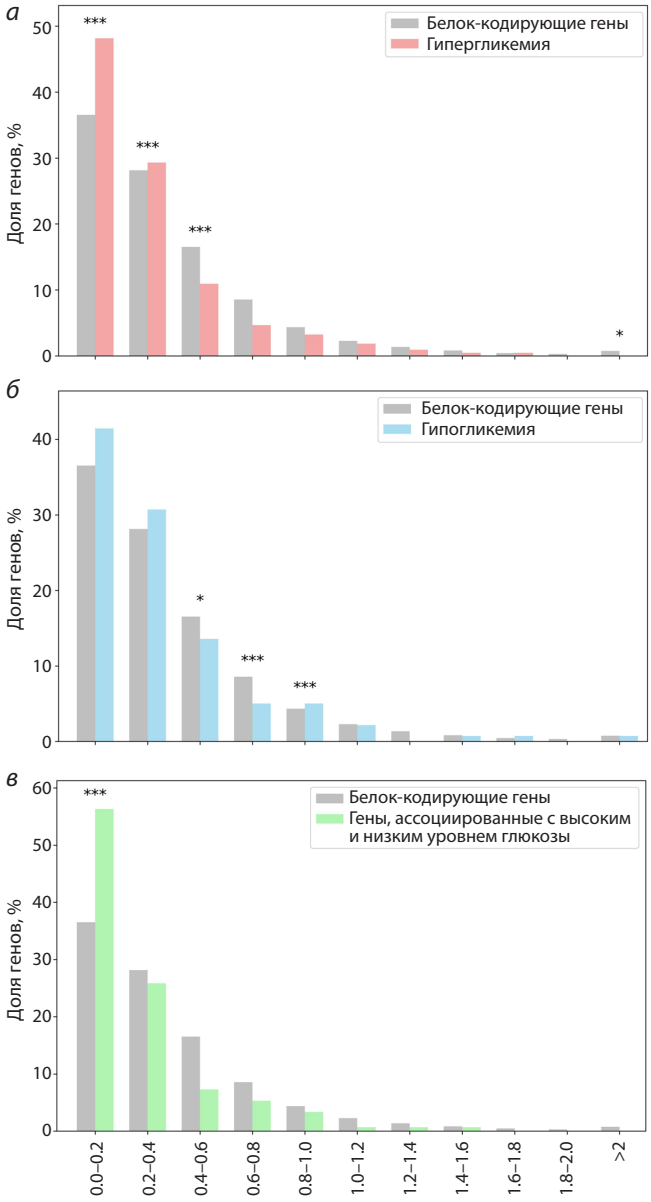


Рис. 5. Распределение белок-кодирующих генов, ассоциированных с НМГ, по значениям DI.

В частности, выявление специфических для заболевания биомаркеров дает новые перспективы в диагностике, мониторинге, прогнозировании заболевания и его исходов, в фармакогенетике современных сахароснижающих препаратов, а также в поиске новых терапевтических средств (Chung et al., 2020). Быстрое накопление данных о молекулярных основах генетической предрасположенности к СД и молекулярных механизмах его осложнений определяет потребность в исследовательских инструментах для структурированного поиска информации в этой области.

Мы создали базу данных генов и белков, для которых были продемонстрированы ассоциации с НМГ: гипергликемией, гипогликемией и обоими нарушениями. Разработанный веб-ресурс под названием GlucoGenes® (https://glucogenes.sysbio.ru/genes/main) может быть использован для сбора, поиска и визуализации информации о генах и белках, так или иначе связанных с НМГ. Доступ к базе данных, интегрированной в GlucoGenes®, осуществляется через программный интерфейс на основе REST для просмотра записей. Графический пользовательский интерфейс позволяет просматривать записи и экспортировать их содержимое в формат Excel. База данных содержит каталоги генов и белков, ассоциированных с НМГ, с информацией о типах ассоциаций и ссылками на рефераты соответствующих публикаций в PubMed или полнотекстовые статьи в PMC. Списки генов и белков доступны для скачивания. Ограничением ресурса является аккумуляция данных только из источников, проиндексированных в Pubmed и PMC. Очевидна необходимость регулярного обновления информации.

Созданный ресурс может быть полезен для решения исследовательских задач в области биоинформатики и молекулярной биологии СД. В частности, он может применяться для отбора генов и белков с целью исследования генетической предрасположенности к СД в различных популяциях, в изучении молекулярных аспектов патогенеза СД, в поиске потенциальных биомаркеров осложнений заболевания, молекул-таргетов для терапии и иных задач. В этой статье мы приводим примеры использования разработанного ресурса для решения научных задач в биоинформатике.

Первое исследование посвящено эволюционному происхождению генов, ассоциированных с НМГ. Эволюционный анализ генов при помощи филостратиграфии является ключевым инструментом в биологии, позволяющим понять фундаментальные механизмы, лежащие в основе разнообразия жизни на Земле. Эволюционная история генов раскрывает информацию о том, как различные функции и структуры развивались и адаптировались к изменениям окружающей среды. Это знание не только помогает в реконструкции филогенетических деревьев, но и способствует идентификации генов, ответственных за адаптивные изменения и специфические физиологические процессы, такие как метаболизм глюкозы. Проведенный филостратиграфический анализ показал, что среди генов, ассоциированных с метаболизмом глюкозы, значительную долю (до 40 %) составляют гены с PAI = 3, что соответствует времени происхождения многоклеточных организмов (Maloof et al., 2010). Большинство изученных генов оказались высококонсервативными (DI < 0.6) или консервативными (DI < 1). Полученные результаты подчеркивают важность генов, ассоциированных с НМГ, в регуляции специализированных метаболических процессов, характерных для сложных организмов.

Во втором исследовании данные веб-ресурса были использованы для анализа SNP в проксимальных районах промоторов генов человека, влияющих на сродство TATA-связывающего белка. Интеграция данных GlucoGenes® с информацией из других баз данных об ассоциациях SNP с рядом заболеваний человека, с одной стороны, и биоинформатических оценок изменения уровня глюкозы у пациентов с этими SNP, с другой стороны, отражает молекулярные механизмы возможного влияния НМГ на течение этих заболеваний.

Таблица 2. SNP, влияющие на сродство TBP, в промоторах генов *ABCC8*, *INSR* и *PGM1*, ассоциированных с НМГ

Интернет-портал GlucoGenes®		База данных ClinVar (Landrum et al., 2014)					База данных Human_SNP_TATAdb (Филонов и др., 2023)						База данных PubMed
		Ассоциации с SNP		База данных dbSNP (Day, 2010)			K_D, нМ/л, *in silico*		Значимость				Влияние на течение болезни (Ref)
Символ гена	Статус гликемии	Заболевание	dbSNP ID:min	5′-фланг, 10 bp	WT → min	3′-фланг, 10 bp	WT	min	z	*p*	ρ	Δ	
							$M_0 \pm SEM$	$M_0 \pm SEM$					
PGM1	Low Glucose	PGM1-congenital disorder of glycosylation, inborn genetic diseases	rs199928376:T	tgccgccaac	G→T	gggtaaggga	46.52 ± 3.87	36.68 ± 3.28	3.89	10^{-3}	B	↑	Избыток PGM1 – ответ на дефицит глюкозы при раке легких (Li et al., 2020)
INSR	High and Low Glucose & High Glucose	Insulin-resistant diabetes mellitus & acanthosis nigricans; Rabson-Mendenhall syndrome, leprechaunism syndrome	rs532052290:T	cgcctcggag	C→T	tcttgggtgg	62.14 ± 4.83	17.87 ± 1.92	18.81	10^{-6}	A	↑	Избыток *Insr* улучшал здоровье мышей db/db (Sasaki et al., 2015)
ABCC8	High and Low Glucose & Low Glucose & High Glucose	ABCC8-related disorder, hypoglycemia, hyperinsulinism, diabetes mellitus, monogenic diabetes	rs1409900082:A	agcactttcc	G→A	tggtcctcag	74.04 ± 5.71	30.75 ± 2.85	14.58	10^{-6}	A	↑	Влияние диеты матери на потомство у мышей (Zhang Q. et al., 2024): избыток Abcc8 защищает потомство
			rs1409900082:T	agcactttcc	G→T	tggtcctcag	74.04 ± 5.71	52.37 ± 3.26	6.99	10^{-6}	A		
			rs1955706366:A	cctgtccagt	G→A	ctaagcgagt	53.2 ± 4.54	12.08 ± 1.4	20.58	10^{-6}	A		
			rs1957150758:A	cttggccgac	C→A	actttccgca	74.04 ± 5.71	47.43 ± 3.75	8.06	10^{-6}	A		
			rs541244107:A	agctgcaagg	G→A	ccgggcggcg	59.19 ± 4.91	50.3 ± 3.4	3.04	10^{-2}	C		
			rs541244107:C	agctgcaagg	G→C	ccgggcggcg	59.19 ± 4.91	47.44 ± 3.82	3.83	10^{-3}	B		
			rs565662182:A	tgccccccat	G→A	gaggagcagt	53.2 ± 4.54	31.21 ± 2.78	8.65	10^{-6}	A		
			rs58241708:T	tgcagagatc	C→T	ttcctgtcca	53.2 ± 4.54	45.33 ± 3.8	2.68	10^{-2}	C		

Примечание. WT и min – анцестральный (норма) и минорный (патология) аллели SNP; K_D – равновесная константа диссоциации комплекса TBP-промотор, выраженная в наномолях на литр (нМ/л); M_0 и SEM – контекстно-зависимая оценка *in silico* и ее стандартизированная ошибка; z, *p* и ρ – значение z-статистики Фишера и уровень ее значимости, а также эвристическая приоритизация оценок *in silico* от наилучшей (A) до наихудшей (D) в алфавитном порядке; Δ – экспрессия выше (↑) или ниже (↓).

Заключение

GlucoGenes® – это ресурс, сочетающий в себе графический интерфейс пользователя и базу данных генов и белков, связанных с гипергликемией, гипогликемией и обоими метаболическими нарушениями. Ресурс использован для биоинформатического анализа эволюционных характеристик генов, ассоциированных с указанными нарушениями, а также при анализе SNP в проксимальных районах промоторов генов, влияющих на сродство TATA-связывающего белка. Показано, что значительная часть генов, ассоциированных с НМГ, являются эволюционно древними и консервативными. Идентифицированы SNP-маркеры, которые могут снижать (45 SNP-маркеров) или повышать (136 SNP-маркеров) экспрессию 52 генов.

Список литературы / References

Климонтов В.В., Бериков В.Б., Сайк О.В. Искусственный интеллект в диабетологии. *Сахар. диабет.* 2021;24(2):156-166. doi 10.14341/DM12665
[Klimontov V.V., Berikov V.B., Saik O.V. Artificial intelligence in diabetology. *Sakharnyi Diabet = Diabetes Mellitus.* 2021;24(2): 156-166. doi 10.14341/DM12665 (in Russian)]

Колчанов Н.А., Игнатьева Е.В., Подколодная О.А., Лихошвай В.А., Матушкин Ю.Г. Генные сети. *Вавиловский журнал генетики и селекции* 2013;17(4/2):833-850
[Kolchanov N.A., Ignatieva E.V., Podkolodnaya O.A., Likhoshvai V.A., Matushkin Yu.G. Gene networks. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2013;17(4/2):833-850 (in Russian)]

Мустафин З.С., Лашин С.А., Матушкин Ю.Г. Филостратиграфический анализ генных сетей заболеваний человека. *Вавиловский журнал генетики и селекции*. 2021;25(1):46-56. doi 10.18699/VJ21.006
[Mustafin Z.S., Lashin S.A., Matushkin Yu.G. Phylostratigraphic analysis of gene networks of human diseases. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2021;25(1):46-56. doi 10.18699/VJ21.006]

Филонов С.В., Подколодный Н.Л., Подколодная О.А., Твердохлеб Н.Н., Пономаренко П.М., Рассказов Д.А., Богомолов А.Г., Пономаренко М.П. Human_SNP_TATAdb – база данных о SNP, статистически достоверно изменяющих сродство TATA-связывающего белка к промоторам генов человека: полногеномный анализ и варианты использования. *Вавиловский журнал генетики и селекции*. 2023;27(7):728-736. DOI 10.18699/VJGB-23-85
[Filonov S.V., Podkolodnyy N.L., Podkolodnaya O.A., Tverdokhleb N.N., Ponomarenko P.M., Rasskazov D.A., Bogomolov A.G., Ponomarenko M.P. Human_SNP_TATAdb: a database of SNPs that statistically significantly change the affinity of the TATA-binding protein to human gene promoters: genome-wide analysis and use cases. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):728-736. doi 10.18699/VJGB-23-85]

Ceriello A., Monnier L., Owens D. Glycaemic variability in diabetes: clinical and therapeutic implications. *Lancet Diabetes Endocrinol.* 2019;7(3):221-230. doi 10.1016/S2213-8587(18)30136-0

Chung W.K., Erion K., Florez J.C., Hattersley A.T., Hivert M.F., Lee C.G., McCarthy M.I., Nolan J.J., Norris J.M., Pearson E.R., Philipson L., McElvaine A.T., Cefalu W.T., Rich S.S., Franks P.W. Precision medicine in diabetes: a Consensus Report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia.* 2020;63(9):1671-1693. doi 10.1007/s00125-020-05181-w

Day I.N. dbSNP in the detail and copy number complexities. *Hum. Mutat.* 2010;31(1):2-4. doi 10.1002/humu.21149

Dhawan P., Vasishta S., Balakrishnan A., Joshi M.B. Mechanistic insights into glucose induced vascular epigenetic reprogramming in type 2 diabetes. *Life Sci.* 2022;298:120490. doi 10.1016/j.lfs.2022.120490

Domazet-Lošo T., Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature.* 2010;468(7325):815-819. doi 10.1038/nature09632

Hall E., Dekker Nitert M., Volkov P., Malmgren S., Mulder H., Bacos K., Ling C. The effects of high glucose exposure on global gene expression and DNA methylation in human pancreatic islets. *Mol. Cell. Endocrinol.* 2018;472:57-67. doi 10.1016/j.mce.2017.11.019

Hanefeld M., Frier B.M., Pistrosch F. Hypoglycemia and cardiovascular risk: is there a major link? *Diabetes Care.* 2016;39(S.2):S205-S209. doi 10.2337/dcS15-3014

International Diabetes Federation. IDF Diabetes Atlas, 10th ed. Brussels, 2021

Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol.* 2015;9(S2):S2. doi 10.1186/1752-0509-9-S2-S2

Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics.* 2019;20(1):34. doi 10.1186/s12859-018-2567-6

Kanehisa M., Sato Y., Kawashima M., Furumichi M., Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44(D1):D457-D462. doi 10.1093/nar/gkv1070

Klimontov V.V., Saik O.V., Korbut A.I. Glucose variability: How does it work? *Int. J. Mol. Sci.* 2021;22(15):7783. doi 10.3390/ijms22157783

Landrum M.J., Lee J.M., Riley G.R., Jang W., Rubinstein W.S., Church D.M., Maglott D.R. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014;42:D980-D985. doi 10.1093/nar/gkt1113

Langenberg C., Lotta L.A. Genomic insights into the causes of type 2 diabetes. *Lancet.* 2018;391(10138):2463-2474. doi 10.1016/S0140-6736(18)31132-2

Li Y., Liang R., Sun M., Li Z., Sheng H., Wang J., Xu P., Liu S., Yang W., Lu B., Zhang S., Shan C. AMPK-dependent phosphorylation of HDAC8 triggers PGM1 expression to promote lung cancer cell survival under glucose starvation. *Cancer Lett*. 2020;478:82-92. doi 10.1016/j.canlet.2020.03.007

Lyssenko V., Vaag A. Genetics of diabetes-associated microvascular complications. *Diabetologia.* 2023;66(9):1601-1613. doi 10.1007/s00125-023-05964-x

Maloof A.C., Porter S.M., Moore J.L., Dudás F.Ö., Bowring S.A., Higgins J.A., Fike D.A., Eddy M.P. The earliest Cambrian record of animals and ocean geochemical change. *Geol. Soc. Am. Bull.* 2010;122(11-12):1731-1774. doi 10.1130/B30346.1

Nielsen J. Systems biology of metabolism: A driver for developing personalized and precision medicine. *Cell Metab*. 2017:25(3):572-579. doi 10.1016/j.cmet.2017.02.002

Orlov Y.L., Anashkina A.A., Klimontov V.V., Baranova A.V. Medical genetics, genomics and bioinformatics aid in understanding molecular mechanisms of human diseases. *Int. J. Mol. Sci.* 2021;22(18):9962. doi 10.3390/ijms22189962

Putra S.E.D., Martriano Humardani F., Antonius Y., Jonathan J., Thalia Mulyanata L. Epigenetics of Diabetes: A bioinformatic approach. *Clin. Chim. Acta.* 2024;557:117856. doi 10.1016/j.cca.2024.117856

Rehni A.K., Dave K.R. Impact of hypoglycemia on brain metabolism during diabetes. *Mol. Neurobiol.* 2018;55(12):9075-9088. doi 10.1007/s12035-018-1044-6

Saik O.V., Klimontov V.V. Bioinformatic reconstruction and analysis of gene networks related to glucose variability in diabetes and its complications. *Int. J. Mol. Sci.* 2020;21(22):8691. doi 10.3390/ijms21228691

Saik O.V., Klimontov V.V. Hypoglycemia, vascular disease and cognitive dysfunction in diabetes: insights from text mining-based reconstruction and bioinformatics analysis of the gene networks. *Int. J. Mol. Sci.* 2021;22(22):12419. doi 10.3390/ijms22222212419

Saik O.V., Klimontov V.V. Gene networks of hyperglycemia, diabetic complications, and human proteins targeted by SARS-CoV-2: what is the molecular basis for comorbidity? *Int. J. Mol. Sci.* 2022;23:7247. doi 10.3390/ijms23137247

Sasaki T., Kuroko M., Sekine S., Matsui S., Kikuchi O., Susanti V.Y., Kobayashi M., Tanaka Y., Yuasa T., Kitamura T. Overexpression of insulin receptor partially improves obese and diabetic phenotypes in db/db mice. *Endocr. J.* 2015;62(9):787-796. doi 10.1507/endocrj.ej15-0255

Shojima N., Yamauchi T. Progress in genetics of type 2 diabetes and diabetic complications. *J. Diabetes Investig.* 2023;14(4):503-515. doi 10.1111/jdi.13970

Vaulont S., Vasseur-Cognet M., Kahn A. Glucose regulation of gene transcription. *J. Biol. Chem.* 2000;275(41):31555-31558. doi 10.1074/jbc.R000016200

Vega M.E., Finlay J., Vasishtha M., Schwarzbauer J.E. Elevated glucose alters global gene expression and tenascin-C alternative splicing in mesangial cells. *Matrix Biol. Plus.* 2020;8:100048. doi 10.1016/j.mbplus.2020.100048

Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 2007;24(8):1586-1591. doi 10.1093/molbev/msm088

Wilmot E.G., Choudhary P., Leelarathna L., Baxter M. Glycaemic variability: The under-recognized therapeutic target in type 1 diabetes care. *Diabetes Obes. Metab.* 2019;21(12):2599-2608. doi 10.1111/dom.13842

Zhang S., Ke Z., Yang C., Zhou P., Jiang H., Chen L., Li Y., Li Q. High glucose causes distinct expression patterns of primary human skin cells by RNA sequencing. *Front. Endocrinol.* 2021;12:603645. doi 10.3389/fendo.2021.603645

Zhang Q., Xiao X., Zheng J., Li M., Yu M., Ping F., Wang T., Wang X. DNA methylation regulates pancreatic gene expression and links maternal high-fat diet to the offspring glucose metabolism. *J. Nutr. Biochem*. 2024;123:109490. doi 10.1016/j.jnutbio.2023.109490

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 02.10.2024. После доработки 08.11.2024. Принята к публикации 12.11.2024.

Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(8):1018-1024

doi 10.18699/vjgb-24-108

Ассоциация аутистических личностных черт у неклинических испытуемых с показателями ЭЭГ в условиях просмотра видеозаписей лица

А.Н. Савостьянов [1, 2, 3], Д.А. Кулешов [1, 4, 5], Д.И. Клемешова[1, 2], М.С. Власов [6], А.Е. Сапрыгин [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Научно-исследовательский институт нейронаук и медицины, Новосибирск, Россия
[3] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[4] Институт нефтегазовой геологии и геофизики им. А.А. Трофимука Сибирского отделения Российской академии наук, Новосибирск, Россия
[5] Сибирский государственный университет телекоммуникаций и информатики, Новосибирск, Россия
[6] Бийский филиал им. В.М. Шукшина Алтайского государственного педагогического университета, Бийск, Россия
a-sav@mail.ru

Аннотация. В рамках проводимого исследования разработан и апробирован программно-информационный модуль экспериментально-компьютерной платформы "EEG_Self-Construct", позволяющий выявлять нейрофизиологические маркеры самореферентных процессов на основе совместного использования ЭЭГ и регистрации видеозаписей лица для индукции функциональных состояний головного мозга, ассоциированных с личностными особенностями участников. Этот модуль был апробирован на группе неклинических участников с разной степенью выраженности аутистических личностных черт (АЛЧ), измеренных с помощью опросника расширенного фенотипа аутизма. Степень индивидуальной выраженности АЛЧ – это количественный показатель, который характеризует затруднения, возникающие у человека при коммуникации с другими людьми. У каждого человека имеется некоторая индивидуальная степень выраженности таких черт. Высокие значения аутистических черт определяются у пациентов с аутизмом. Однако существуют также люди, у которых высокие значения АЛЧ не сопровождаются клинической симптоматикой. Разработанный нами модуль дает возможность индуцировать функциональные состояния головного мозга, в которых ЭЭГ-показатели людей с разным уровнем АЛЧ достоверно различаются. Кроме того, модуль включает комплект программного обеспечения для регистрации и анализа индексов мозговой активности. Нами установлено, что зависимости между мозговой активностью и индивидуальным уровнем выраженности АЛЧ у неклинических испытуемых могут быть выявлены в условиях функционального покоя, следующих за распознаванием самоотнесенной информации, тогда как распознавание социально нейтральной информации не индуцирует процессы, связанные с аутистичностью. Показано, что у людей с высокими значениями АЛЧ наблюдаются повышенные показатели спектральной плотности в диапазонах дельта- и тета-ритмов в лобных отделах обоих полушарий в сравнении с людьми с низкой степенью аутистичности. Это может быть гипотетически интерпретировано как индекс сниженной мозговой активности, ассоциированной с распознаванием самоотнесенной информации у людей с высокой аутистичностью. Разрабатываемый нами программный модуль может быть интегрирован с модулями, позволяющими выявлять молекулярно-генетические маркеры личностных черт, включая черты, определяющие предрасположенность к психиатрическим патологиям.

Ключевые слова: информационно-цифровые платформы в медицине; нейровычислительные технологии; ЭЭГ покоя; аутистические черты; расширенный аутистический фенотип; самореференция; дефолт-система мозга.

Для цитирования: Савостьянов А.Н., Кулешов Д.А., Клемешова Д.И., Власов М.С., Сапрыгин А.Е. Ассоциация аутистических личностных черт у неклинических испытуемых с показателями ЭЭГ в условиях просмотра видеозаписей лица. *Вавиловский журнал генетики и селекции*. 2024;28(8):1018-1024. doi 10.18699/vjgb-24-108

Финансирование. Часть исследования, касающаяся подготовки психологических тестов, подбора экспериментальных групп и регистрации ЭЭГ, выполнена при финансовой поддержке Российского научного фонда, научный проект № 22-15-00142: «фМРТ и ЭЭГ корреляты фокуса внимания на собственной персоне как фактора предрасположенности к аффективным расстройствам».

Благодарности. Разработка программно-аппаратного модуля проведена в рамках бюджетного проекта ИЦиГ СО РАН № FWNR-2022-0020.

Association of autistic personality traits with the EEG scores in non-clinical subjects during the facial video viewing

A.N. Savostyanov [1, 2, 3], D.A. Kuleshov [1, 4, 5], D.I. Klemeshova[1, 2], M.S. Vlasov [6], A.E. Saprygin [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Scientific Research Institute of Neurosciences and Medicine, Novosibirsk, Russia
[3] Novosibirsk State University, Novosibirsk, Russia
[4] Trofimuk Institute of Petroleum Geology and Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[5] Siberian State University of Telecommunications and Informatics, Novosibirsk, Russia
[6] Altai State Pedagogical University, Biysk Branch named after V.M. Shukshin, Biysk, Russia
a-sav@mail.ru

© Савостьянов А.Н., Кулешов Д.А., Клемешова Д.И., Власов М.С., Сапрыгин А.Е., 2024
Контент доступен под лицензией Creative Commons Attribution 4.0

Abstract. A software information module of the experimental computer platform "EEG_Self-Construct" was developed and tested in the framework of this study. This module can be applied for identification of neurophysiological markers of self-referential processes based on the joint use of EEG and facial video recording to induce the brain's functional states associated with participants' personality traits. This module was tested on a group of non-clinical participants with varying degrees of severity of autistic personality traits (APT) according to the Broad Autism Phenotype Questionnaire. The degree of individual severity of APT is a quantitative characteristic of difficulties that a person has when communicating with other people. Each person has some individual degree of severity of such traits. Patients with autism are found to have high rates of autistic traits. However, some individuals with high rates of autistic traits are not accompanied by clinical symptoms. Our module allows inducing the brain's functional states, in which the EEG indicators of people with different levels of APT significantly differ. In addition, the module includes a set of software tools for recording and analyzing brain activity indices. We have found that relationships between brain activity and the individual level of severity of APT in non-clinical subjects can be identified in resting-state conditions following recognition of self-referential information, while recognition of socially neutral information does not induce processes associated with APT. It has been shown that people with high scores of APT have increased spectral density in the delta and theta ranges of rhythms in the frontal cortical areas of both hemispheres compared to people with lower scores of APT. This could hypothetically be interpreted as an index of reduced brain activity associated with recognition of self-referential information in people with higher scores of autistic traits. The software module we are developing can be integrated with modules that allow identifying molecular genetic markers of personality traits, including traits that determine the predisposition to mental pathologies.

Key words: information-digital platforms in medicine; neurocomputation technologies; resting-state EEG; autistic personality traits; Broad Autism Phenotype; self-referential processing; default-mode network.

For citation: Savostyanov A.N., Kuleshov D.A., Klemeshova D.I., Vlasov M.S., Saprygin A.E. Association of autistic personality traits with the EEG scores in non-clinical subjects during the facial video viewing. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):1018-1024. doi 10.18699/vjgb-24-108

Введение

Разработка новых подходов выявления предрасположенности к определенным типам поведения, включая повышенный риск развития психических заболеваний, основана на обследовании людей при помощи генетических, нейрофизиологических и поведенческих методик, накоплении экспериментальной информации в базах данных и ее анализе с использованием широкого круга информационных технологий (Ivanov et al., 2022; Lin et al., 2022).

Согласно современным представлениям, аутизм – это заболевание, которое связано с нарушениями в работе мозга и проявляется в социальной сфере (Baron-Cohen, 2002; Lavenne-Collot et al., 2023). Это заболевание обнаруживается в трех сферах: социальном взаимодействии, коммуникации (использовании вербальных и невербальных стимулов), а также ограниченными и повторяющимися моделями в поведении, интересах и деятельности (Baron-Cohen, 2009; Murray et al., 2017). В 1980-х годах аутизм был признан спектром состояний (нарушений), который может быть индивидуальным для каждого пациента (Lovaas, 1987).

Жесткой границы между «здоровым человеком» и «аутистом» не существует, так как каждому человеку можно присвоить определенную величину некоторых аутистических личностных черт (АЛЧ) (Piven et al., 1997). Чем выше их уровень, тем больше поведение субъекта напоминает поведение аутиста. Известно, что АЛЧ может приводить к клиническим нарушениям в поведении, если их величина превышает некоторый порог. Однако существует феномен «неклинического аутизма», когда человек с сильно выраженными АЛЧ не обращается за врачебной помощью. При этом значительная часть таких «неклинических аутистов» в ходе своей жизни демонстрируют уровень социальной успешности, не отличающийся от людей с низкой степенью выраженности аутистичности. Предполагается, что имеются некие компенсаторные механизмы, которые, возможно, формируются в зависимости от влияния окружающей среды и могут как ослаблять, так и усиливать проявление АЛЧ субъектов (Frith, 1991; Georgiades et al., 2017).

Поскольку аутизм и АЛЧ ассоциированы с поведенческими затруднениями в социальной коммуникации, в большинстве нейрофизиологических (Tsai el al., 2013; Tseng et al., 2015) и генетических (Genovese, Butler, 2023) исследований проводится сравнение мозговых реакций людей с разной степенью аутистичности на предъявление им внешних стимулов, распознавание которых существенно для регуляции межличностной коммуникации. Для экспериментального исследования феномена аутизма применяются такие подходы, как психологическое тестирование при помощи опросников, запись и анализ ЭЭГ в условиях стимуляции. В качестве таких стимулов обычно используются либо фотографии человеческих лиц (Harms et al., 2010; Tseng et al., 2015), либо речевые задания (Tsai et al., 2013). Однако существуют также исследования, согласно которым выраженность аутизма ассоциирована с мозговой активностью в условиях функционального покоя без распознавания внешних стимулов (Harikumar et al., 2021).

Один из эффективных методов – регистрация видеозаписей лиц для индукции психологических состояний, различающихся у участников обследования с разной степенью выраженности личностных черт (Сы и др., 2024).

Применяется также подход, при котором ЭЭГ записывается без какой-либо дополнительной стимуляции. В его основе лежит гипотеза о функциональной роли дефолт-системы головного мозга в организации процессов самореференции. Дефолт-система – это совокупность участков коры, которые демонстрируют повышенную активацию в условиях покоя, но снижают уровень активации при

выполнении заданий, связанных с вниманием к внешней стимуляции. Дефолт-система рассматривается как мозговая структура, вовлеченная в оценку социально значимых стимулов, которые субъект относит к собственной личности (Northoff et al., 2005). Предполагается, что клинические формы аутизма сопровождаются снижением активности дефолт-системы (Ronde et al., 2024). Функции дефолт-системы могут быть ассоциированы не только с особенностями социального поведения людей, но и с особенностями их генома (Fanelli et al., 2024).

Ранее мы предложили подход для совместной регистрации и обработки ЭЭГ и видеозаписей лица, позволяющий совмещать анализ мозговой активности с оценкой динамики мимических мышц (Савостьянов и др., 2022). В этом исследовании мы предлагаем методику, которая предполагает использование фрагментов видеозаписей, полученных на первом этапе обследования, для стимуляции участников на более поздних этапах обследования. Как показано ниже, такой подход дает полезную информацию для выявления маркеров аутистических черт у неклинических испытуемых.

Для информационной поддержки проводимых исследований нами разрабатывается информационно-программный блок “EEG_AutisticTrait”, представляющий собой важную компоненту экспериментально-компьютерного модуля “EEG_Self-Construct”. Он обеспечивает полный цикл информационной поддержки исследований, включая: (а) накопление и хранение результатов обследования людей при помощи психологических, нейрофизиологических и генетических методик, позволяющих обнаружить индивидуальные особенности социальных коммуникаций, ассоциированных с аутизмом; (б) компьютерную обработку экспериментальных данных с применением методов регрессионного, корреляционного и факторного анализа, сопоставляющих поведенческие и нейрофизиологические показатели (Сы и др., 2024); (в) визуализацию первичных экспериментальных данных и результатов анализа.

Принципиальная новизна предлагаемого подхода состоит в том, что для выявления нейрофизиологических маркеров АЛЧ используются временные интервалы ЭЭГ-записей в условиях покоя в промежутках между распознаванием самоотнесенных или несамотнесенных стимулов. Такой подход дает возможность индуцировать состояния психики, связанные с самореференцией, в интервалах функционального покоя.

Материалы и методы

Последовательность этапов работы экспериментально-компьютерного модуля “EEG_Self-Construct” и перечень необходимого для реализации этих этапов программного обеспечения представлены в табл. 1. Модуль содержит как программные продукты, разработанные сотрудниками Института цитологии и генетики СО РАН, так и программы, взятые из открытых источников. Все блоки модуля позволяют осуществлять полный цикл сбора и обработки данных, необходимых для установления маркеров аутистических личностных черт.

Испытуемые. В исследовании принимали участие добровольцы, которые в большинстве были студентами Новосибирского государственного университета. Экспериментальная выборка включала 43 участника, возраст которых варьировал от 18 до 48 лет (19 мужчин и 24 женщины). Все участники на момент обследования не имели неврологических или психических заболеваний и не использовали никаких психоактивных веществ или фармакологических препаратов. Участники давали информированное согласие на прохождение экспериментального обследования в соответствии с Хельсинкской деклара-

Таблица 1. Перечень этапов работы модуля и программных средств, необходимых для их выполнения

Название этапа работы модуля	Программные пакеты, необходимые для реализации этапа	Результат прохождения этапа работы модуля
Этап 1. Извлечение списков генов-кандидатов и мозговых структур, ассоциированных с личностными особенностями из текстов на естественном языке	ANDSystem Software	Перечень генов-кандидатов для психогенетических исследований, списки мозговых структур для нейрофизиологических исследований
Этап 2. Планирование экспериментального дизайна и обработки данных	EventIDE	Протоколы поведенческих и нейрофизиологических экспериментов, протоколы обработки данных
Этап 3. Разработка экспериментальных парадигм психофизиологического обследования	EventIDE, Millisecond Software	Программные скрипты для проведения экспериментов
Этап 4. Регистрация ЭЭГ/ЭКГ-сигналов	NeoRec System	Записи ЭЭГ- и ЭКГ-сигналов с расстановкой меток событий
Этап 5. Разработка базы экспериментальных данных	ICBrainDB	Сетевая база психологических, нейрофизиологических и генетических данных
Этап 6. Предобработка ЭЭГ/ЭКГ-сигналов	EEGlab_toolbox	Нейрофизиологические сигналы, очищенные от нерелевантного шума
Этап 7. Локализация источников сигналов на поверхности коры головного мозга	EEGLab_toolbox, eLoreta	Карты локализации источников мозговой активности
Этап 8. Статистическая обработка поведенческих, физиологических и генетических данных	IBM SPSS Statistics, GNU PSPP	Результаты статистических сравнений экспериментальных выборок, перечень нейрофизиологических и генетических маркеров личностных черт

цией по биомедицинской этике. Протокол эксперимента был одобрен этическим комитетом НИИ нейронаук и медицины.

Психологическое тестирование выполнено при помощи специальной интернет-формы, реализованной на платформе Yandex сотрудниками ИЦиГ СО РАН. Все участники заполняли русскоязычную версию теста BAPQ для оценки выраженности аутистических черт (Hurley et al., 2007; перевод М.С. Власова). Этот тест включает 36 вопросов, касающихся способности человека контролировать свое поведение в ситуациях социального общения. Кроме того, участники заполняли психологические опросники на личностную и ситуативную тревожность Ч. Спилбергера (Spielberger, 1970; русскоязычная адаптация (Ханин, 1976)), опросник для оценки личностных черт Л. Голдберга «Маркеры факторов “Большой пятерки”» (перевод и валидация Г.Г. Князева с соавторами (2010)), опросник на аффилиацию со своей семьей (Cross et al., 2000), опросник на эмоциональный интеллект (Князев и др., 2012).

Экспериментальный метод. Программа для проведения эксперимента реализована на платформе Inquisit (https://www.millisecond.com/). В эксперименте участники выполняли три условия. В первом условии ЭЭГ регистрировали 12 мин без функциональной нагрузки. У испытуемого регистрировали три интервала по 2 мин с закрытыми глазами и три интервала по 2 мин с открытыми глазами. В интервалах, когда испытуемый открывал глаза, он/она видели черный экран монитора компьютера. В этот период у испытуемого совместно с ЭЭГ записывалось видеоизображение лица в течение всех 12 минут. Второе и третье условие отличались от первого тем, что при втором условии при открытых глазах испытуемый наблюдал видеозапись своего собственного лица, полученную при выполнении первого фонового задания, а при третьем условии ему/ей предъявлялась видеозапись лица незнакомого человека (всегда мужчины для испытуемого-мужчины, и женщины – для испытуемой-женщины). Порядок второго и третьего заданий менялся случайно.

Регистрация и обработка ЭЭГ. Для регистрации нейрофизиологических данных использовалась программа NeoRec, разработанная компанией «Медицинские компьютерные системы» (https://mks.ru/). ЭЭГ регистрировалась при помощи 130-канального усилителя NVX-132, Россия, 128 каналов ЭЭГ, расположенных по международной схеме 5-5 % с референтным электродом Cz, заземляющим электродом AFz; полоса пропускания 0.1–100 Гц, частота дискретизации сигнала 1000 Гц. Кроме ЭЭГ, дополнительно регистрировались ЭОГ и ЭКГ.

Мышечные и иные артефакты были удалены из ЭЭГ при помощи анализа независимых компонент с использованием программного пакета EEGlab_toolbox (Delorme, Makeig, 2004; https://sccn.ucsd.edu/eeglab/index.php). Затем из записей ЭЭГ извлекались фрагменты, соответствующие периодам, когда участник сидел с закрытыми глазами. Анализ далее проводился только для тех интервалов ЭЭГ-записей, в которых участник не видел ни видеозаписей, ни пустого экрана, но которые были записаны непосредственно после наблюдения соответствующих стимулов. После извлечения этих фрагментов ЭЭГ они разделялись на двухсекундные временные интервалы. Дальнейший анализ проводился с помощью программного пакета eLoreta (Pascual-Margui, 2002; https://www.uzh.ch/keyinst/loreta.htm).

В нашем случае выявленные при использовании eLoreta нейрофизиологические состояния сопоставлялись с психологическими особенностями испытуемых для определения маркеров АЛЧ. Для каждого двухсекундного интервала вычислялись значения спектральной плотности в частотных диапазонах дельта (2–4 Гц), тета (4–8 Гц), альфа-1 (8–10 Гц), альфа-2 (10–12 Гц), бета-1 (12–16 Гц), бета-2 (16–20 Гц), бета-3 (20–25 Гц) и гамма (25–35 Гц) ритмов. Далее для каждого участника вычислялся показатель суммарного спектра на всем интервале ЭЭГ-тестирования отдельно для каждого из трех экспериментальных условий (у каждого участника для этого использовалось от 150 до 170 двухсекундных интервалов). Спектры считались независимо для каждого из 128 каналов ЭЭГ, включенных в обработку данных. Отрезок записи ЭЭГ с частотой дискредитации 1000 Гц длительностью 3000 мс после появления блока был использован для расчета спектральной плотности источников в программе eLoreta (Pascual-Margui, 2002).

Статистический анализ. Оценку валидности психологических тестов проводили при помощи программного пакета IBM SPSS software, IBM, https://www.ibm.com/spss. Регрессионный анализ был выполнен в пакете eLoreta для поиска зависимости спектральной плотности от показателей индивидуальной выраженности аутистических черт независимо для каждого из трех экспериментальных условий. Дополнительную поправку на множественные сравнения не проводили.

Результаты

Результаты психологического тестирования

Для установления валидности русской версии теста BAPQ мы определили согласованность ответов на 36 вопросов этого теста по критерию Кронбаха. Значение альфа Кронбаха составило 0.838, что указывает на достаточно высокую согласованность разных пунктов опросника между собой. Кроме того, мы провели анализ корреляции индивидуальной выраженности АЛЧ с оценками по различным шкалам хорошо валидизированных психологических тестов. В табл. 2 представлена соотнесенность результатов теста BAPQ с другими рассмотренными в этом исследовании личностными чертами. Выраженность АЛЧ достоверно положительно коррелирует с тревожностью и отрицательно – с экстраверсией, способностью выражать позитивные эмоции и аффилиацию с семьей.

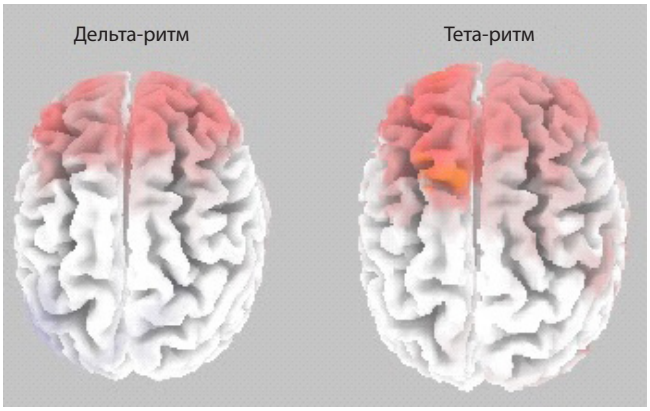
Результаты eLoreta при выявлении эффектов аутистических черт

Корреляции между показателем АЛЧ по тесту BAPQ были статистически достоверны только для условия «свое лицо» ($p = 0.0340$) в частотных диапазонах дельта (2–4 Гц) и тета (4–8 Гц) ритмов (см. рисунок). Для обоих частотных диапазонов eLoreta выявила положительную ассоциацию между показателями спектральной плотности и индивидуальной выраженностью аутистических черт

Таблица 2. Коэффициенты корреляции Пирсона между значениями по тесту BAPQ и шкалами других личностных опросников

BAPQ	Личностная тревожность	Экстраверсия	Семейные ценности	Способность выражать положительные эмоции
Корреляция Пирсона	0.407**	–0.524**	–0.351**	–0.278*
Достоверность (двухсторонняя)	0.002	0.003	0.007	0.036
N	43	43	43	43

* Корреляция значима на уровне 0.05 (двухсторонняя).
** Корреляция значима на уровне 0.01 (двухсторонняя).



Корреляция спектральной плотности в диапазоне дельта (2–4 Гц) и тета (4–8 Гц) ритмов с выраженностью аутистических черт (BAPQ) в группе из 43 участников для ЭЭГ-интервалов с закрытыми глазами между наблюдением своего лица.

Области коры, в которых найдены положительные корреляции АЛЧ со спектральной плотностью ($p < 0.04$), отмечены красным цветом. Наблюдается достоверно положительная ассоциация между АЛЧ и спектральной плотностью в лобных отделах обоих полушарий.

в лобных отделах коры обоих полушарий, т. е. более высокие значения АЛЧ соответствовали большим показателям спектральной плотности. Для условия «пустой экран» достоверность отсутствовала ($p = 0.28640$). Для условия «чужое лицо» ($p = 0.0932$) p-уровень был близок к достоверному значению, но не достигал его.

Обсуждение

Выявление нейрофизиологических маркеров личностных черт, включая черты, ассоциированные с предрасположенностью к болезням, предполагает использование сложного многокомпонентного инструментария для планирования экспериментов, сбора, хранения и анализа данных, сопоставления результатов разных исследований и организацию доступа к разным программам и полученным при их помощи данным. Важная составляющая такого инструментария – возможность разрабатывать и реализовывать новые парадигмы проведения нейрофизиологических исследований. Например, Ц. Сы с коллегами (2024) разработали программный модуль для выявления межнациональных особенностей в процессах самоотнесения информации к самому испытуемому или другим людям, имеющий существенное значение для поиска маркеров депрессии.

В рамках поиска маркеров предрасположенности к психиатрическим расстройствам важными задачами являются реконструкция и анализ генных сетей, лежащих в основе регуляции психоэмоциональных состояний у человека и животных (Savostyanov, Makarova, 2024). Пример модуля, нацеленного на реконструкцию и сопоставление генных сетей личностной тревожности у мышей и человека, описан в статье В.А. Савостьянова и А.А. Макаровой (Savostyanov, Makarova, 2024). Благодаря этому модулю могут быть выявлены структуры головного мозга, в которых обнаруживается дифференциальная экспрессия генов у животных, различающихся по своему уровню личностной тревожности. В дальнейшем такие структуры могут быть рассмотрены в качестве областей интереса для нахождения нейрофизиологических маркеров тревожного расстройства у человека.

Была проведена апробация информационно-программного блока “EEG_AutisticTrait” для выявления нейрофизиологических маркеров аутистических личностных черт. При использовании специальной платформы Yandex было осуществлено комплексное тестирование участников с помощью нескольких опросников, включая тесты на индивидуальную выраженность АЛЧ (русская версия теста BAPQ). Значение альфа Кронбаха для разных вопросов теста BAPQ оказалось равным 0.83, что говорит о достаточно высокой внутренней согласованности этого теста. Также были найдены отрицательные корреляции АЛЧ с экстраверсией, эмоциональным интеллектом и важностью семейных ценностей и положительные корреляции между АЛЧ и тревожностью, что хорошо согласуется с общим представлением психологов об аутистических чертах.

На нейрофизиологическом уровне были обнаружены положительные корреляции между АЛЧ и спектральной плотностью в диапазонах дельта- и тета-ритма для экспериментального условия, ассоциированного с самореференцией визуальной информации, но не было установлено достоверных взаимосвязей при условиях, следующих за просмотром социально нейтрального стимула (пустой экран), или информации, отнесенной к другим людям. Согласно литературным данным (Knyazev, 2007), высокие фоновые значения спектральной плотности дельта- и тета-ритма чаще всего интерпретируются как индикатор сниженной функциональной активности мозга. При таком подходе наши результаты могут быть гипотетически объяснены как коррелят пониженной активности мозга в условиях, следующих за предъявлением самоотнесенной информации, у людей с высокими аутистическими чертами в сравнении с людьми с низким уровнем аутистичности.

Существенно, что мы выявили нейрофизиологические корреляты АЛЧ только для самореферентного условия. В социально нейтральном условии какие-либо тенденции к зависимости между АЛЧ и мозговой активностью полностью отсутствовали, тогда как для условия «чужое лицо»

наблюдалась маргинальная статистическая тенденция к достоверности результата. Можно допустить, что ЭЭГ-активность покоя у неклинических испытуемых слабо ассоциирована с их уровнем аутистичности, что объясняет неудачу предыдущих попыток найти какие-либо взаимосвязи АЛЧ с фоновой ЭЭГ у таких участников. Однако просмотр видеозаписей, отнесенных к самому участнику (и в меньшей степени к другим людям), активирует в головном мозге процессы, связанные с распознаванием социально значимой информации, что делает ЭЭГ-показатели более зависимыми от АЛЧ, чем в случае просмотра социально нейтральных стимулов.

Заключение

Предлагаемый нами подход основан на интеграции психологических и нейрофизиологических методов сбора и анализе данных. В дальнейшем предполагается провести исследование зависимости АЛЧ от генетических особенностей испытуемых. Также желательно изучить влияние уровня экспрессии различных генов в головном мозге на выраженность личностных черт. Строгая количественная оценка уровня экспрессии генов в головном мозге не может быть выполнена на живом человеке, что предполагает необходимость совмещения данных, полученных на людях и на экспериментальных животных (Savostyanov, Makarova, 2024). Такое исследование требует разработки особого инструментария для накопления, хранения и анализа данных, который будет создан на базе платформы «Биоинформатика и системная компьютерная биология». В дальнейшем этот инструментарий может быть применен для изучения нейрофизиологических коррелятов различных личностных черт у людей без клинических нарушений и при патологиях, что даст возможность проводить новые комплексные исследования в рамках системной нейробиологии.

Список литературы / References

Князев Г.Г., Митрофанова Л.Г., Бочаров А.В. Валидизация русскоязычной версии опросника Л. Голдберга «Маркеры факторов “Большой пятерки”» *Психол. журнал*. 2010;31(5): 100-110
[Knyazev G.G., Mitrofanova L.G., Bocharov A.V. Validization of Russian version of Goldberg’s “Big-five factor markers” inventory. *Psikhologicheskii Zhurnal.* 2010;31(5):100-110 (in Russian)]

Князев Г.Г., Митрофанова Л.Г., Разумникова О.М., Барчард К. Адаптация русскоязычной версии «Опросника эмоционального интеллекта» К. Барчард. *Психол. журнал.* 2012;33(4): 112-120
[Knyazev G.G., Mitrofanova L.G., Razumnikova O.M., Barchard K. Adaptation of Russian language version of K. Barchard’s Emotional Intelligence Questionnaire. *Psikhologicheskii Zhurnal.* 2012;33(4): 112-120 (in Russian)]

Савостьянов А.Н., Вергунов Е.Г., Сапрыгин А.Е., Лебедкин Д.А. Апробация технологии оценки мимики лиц для изучения динамики функциональных состояний человека в ЭЭГ-парадигме покоя. *Вавиловский журнал генетики и селекции*. 2022;26(8): 765-772. doi 10.18699/VJGB-22-92
[Savostyanov A.N., Vergunov E.G., Saprygin A.E., Lebedkin D.A. Validation of a face image assessment technology to study the dynamics of human functional states in the EEG resting-state paradigm. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2022;26(8):765-772. doi 10.18699/VJGB-22-92]

Сы Ц., Тянь Ц., Савостьянов В.А., Лебедкин Д.А., Бочаров А.В., Савостьянов А.Н. Сопоставление показателей мозговой активности у китайских и российских студентов в условиях распознавания информации, отнесенной к себе и другим людям. *Вавиловский журнал генетики и селекции*. 2024;28(8):982-992. doi 10.18699/vjgb-24-105
[Si Q., Tian J., Savostyanov V.A., Lebedkin D.A., Bocharov A.V., Savostyanov A.N. Comparison of brain activity indexes in the Chinese and Russian students under recognition of self- and other-related information. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding.* 2024;28(8):982-992. doi 10.18699/vjgb-24-105]

Ханин Ю.Л. Краткое руководство к шкале реактивной и личностной тревожности Ч.Д. Спилбергера. Л.: ЛНИИФК,1976
[Khanin Yu.L. Quick Guide to C.D. Spielberger’s Scale of State and Trait Anxiety. Leningrad, 1976 (in Russian)]

Baron-Cohen S. The extreme male brain theory of autism. *Trends Cogn. Sci*. 2002;6(6):248-254. doi 10.1016/s1364-6613(02)01904-6

Baron-Cohen S. Autism: the empathizing-systemizing (E-S) theory. *Ann. N.Y. Acad. Sci.* 2009;1156:68-80. doi 10.1111/j.1749-6632.2009.04467.x

Cross S.E., Bacon P.L., Morris M.L. The relational-interdependent self-construal and relationships. *J. Pers. Soc. Psychol*. 2000;78(4): 791-808

Delorme A., Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*. 2004;134(1):9-21. doi 10.1016/j.jneumeth.2003.10.009

Fanelli G., Robinson J., Fabbri C., Bralten J., Mota N.R., Arenella M., Sprooten E., Franke B., Kas M., Andlauer T.F., Serretti A. Shared genetics linking sociability with the brain’s default mode network. *medRxiv*. [Preprint]. 2024. May 25:2024.05.24.24307883. doi 10.1101/2024.05.24.24307883

Frith U. Asperger and his syndrome. In: Frith U. (Ed.). Autism and Asperger Syndrome. Cambridge University Press, 1991;1-36

Genovese A., Butler M.G. The autism spectrum: behavioral, psychiatric and genetic associations. *Genes* (*Basel*). 2023;14(3):677. doi 10.3390/genes14030677

Georgiades S., Bishop S.L., Frazier T. Editorial perspective: longitudinal research in autism − introducing the concept of ‘chronogeneity’. *J. Child Psychol*. *Psychiatry*. 2017;58:634-636. doi 10.1111/jcpp.12690

Harikumar A., Evans D.W., Dougherty C.C., Carpenter K.L.H., Michael A.M. A review of the default mode network in autism spectrum disorders and attention deficit hyperactivity disorder. *Brain Connect*. 2021;11(4):253-263. doi 10.1089/brain.2020.0865

Harms M.B., Martin A., Wallace G.L. Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychol. Rev*. 2010;20(3):290-322. doi 10.1007/s11065-010-9138-6

Hurley L., Parlier M., Reznick J., Piven J. The broad autism phenotype questionnaire. *J. Autism Dev. Disord.* 2007;37(9):1679-1690. doi 10.1007/s10803-006-0299-3

Ivanov R., Kazantsev F., Zavarzin E., Klimenko A., Milakhina N., Matushkin Y.G., Savostyanov A., Lashin S. ICBrainDB: An integrated database for finding associations between genetic factors and EEG markers of depressive disorders. *J. Pers. Med.* 2022;12(1):53. doi 10.3390/jpm12010053

Knyazev G.G. Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neurosci. Biobehav. Rev.* 2007;31(3): 377-395. doi 10.1016/j.neubiorev.2006.10.004

Lavenne-Collot N., Tersiguel M., Dissaux N., Degrez C., Bronsard G., Botbol M., Berthoz A. Self/other distinction in adolescents with autism spectrum disorder (ASD) assessed with a double mirror paradigm. *PLoS One*. 2023;18(3):e0275018. doi 10.1371/journal.pone.0275018

Lin M., Wang Y., Lopez-Naranjo C., Hu S., Reyes R.C.G., Paz-Linares D., Areces-Gonzalez A., Hamid A.I.A., Evans A.C., Savostyanov A.N., Calzada-Reyes A., Villringer A., Tobon-Quinero C.A., Garcia-Agustin D., Yao D., Dong L., Aubet-Vazquez E., Reza F.,

Razzaq F.A., Omar H., Abdullah J.M., Galler J.R., Ochoa-Gomez J.F., Prichep L.S., Galan-Garcia L., Morales-Chacon L., Valdes-Sosa M.J., Trondle M., Zulkifly M.F.M., Rahman M.R.B.A., Milakhina N.S., Langer N., Rudych P., Koenig T., Virues-Alba T.A., Lei X., Bringas-Vega M.L., Bosch-Bayard J.F., Valdes-Sosa P.A. Harmonized-Multinational qEEG norms (HarMNqEEG). *NeuroImage*. 2022;256:119190. doi 10.1016/j.neuroimage.2022.119190

Lovaas O.I. Behavioral treatment and normal educational and intellectual functioning in young autistic children. *J. Consult. Clin. Psychol.* 1987;55(1):3-9. doi 10.1037/0022-006x.55.1.3

Murray K., Johnston K., Cunnane H., Kerr Ch., Spain D., Gillan N., Hammond N., Murphy D., Happe F. A new test of advanced theory of mind: The "Strange Stories Film Task" captures social processing differences in adults with autism spectrum disorders. *Autism Res*. 2017;10(6):1120-1132. doi 10.1002/aur.1744

Northoff G., Heinzel A., De Greck M., Bermpohl F., Dobrowolny H., Panksepp J. Self-referential processing in our brain – a meta-analysis of imaging studies on the self. *NeuroImage*. 2005;31(1):440-457. doi 10.1016/j.neuroimage.2005.12.002

Pascual-Margui R.D. Standardized low-resolution brain electromagnetic tomography (sLORETA). Technical details. *Methods Find. Exp. Clin. Pharmacol*. 2002;24(Suppl. D):5-12

Piven J., Palmer P., Jacobi D., Childress D., Arndt S. Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *Am. J. Psychiatry*.1997;154(2):185-190. doi 10.1176/ajp.154.2.185

Ronde M., van der Zee E.A., Kas M.J.H. Default mode network dynamics: an integrated neurocircuitry perspective on social dysfunction in human brain disorders. *Neurosci. Biobehav. Rev*. 2024;164: 105839. doi 10.1016/j.neubiorev.2024.105839

Savostyanov V.A., Makarova A.A. Reconstruction and analysis of the gene network for regulation of trait anxiety level in mice by means of ANDSystem software. In: IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM), Altai, Russian Federation, 2024;2340-2343. doi 10.1109/EDM61683.2024.10615053

Spielberger C.D., Gorsuch R.L., Lushene R.E. Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press, 1970

Tsai A.C., Savostyanov A.N., Wu A., Evans J.P., Chien V.S.C., Yang H.-H., Yang D.-Y., Liou M. Recognizing syntactic errors in Chinese and English sentences: brain electrical activity in Asperger's syndrome. *Res. Autism Spectr. Disord*. 2013;7(7):889-905. doi 10.1016/j.rasd.2013.02.001

Tseng Y.L., Yang H.H., Savostyanov A.N., Chien V.S., Liou M. Voluntary attention in Asperger's syndrome: brain electrical oscillation and phase-synchronization during facial emotion recognition. *Res. Autism Spectr. Disord*. 2015;13-14:32-51. doi 10.1016/j.rasd.2015.01.003

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Поступила в редакцию 28.10.2024. После доработки 12.11.2024. Принята к публикации 13.11.2024.

Алфавитный указатель авторов статей, опубликованных в журнале в 2024 г.

Агафонов А.В. 1,5
Агеева Е.В. 5, 523
Адамовская А.В. 8, 882, 960
Адонина И.Г. 4, 377; 5, 506
Адоньева Н.В. 2, 185
Азовцева А.И. 1, 108
Айтназаров Р.Б. 1, 117
Алексеев В.Ю. 3, 276
Алпатьева Н.В. 2, 175
Амстиславский С.Я. 7, 744
Андреенков О.В. 2, 131
Андреенкова О.В. 2, 185
Андреюшкова Д.А. 7, 759
Антоненко О.В. 2, 131
Антропова Е.А. 8, 904, 960
Арешева О.М. 1, 44
Арипов В.С. 2, 249
Арссан М.А. 4, 407
Афоничева К.В. 6, 667
Афонников Д.А. 4, 443; 8, 854, 874
Афонникова С.Д. 2, 138

Багмет В.Б. 7, 706
Бадаева Е.Д. 5, 506; 7, 716
Баканова М.Л. 4, 424
Бакоев С.Ю. 6, 640
Баранова Е.Д. 2, 204
Баранова О.А. 4, 377
Басит С. 3, 326
Басов В.И. 5, 506
Басов Н.В. 8, 882
Баттулин Н.Р. 2, 138; 5, 487
Бебякина И.В. 5, 506
Беклемишев А.Б. 5, 554
Беклемишева В.Р. 7, 759
Белокопытова И.И. 4, 407
Бердникова А.А. 6, 628
Боброва М.М. 6, 640
Бобрышева А.А. 3, 332
Богачев А.С. 8, 993
Богачев С.С. 7, 696; 8, 993
Богомолов А.Г. 8, 950
Болдаков Д.М. 5, 506
Боме Н.А. 3, 263
Бондарь А.А. 5, 554
Борисенко Н.В. 1, 63
Бородин П.М. 6, 592
Бочарникова М.Е. 8, 854
Бочаров А.В. 8, 982
Брусенцев Е.Ю. 7, 744
Бужан Э. 7, 752
Булатов И.О. 6, 602
Бурханова Г.Ф. 3, 276
Бурыгин Г.Л. 3, 308
Бусов И.Д. 4, 443
Бутикова Е.А. 8, 882

Валихова Л.В. 6, 659
Васильев Г.В. 1,80; 7, 780
Васильев С.А. 2, 198
Васильева О.Ю. 1, 55; 2, 198
Вахрамеев А.Б. 1, 108
Велич Л. 7, 752
Венгушт Г. 7, 752
Венгушт Д.Ж. 7, 752
Вензель А.С. 8, 882
Веселова С.В. 3, 276
Власов М.С. 8, 1018
Воевода М.И. 5, 554
Волкова Е.И. 2, 131
Волкова Н.В. 2, 249
Волянская А.Р. 8, 960
Ворожейкин П.С. 8, 834
Вьюшина А.В. 4, 387
Вяткин Ю.В. 7, 679

Гайслер Е.В. 8, 882
Ганчева М.С. 2, 175
Генаев М.А. 4, 443; 6, 602
Гетманцева Л.В. 6, 640
Гилинский М.А. 3, 299
Глаголева А.Ю. 4, 443
Глушков А.Н. 4, 424
Голованова Е.В. 5, 563
Голубенко М.В. 5, 467
Гончаров Н.П. 2, 155; 7, 716
Гордеева Е.И. 5, 495; 6, 602
Графодатский А.С. 7, 759
Гребенчук А.Е. 1, 98
Грибкова О.В. 2, 239
Григорьева Е.В. 7, 679
Гришко Е.О. 6, 592
Грунтенко Н.Е. 2, 185
Гурина А.А. 2, 175
Гурков А.Н. 3, 317
Гусев О.А. 2, 138

Давоян Р.О. 5, 506; 7, 716
Давоян Э.Р. 5, 506
Даркова Я.А. 2, 198
Дашкевич С.М. 3, 263
Деведжич А. 7, 752
Девяткин В.А. 3, 351
Демаков С.А. 2, 131
Деменков П.С. 2, 204; 8, 808, 864, 882, 904, 927, 960
Дементьева Н.В. 1, 108
Денисова Е.И. 3, 288
Деренко М.В. 1, 90

Держинский Е.А. 5, 563
Джос Е.А. 6, 619
Дидоренко С.В. 5, 515
Дмитриев А.Э. 2, 198
Долгова Е.В. 8, 993
Дорогина О.В. 1, 5, 55
Досжанова Б.Н. 5, 515
Доценко П.А. 8, 808
Дроздова П.Б. 3, 317
Дружинин В.Г. 2, 204
Дыкман Л.А. 3, 308
Дысин А.П. 1, 108

Евдокимов А.Н. 7, 688
Елгаева Е.Е. 6, 628
Емельянов В.В. 1, 44
Ефимов В.М. 2, 155, 185
Ефремов Я.Р. 8, 993

Жимулев И.Ф. 2, 131
Журавлева А.А. 4, 365

Закиян С.М. 7, 679
Заливина Е.А. 4, 398, 407
Зарецкий А.М. 7, 731
Зарубин А.А. 6, 659
Затыбеков А.К. 5, 515
Захарова Я.А. 4, 424
Землянская Е.В. 8, 918, 950, 974
Зинченко А.Н. 5, 506
Злобин А.С. 4, 456
Золотарева К.А. 8, 808, 1008
Зоркольцева И.В. 6, 628
Зощук С.А. 7, 716
Зубанова Ю.С. 5, 506
Зыкова Т.Е. 4, 443

Ибрагимов П.Ш. 4, 416
Ибрагимова Н.Н. 1, 33
Иванисенко В.А. 8, 808, 864, 882, 904, 927, 940, 960
Иванисенко Т.В. 8, 882, 904, 927, 960
Иванов М.К. 7, 780
Иванов Р.А. 8, 808, 874, 1008
Игнатьева Е.В. 8, 864
Игошин А.В. 1, 117; 2, 190
Ильина А.В. 1, 117
Ильичев А.А. 2, 249
Ильчибаева Т.В. 4, 398
Ильясов П.В. 2, 239
Интересова Е.А. 7, 759

Кабиева Ш.Ш. 6, 640
Казанцев Ф.В. 8, 874, 897
Карамышева Т.В. 8, 993
Карманов С.Ю. 8, 927
Катохин А.В. 5, 554
Кашеварова Н.М. 1, 15
Кенжегулов О.А. 1, 63
Кескинов А.А. 6, 640
Кирикович С.С. 8, 993
Клемешева Д.И. 8, 1018
Клещев М.А. 7, 780; 8, 904
Климонтов В.В. 8, 1008
Кнауб В.В. 5, 536
Коваль В.В. 8, 927
Коваль В.С. 4, 443
Кожевникова О.С. 3, 351
Козенева В.С. 7, 744
Колесников Н.А. 6, 659
Колчанов Н.А. 7, 696; 8, 807, 904, 927, 993
Комарова А.А. 4, 398
Комышев Е.Г. 4, 443
Кондаурова Е.М. 4, 398, 407
Коненков В.И. 4, 433
Кораблёва С.Ю. 5, 487
Коробейникова А.В. 6, 640
Корытина Г.Ф. 7, 792
Кочиева Е.З. 6, 619
Крадецкая О.О. 3, 263
Кулаева Е.Д. 3, 342
Кулешов Д.А. 8, 1018
Куликова Е.А. 4, 407
Куслий М.А. 5, 571

Лавреха В.В. 8, 950
Лаврик И.Н. 8, 882, 927
Лаврик О.И. 7, 688
Лактионов П.П. 2, 215
Ларионов А.В. 2, 204
Ларкин Д.М. 1, 117; 2, 190; 4, 416
Ласточкин В.В. 1, 44
Лашин С.А. 8, 808, 864, 874, 897, 1008
Лебедев И.Н. 2, 198
Лебедева Д.А. 7, 744
Лебедкин Д.А. 8, 982
Левитес Е.В. 8, 993
Левицкий В.Г. 8, 822
Леонова И.Н. 4, 456; 5, 523
Леплина О.Ю. 8, 993
Ли В. 6, 602
Лимарева Л.В. 2, 239
Лукъянчук И.В. 2, 166
Лыжин А.С. 2, 166
Лялина Е.В. 7, 716

Мадьярова Е.В. 3, 317
Макаров В.В. 6, 640
Макарова А.А. 8, 927
Макарова А.В. 1, 108
Макарова Е.Н. 3, 288
Максимов И.В. 3, 276
Малахова А.А. 7, 679
Маликов Д.Г. 5, 571
Мальцева А.В. 8, 904
Мальцева Е.К. 2, 138
Малярчук Б.А. 1, 90; 6, 650
Мамаев А.Л. 8, 993
Маркель А.Л. 3, 299
Мартынов А.А. 2, 190
Марченко И.В. 6, 667

Марченко М.А. 8, 940
Масленникова В.С. 5, 554
Матора Л.Ю. 3, 308
Матушкин Ю.Г. 8, 874, 897
Махмуд М. Имад А.М. 3, 326
Мацкова Л.В. 2, 204
Машкина Е.В. 3, 342
Медведев С.П. 7, 679
Мельникова П.М. 8, 927
Меркулова Т.И. 8, 822
Мещанинова М.И. 8, 993
Минина В.И. 4, 424
Минина Ю.М. 7, 679
Митрофанова О.В. 1, 108
Михайлова А.Д. 8, 864
Мишакова Т.М. 1, 117
Мишина А.И. 6, 640
Мишинов С.В. 8, 882
Модина С.А. 5, 571
Мокшина Н.Е. 1, 33
Молодцева А.С. 5, 571
Мосалев К.И. 5, 554
Музлаева Е.С. 3, 342
Мустафин З.С. 8, 874
Мустафин Р.Н. 2, 228; 5, 476
Мухин А.М. 8, 874

Надточий Ю.А. 7, 679
Назаров К.Д. 7, 688
Насибуллин Т.Р. 7, 792
Науменко В.С. 4, 398, 407
Науменко К.Н. 7, 688
Никитина Т.В. 2, 198
Николаева О.А. 1, 108
Никонов С.Д. 7, 696; 8, 993
Никулин А.Ю. 7, 706
Никулин В.Ю. 7, 706
Нимаев В.В. 4, 433
Нурисламов А.Р. 7, 688

Окотруб С.В. 7, 744
Олешко О.С. 8, 882
Омельянчук Н.А. 8, 918, 950, 974
Ооржак А.Ю. 6, 640
Ордян Н.Э. 4, 387
Орлов Ю.Л. 8, 904, 960
Осадчук А.В. 1,80; 7, 780
Осадчук Л.В. 1,80; 7, 780
Осечкова А.Е. 8, 882
Останин А.А. 7, 696; 8, 993
Отман Н. 3, 326
Ошихмина С.Г. 8, 993

Пальянов А.Ю. 8, 843
Пальянова Н.В. 8, 843
Панин В.М. 1, 63
Патрушев Ю.В. 8, 882
Перетолчина Т.Е. 3, 317
Першина Л.А. 6, 610
Петинцева А.А. 7, 792
Петрусева И.О. 7, 688
Петунина Ж.В. 3, 317
Пивина С.Г. 4, 387
Плотникова Л.Я. 5, 536
Подколодная О.А. 8, 940
Подколодный Н.Л. 8, 808, 940
Поздняков А.С. 8, 882
Покорны Б. 7, 752
Покровский А.Г. 8, 882, 927
Политыко Ю.К. 3, 299
Пономаренко М.П. 8, 808, 1008
Попов А.А. 7, 688
Потапова Н.А. 4, 456
Правикова П.Д. 4, 407
Прасолова М.А. 7, 780
Предтеченская Е.В. 8, 927
Приказюк Е.Г. 1, 44
Пристяжнюк И.Е. 2, 138; 5, 487
Притворова А.В. 4, 387
Прокофьев В.Ф. 4, 433
Проскурина А.С. 8, 993
Проскурякова А.А. 7, 759
Прудникова М.М. 2, 131
Пузырёв В.П. 5, 467
Пшеничникова Т.А. 6, 602
Пылаев Т.Е. 1, 63
Пыхтина М.Б. 5, 554

Рандзуан Х.М. 3, 326
Рахманова Т.А. 7, 744
Рейнбах Н.Р. 1, 108
Речкин Д.В. 2, 155
Рзаев Дж.А. 7, 679
Риттер Г.С. 8, 993
Рогачев А.Д. 8, 882, 927
Рогозина Е.В. 2, 175
Родный А.Я. 4, 398
Рожкова И.Н. 7, 744
Романенко С.А. 7, 759
Романов Е.М. 2, 148
Романов С.Е. 2, 215
Романова Е.В. 3, 317
Ромашов Г.А. 2, 190
Рузанова В.С. 8, 993
Румянцев С.Д. 3, 276
Румянцева Ю.В. 3, 351
Рыбаков М.А. 8, 974
Рыжкова А.С. 6, 583
Рябова А.Е. 1, 108

Савостьянов А.Н. 8, 982, 1018
Савостьянов В.А. 8, 982
Савченко Р.Р. 7, 770
Саженова Е.А. 2, 198
Салина Е.А. 4, 456; 5, 506
Самойлова Е.М. 2, 215
Сапрыгин А.Е. 8, 1018
Саранчина А.Е. 3, 317
Сарсенова С.Х. 1, 63
Селиванов Н.Ю. 1, 63

Серяпина А.А. 3, 299
Сибикеев С.Н. 4, 377
Сидоров С.В. 7, 696; 8, 993
Силкова О.Г. 4, 365
Симонов А.В. 6, 602
Скрябин Н.А. 7, 770
Смирнов А.В. 2, 138; 5, 487; 6, 583
Смольникова М.В. 6, 667
Соболева О.А. 4, 424
Соколова Д.В. 7, 731
Соловьева А.Е. 7, 731
Сорокан А.В. 3, 276
Сорокоумова А.А. 3, 299
Сотникова Ю.С. 8, 882
Степанов В.А. 6, 659
Стройл Б.К. 7, 752
Ступак В.В. 8, 882
Сузуки Т. 5, 515
Сухомясова А.Л. 6, 659
Сущенко Р.З. 7, 706
Сы Ц. 8, 982

Табанюхов К.А. 5, 554
Таранов О.С. 7, 696; 8, 993
Терещенко Н.А. 7, 716
Терещенко С.Ю. 6, 667
Тимашева Я.Р. 7, 792
Тимофеев М.А. 3, 317
Тимофеева А.А. 4, 424
Титов И.И. 8, 834
Титов Р.А. 4, 424
Тишакова К.В. 7, 759
Ткаченко А.Г. 1, 15, 24
Толмачева Е.Н. 2, 198
Торгунакова А.В. 4, 424
Трофимова М.Ф. 8, 897
Трубачеева Н.В. 6, 610
Туктарова И.А. 7, 792
Турнаев И.И. 8, 854
Туруспеков Е.К. 5, 515
Тянь Ц. 8, 982

Урзи Ф. 7, 752
Усенбеков Е.С. 4, 416
Утебаев М.У. 3, 263

Фергюсон-Смит М.А. 7, 759
Филюшин М.А. 6, 619
Фишман В.С. 7, 688
Фурсова А.Ж. 3, 351

Хабарова Е.А. 7, 679
Хамзина А.К. 4, 416
Хан Я.Н. 3, 326
Хаова Е.А. 1, 15, 24
Харьков В.Н. 6, 659
Хитринская И.Ю. 6, 659
Хлебодарова Т.М. 8, 897
Хлесткина Е.К. 5, 495
Хоцкин Н.В. 7, 744
Хуснутдинова Э.К. 2, 228

Центалович Ю.П. 3, 299
Цепилов Я.А. 4, 456; 6, 628
Цуканов А.В. 8, 822
Цыбовский И.С. 1, 98

Чадаева И.В. 8, 808
Чао Х. 8, 904, 960
Часовских Н.Ю. 3, 332
Черенко В.А. 8, 918
Черепанова Е.А. 3, 276
Чересиз С.В. 8, 882
Черных Е.Р. 7, 696; 8, 993
Чижик Е.Е. 3, 332
Чилимова И.В. 3, 263
Чиркова Т.В. 1, 44
Чудакова Д.А. 2, 215
Чэнь М. 8, 904, 960

Шабанова Е.В. 1, 5
Шавшаева Н.А. 7, 744
Шайхутдинов И.Х. 2, 239
Шамустакимова А.О. 1, 74
Шварц М.Б. 2, 131
Шевченко А.В. 4, 433
Шейкина О.В. 2, 148
Шеленга Т.В. 7, 731
Шелихова Е.В. 5, 554
Шестаков Д.А. 6, 640
Шеховцов С.В. 5, 563
Шишин К.С. 8, 1008
Шкляр А.А. 3, 351
Шнайдер Т.А. 2, 138; 5, 487
Шоева О.Ю. 5, 495
Шубина М.В. 6, 667
Шумный В.К. 5, 523; 6, 610

Щеголев С.Ю. 3, 308
Щенникова А.В. 6, 619
Щербаков Д.Н. 2, 249
Щербаков Д.Ю. 3, 317
Щербаков Ю.С. 1, 108

Эльконин Л.А. 1, 63
Эрдман В.В. 7, 792
Этерович Т. 7, 752

Юданова С.С. 1, 55
Юдин В.С. 6, 640
Юдин Н.С. 1, 117; 2, 190; 4, 416
Юдкин В.А. 7, 759
Юнусова А.М. 2, 138; 5, 487; 6, 583
Юрченко А.А. 4, 416

Якубов Л.А. 7, 696
Ямашита Й. 5, 515
Янг Ф. 7, 759
Янжекович Ф. 7, 752
Яньшоле Л.В. 3, 299
Яркова Е.С. 7, 679
Яцык И.В. 8, 960

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» до 2011 г. выходил под названием «Информационный вестник ВОГиС»/ "The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Сетевое издание «Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» – реестровая запись СМИ Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Russian Science Citation Index, Российский индекс научного цитирования, ВИНИТИ, Web of Science CC, Scopus, PubMed Central, DOAJ, ROAD, Ulrich's Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:
русскоязычная версия – на сайте https://vavilovj-icg.ru/
и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440

англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour
и платформе PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

При перепечатке материалов ссылка обязательна.

✉ email: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.
Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.
Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.
Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.
Дата публикации 26.12.2024. Формат 60 × 84 1/8. Уч.-изд. л. 30.5.