Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2026;30(2):299-310

doi 10.18699/vjgb-26-34

Разработка и валидация программы PipeSeq для анализа данных секвенирования РНК на модели *Chlamydomonas reinhardtii*

А.М. Нерезенко, П.А. Виролайнен, С.А. Тупицына, Е.М. Чекунова

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
alexnerezenko@gmail.com

Аннотация. Секвенирование РНК (РНК-сек) – высокочувствительный метод анализа транскриптома, позволяющий одновременно оценивать экспрессию тысяч генов и выявлять паттерны экспрессии в различных условиях. Существующее разнообразие форматов данных РНК-сек, методов нормализации и подходов к статистической обработке результатов затрудняет сопоставление данных разных исследований и снижает воспроизводимость анализа. В настоящей работе представлен автоматизированный пайплайн PipeSeq, объединяющий стандартные этапы обработки данных РНК-сек – от загрузки (SRA Toolkit), выравнивания прочтений на референсный геном (HISAT2) и сборки транскриптов (StringTie) до подсчета транскриптов (FeatureCounts) и статистического анализа дифференциальной экспрессии генов в различных экспериментальных условиях (DESeq2). Программа PipeSeq имеет простой визуальный интерфейс, поддерживает многопоточность и формирует готовые для анализа тепловые карты экспрессии генов и отчеты в форме таблиц и графиков. Функциональность пайплайна продемонстрирована на трех наборах пакетов сырых данных секвенирования РНК клеток зеленой водоросли *Chlamydomonas reinhardtii*, доступных в открытой базе данных NCBI SRA. Результаты этих экспериментов были использованы для анализа дифференциальной экспрессии генов *C. reinhardtii*, кодирующих факторы транскрипции семейства GATA, в различных световых условиях культивирования. Полученные методами *in silico* данные верифицированы методом полимеразной цепной реакции в реальном времени с обратной транскрипцией (ОТ-ПЦР-РВ) по 12 генам GATA, что позволило выдвинуть предположения об их функциях, а также оценить степень согласованности между массовым (РНК-сек) и таргетным (ОТ-ПЦР-РВ) подходами. Результаты нашего исследования показали, что методы секвенирования РНК и ОТ-ПЦР-РВ выявляют схожие направления изменения экспрессии генов, но демонстрируют различия по оценке степени размера эффекта и чувствительности, что подчеркивает необходимость совместного применения двух подходов. Таким образом, программа PipeSeq представляет собой инструмент для проведения полного цикла биоинформатического анализа данных РНК-сек, дает возможность обрабатывать данные ОТ-ПЦР-РВ и выполнять сравнительный статистический анализ полученных результатов.

Ключевые слова: секвенирование РНК; РНК-сек; ОТ-ПЦР-РВ; пайплайн; транскриптом; экспрессия генов; факторы транскрипции семейства GATA; ФТ GATA; *Chlamydomonas reinhardtii*

Для цитирования: Нерезенко А.М., Виролайнен П.А.,Тупицына С.А., Чекунова Е.М. Разработка и валидация программы PipeSeq для анализа данных секвенирования РНК на модели *Chlamydomonas reinhardtii*. *Вавиловский журнал генетики и селекции*. 2026; 30(2):299-310. doi 10.18699/vjgb-26-34

Финансирование. Работа выполнена при поддержке СПбГУ, шифр проекта 124032000041-1.

Development and validation of the PipeSeq program for RNA-seq data analysis in the *Chlamydomonas reinhardtii* as a model

A.M. Nerezenko, P.A. Virolainen, S.A. Tupitsyna, E.M. Chekunova

Saint-Petersburg University, St. Petersburg, Russia
alexnerezenko@gmail.com

Abstract. RNA sequencing (RNA-seq) is a highly sensitive method for transcriptome analysis that allows simultaneous assessment of expression of thousands of genes and identification of expression patterns under various conditions. The existing variety of RNA-seq data formats, normalization methods, and approaches to statistical processing of results complicates comparison of data from different studies and reduces reproducibility of the analysis. This study presents an automated pipeline PipeSeq that combines standard steps of RNA-seq data processing: loading (SRA Toolkit), read alignment to the reference genome (HISAT2), transcript assembly (StringTie), transcript counting (FeatureCounts) and statistical analysis of differential gene expression under various experimental conditions (DESeq2). PipeSeq has a simple visual interface, supports multithreading, and generates ready-to-analyze gene expression heat maps, tables and graphs. The functionality of the pipeline is demonstrated on three sets of raw RNA-seq data from the green alga *Chlamydomonas reinhardtii* cells available in the NCBI SRA database. The data from these experiments were used to

analyze the differential expression of *C. reinhardtii* genes encoding the GATA family transcription factors under different light cultivation conditions. The data obtained by *in silico* methods were verified by real-time reverse transcription polymerase chain reaction (RT-qPCR) for 12 GATA genes, which allowed us to hypothesize their functions and evaluate the correlation between the bulk (RNA-seq) and targeted (RT-qPCR) approaches. Our results showed that RNA-seq and RT-qPCR methods reveal similar directions of gene expression changes, but demonstrate differences in the effect size and sensitivity, which emphasizes the need for a combined use of the two approaches. Thus, the PipeSeq program is a tool for conducting a full cycle of bioinformatic analysis of RNA-seq data, additionally providing the opportunity to process RT-qPCR data and perform a comparative statistical analysis of the results obtained.
Key words: RNA sequencing; RNA-seq; RT-qPCR; pipeline; transcriptome; gene expression; GATA family transcription factors; GATA TFs; *Chlamydomonas reinhardtii*

For citation: Nerezenko A.M., Virolainen P.A., Tupitsyna S.A., Chekunova E.M. Development and validation of the PipeSeq program for RNA-seq data analysis in the *Chlamydomonas reinhardtii* as a model. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(2):299-310. doi 10.18699/vjgb-26-34

Введение

В последние годы метод секвенирования РНК (РНК-сек) получил широкое распространение как надежный подход для массового количественного анализа экспрессии генов в различных биологических системах (Marioni et al., 2008; Mortazavi et al., 2008; Conesa et al., 2016; Li X., Wang, 2021). Благодаря «цифровой» природе данных метод РНК-сек позволяет напрямую подсчитывать количество прочтений для каждого транскрипта, обеспечивая широкий динамический диапазон измерений и высокую воспроизводимость результатов эксперимента. Кроме того, этот метод дает возможность обнаруживать ранее не аннотированные транскрипты и варианты альтернативного сплайсинга (Wang et al., 2009; Li X., Wang, 2021). Эти особенности делают секвенирование РНК мощным инструментом для системных исследований транскриптома, позволяя надежно измерять уровни транскриптов в различных условиях эксперимента и выявлять дифференциальную экспрессию генов объекта исследования (Wang et al., 2009).

Обработка, накопление и объединение этих данных могут способствовать получению новых знаний о функционировании живых систем. Для реализации подобного подхода необходимо решить проблему стандартизации процесса обработки сырых данных РНК-сек (Conesa et al., 2016; Li X., Wang, 2021). Для оценки экспрессии генов до сих пор применяются устаревшие методы нормализации, которые характеризуются низкой воспроизводимостью, не позволяют корректно нормализовывать композиционные смещения и проводить прямое сравнение результатов между независимыми исследованиями, – это RPKM (reads per kilobase per million mapped reads), FPKM (fragments per kilobase of exon model per million mapped fragments) и TPM (transcripts per million). В настоящее время стандартом для проведения анализа дифференциальной экспрессии генов считаются методы, основанные на отрицательном биномиальном распределении, – DESeq2 и edgeR. Нормализованные счетчики (DESeq2) продемонстрировали наименьший коэффициент вариации и максимальную воспроизводимость (Zhao S. et al., 2020; Zhao Y. et al., 2021; Elahimanesh, Najafi, 2024).

Существующее разнообразие методов обработки данных РНК-сек, форматов и способов их представления в различных базах данных затрудняет консолидацию сведений для проведения комплексного анализа. Для решения этой проблемы требуется разработка интегрированных инструментов, способных минимизировать ручной труд, обеспечить высокую степень воспроизводимости результатов и простоту использования для биологов без специальной подготовки в области информационных технологий (Conesa et al., 2016).

Проведение сравнительного анализа результатов обработки данных РНК-сек и молекулярно-биологических данных осложнено рядом методологических проблем. Хотя секвенирование РНК признано достоверным методом глобального профилирования экспрессии, полученные с его помощью результаты обычно представляют относительные изменения уровня транскриптов в масштабе всего генома. Исторически сложилась практика применения метода полимеразной цепной реакции в реальном времени с обратной транскрипцией (ОТ-ПЦР-РВ) как «золотого стандарта», позволяющего подтвердить данные, полученные по результатам транскриптомных исследований (Derveaux et al., 2010; Coenye, 2021). Однако методы на основе ПЦР-РВ и технология секвенирования РНК основаны на разных протоколах, что затрудняет их прямое количественное сравнение: на получаемые результаты могут оказывать влияние эффективность процессов обратной транскрипции и амплификации, методы нормализации данных, чувствительность методик.

В настоящее время метод РНК-сек используется для охвата всего транскриптома объекта исследования и выявления круга дифференциально экспрессируемых генов-кандидатов в ответ на определенное воздействие, тогда как метод ОТ-ПЦР-РВ применяется для точной количественной оценки этих изменений прицельно для ограниченного пула генов интереса (Shi, He, 2014; He et al., 2015; Coenye, 2021).

Для понимания механизмов регуляции метаболизма особое внимание уделяется изучению факторов транскрипции (ФТ) как ключевых регуляторов активности генов. ФТ представляют собой белки, способные связываться с определенными последовательностями ДНК в промоторных областях и тем самым усиливать или подавлять транскрипцию целевых генов. Эти регуляторы координируют работу генов в ответ на различные изменения среды и сигнальные воздействия, участвуя в глобальных процессах роста, развития и адаптации к стрессовым факторам (Riechmann et al., 2000).

У фотосинтезирующих организмов особый интерес представляют гены, кодирующие ФТ семейства GATA, – белки, несущие консервативный домен цинкового пальца типа IV (общая формула: $CX_2CX_{18–20}CX_2C$). Этот домен осуществляет связывание консенсусной последовательности (A/T)GATA(A/G) в промоторах целевых генов (Reyes et al., 2004). Факторы GATA растений участвуют в регуляции процессов фотоморфогенеза, метаболизма азота и углерода, гормонального контроля (Manfield et al., 2007; Naito et al., 2007; Luo et al., 2010; Schwechheimer et al., 2022; Schröder et al., 2023; Ren et al., 2025). В последние годы наблюдается существенный рост научного интереса к исследованию ФТ GATA у мохообразных и водорослей, поскольку их биологические функции и эволюционная роль остаются недостаточно изученными (Schwechheimer et al., 2022; Virolainen, Chekunova, 2024).

Целью настоящего исследования стала разработка интегрированного решения для унификации процессов обработки сырых данных секвенирования РНК и результатов ОТ-ПЦР-РВ.

Для реализации этой цели необходимо:

1. Разработать автоматизированный алгоритм (пайплайн) для проведения полного цикла анализа данных РНК-сек.
2. Апробировать пайплайн на сырых данных РНК-сек, доступных в открытой базе NCBI SRA, на модели генов, кодирующих ФТ семейства GATA у модельного объекта генетики фотосинтеза – зеленой водоросли *Chlamydomonas reinhardtii*, в ответ на изменение световых условий выращивания.
3. Провести анализ экспрессии генов GATA *C. reinhardtii* методом ОТ-ПЦР-РВ в ответ на изменение световых условий выращивания.
4. Провести сравнительный статистический анализ полученных данных РНК-сек и ОТ-ПЦР-РВ.

Материалы и методы

Количественный анализ экспрессии генов в программе PipeSeq. Для анализа экспрессионного профиля 12 генов, кодирующих ФТ GATA у *C. reinhardtii* в различных условиях культивирования методами *in silico*, были отобраны следующие доступные пакеты сырых данных РНК-сек из базы данных NCBI SRA (Wheeler et al., 2005): SRX8380269, SRX8380270, SRX8380271 (повышенная освещенность – 600 мкмоль/м^2/с, 1 ч), SRX7413406, SRX7413407, SRX7413412, SRX7413413, SRX7413414, SRX7413415 (акклиматизация к свету, 30 мин), SRX5120530, SRX5120531, SRX5120532, SRX5120533, SRX5120534, SRX5120535 (акклиматизация к темноте).

Выбранные три набора, PRJNA634446, PRJNA596622, PRJNA509798, содержат данные эксперимента (выращивание в измененных условиях) и данные контроля (выращивание в стандартных условиях) для штамма *C. reinhardtii* дикого типа CC-124 (*wt*, *mt*-), выполненные в трехкратной биологической и технической повторности. Для выравнивания прочтений с портала Phytozome (Goodstein et al., 2012) были загружены геном *C. reinhardtii* (v5.6) (Merchant et al., 2007) и его аннотация в формате GTF.

Загрузку и обработку данных осуществляли с помощью PipeSeq – автоматизированного программного комплекса из 17 скриптов и более 2000 строк кода, предназначенного для системного анализа транскриптомных данных, полученных методом РНК-сек, а также обработки данных ПЦР-РВ методом ΔΔCt (Livak, Schmittgen, 2001; Schmittgen, Livak, 2008) и проведения сравнительного анализа результатов (с расчетом коэффициентов Пирсона, Спирмена, Кендалла). Программа написана на языке программирования Python (версия 3.9) с использованием библиотек Pandas (McKinney, 2011), Matplotlib (Hunter, 2007), PyQt6 (https://www.riverbankcomputing.com/software/pyqt/), PyDESeq2 (Muzellec et al., 2023) и инструментов биоинформатики SRA Toolkit, FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), MultiQC (Ewels et al., 2016), Cutadapt (Martin, 2011), HISAT2 (Kim et al., 2019), SAMtools (Li H. et al., 2009), FeatureCounts (Liao et al., 2014), StringTie (Pertea et al., 2015), DESeq2 (Love et al., 2014).

Штаммы и условия культивирования. Штамм дикого типа *C. reinhardtii* CC-124 (*wt*, *mt*-) из Петергофской генетической коллекции СПбГУ (Квитко и др., 1983) выращивали в чашках Петри на агаризованной среде TAP (1.5 %) (Harris, 1989) с добавлением аргинина (50 мг на литр среды) и дрожжевого автолизата (4 г на литр среды) при температуре 20–25 °C в цикле день (14 ч, освещенность 90 мкмоль/м^2/с) – ночь (10 ч) с пересевом каждые три дня. Для работы отбирали культуры в стандартных условиях освещенности (90 мкмоль/м^2/с) и темноты (контрольные условия), в условиях повышенной освещенности (215 мкмоль/м^2/с) в течение 30 мин и 2 ч (перенос культур из стандартных условий), в условиях стандартной освещенности в течение 30 мин и 2 ч (перенос культур из темноты), в условиях темноты в течение 30 мин и 2 ч (перенос культур из стандартных условий) (экспериментальные условия).

Фиксация культур и выделение РНК. Фиксацию выращенных в контрольных и экспериментальных условиях культур клеток *C. reinhardtii* и выделение суммарной РНК проводили с использованием реагента “ExtractRNA” («Евроген», Россия) в строгом соответствии с инструкцией производителя. Препараты РНК очищали от примеси геномной ДНК с помощью фермента ДНКаза I (Thermo Fisher Scientific, США) и переосаждали этанолом. Концентрацию суммарной РНК измеряли на спектрофотометре Eppendorf BioPhotometer plus (Eppendorf, Германия).

Дизайн праймеров. Ген-специфичные праймеры подбирали с помощью программ IDT PrimerQuest Tool (https://www.idtdna.com/pages/tools/primerquest) и NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) таким образом, чтобы место посадки одного из праймеров располагалось на стыке экзонов либо между сайтами отжига праймеров находилась последовательность интрона. В качестве референс-генов для нормализации получаемых значений задействовали конститутивно экспрессируемые гены *RPL19* (*ribosomal protein L19*) и *RPL32* (*ribosomal protein L32*) *C. reinhardtii* (Liu et al., 2012); для контроля

Последовательности праймеров для количественного анализа экспрессии генов GATA

Ген	Идентификатор	Праймер	Последовательность (5′–3′)	Литературный источник
GATA-1	Cre01.g025050	F	GTTGTTTGGCGACCTCTTTGTG	Настоящее исследование
		R	GATCAGCGGCGGCTATGTC	
GATA-2	Cre10.g435450	F	ACTACGACGAGCGGGAAGA	
		R	GCCTTCTTCGCCATGTACTCC	
GATA-3	Cre08.g378800	F	GAGCTGGACGGGAACGAAAC	
		R	GTGCGGTGCCGAGTAGTTT	
GATA-4	Cre11.g467581	F	GATCCTATCACCACCAAGGTTGC	
		R	CCATGCCGCCCATGTTCA	
GATA-5	Cre03.g160600	F	TCACGGGACGACGACATCA	
		R	CGGGTGAAGAATATGCCACAGG	
GATA-6	Cre03.g160700	F	GAAAAGGCAGGACAAGTCCAAG	
		R	TGTGAGGCGGGATGAAGAT	
GATA-7	Cre05.g242600	F	AGGAGCAGCAGCAGCAATC	
		R	CTGGTTAGTGCGGCGGTATC	
GATA-8	Cre06.g266850	F	TGTGCAACGCATGTGGGATA	
		R	CGGTCTTGGCTGACACATAGTT	
GATA-9	Cre06.g266950	F	ACATCAGCGGCTGCGATAAT	
		R	CGCCTGAGCCACTTTCGG	
GATA-10	Cre07.g319701	F	TCCGCTGCTGCGTAGAGT	
		R	GCAAAGACATCCTCGTCGGC	
GATA-11	Cre08.g358532	F	TCAGCAACAGCCCTCACTTC	
		R	CGCTCAAACCACTTGACCTCTAT	
GATA-12	Cre08.g358534	F	TGTCAAGTGTTTCCACGACAAGA	
		R	GCACCAGAACCACTCGCA	
RPL32	Cre06.g289550	F	CCCAACGGCTTCCTGAAGTA	
		R	AAGCGACGGTTGTGCATCAT	
RPL19	Cre02.g075700	F	CCTGAAGAAGTACCGCGACTC	Liu et al., 2012
		R	AACACGTTACCCTTGACCTTCA	
RBCS	Cre02.g120150	F	ACCCCGGTCAACAACAAGATG	Sanchez-Tarre, Kiparissides, 2021
		R	GTCGTAGTACAGGCAAGACACG	

изменения условий эксперимента использовали ген *RBCS* (*Ribulose bisphosphate carboxylase small subunit*) *C. reinhardtii* (Sanchez-Tarre, Kiparissides, 2021). Последовательности праймеров приведены в таблице.

Для оценки специфичности использовали анализ кривых плавления праймеров на серии контролей (контроль без матрицы, контроль без обратной транскрипции, положительный контроль), а также визуализацию ПЦР-продуктов методом гель-электрофореза (Derveaux et al., 2010). Эффективность ПЦР оценивали средствами программного обеспечения амплификатора QuantStudio 5 (Thermo Fisher Scientific).

Количественный анализ экспрессии генов методом ОТ-ПЦР-РВ. Реакции ПЦР-РВ проводили в один шаг с использованием компонентов набора реагентов “OneTube RT-PCR SYBR” («Евроген», Россия), строго следуя инструкции производителя, на амплификаторе QuantStudio 5 (Thermo Fisher Scientific) со считыванием флюоресценции на шаге элонгации и плавления по следующему протоколу: 55 °C – 15 мин, 95 °C – 1 мин, далее 50 циклов: 95 °C – 15 с, 62 °C – 20 с, 72 °C – 20 с, плавление: 55–95 °C с шагом 0.5 °C. Каждый образец анализировали в трехкратной усредненной биологической повторности (Derveaux et al., 2010). Обработку и визуализацию данных осуществляли методом ΔΔCt (Livak, Schmittgen, 2001; Schmittgen, Livak, 2008) в разработанной программе PipeSeq.

Сравнительный анализ результатов РНК-сек и ОТ-ПЦР-РВ. Сравнительный статистический анализ полученных разными методами данных выполняли в программе PipeSeq.

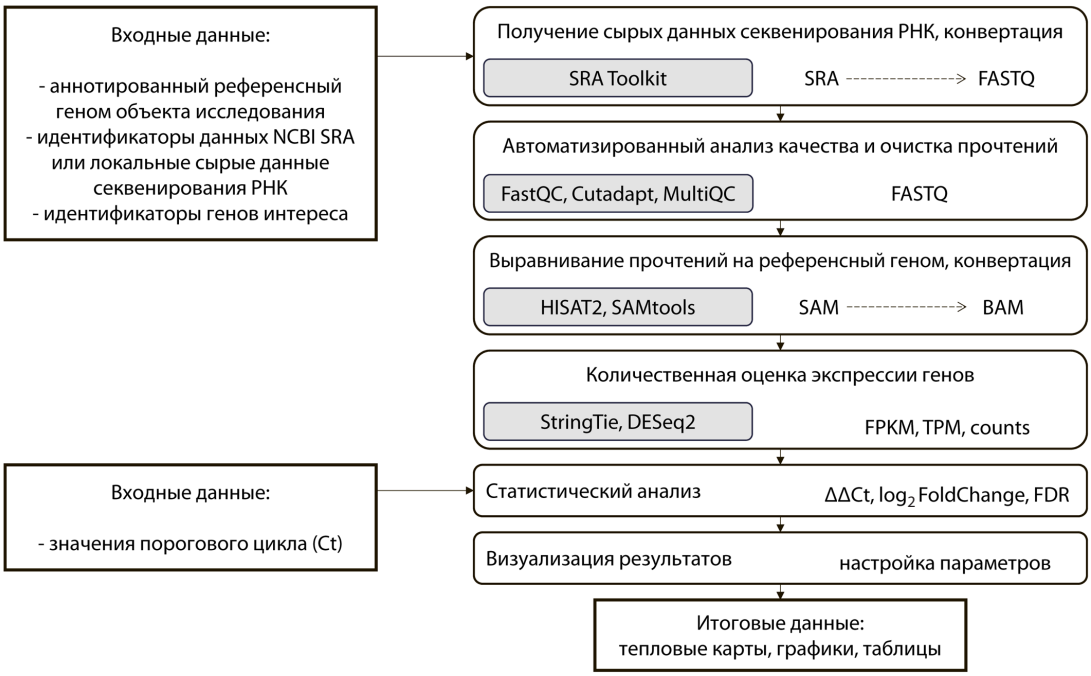

Рис. 1. Схема работы программы PipeSeq.

Результаты

Разработка пайплайна PipeSeq

Целью работы было создание такого инструмента (пайплайна), который позволил бы проводить полный цикл анализа данных РНК-сек через простой визуальный интерфейс с минимальной потребностью для пользователя в системном администрировании и был ориентирован на запуск в локальной среде Windows с использованием подсистемы WSL для выполнения команд Linux. Схема работы программы PipeSeq представлена на рис. 1.

Все шаги исполнения алгоритма полностью автоматизированы и позволяют анализировать большие объемы данных за счет оптимизации команд и параллельных вычислений. Входными данными являются файлы формата FASTQ, содержащие необработанные короткие прочтения секвенирования, получаемые после конвертации из формата SRA с помощью входящего в пакет программы инструмента SRA Toolkit. На начальном этапе выполняется автоматизированная подготовка прочтений, включающая контроль качества (FastQC), удаление последовательностей адаптеров, нуклеотидов с низким качеством, коротких прочтений (Cutadapt), агрегирование сводных отчетов (MultiQC). Далее проводится выравнивание прочтений на референсный геном с помощью инструмента HISAT2 (Kim et al., 2019), при этом автоматически создается индекс референсного генома в случае его отсутствия. Преимущества HISAT2 – учет сплайсинга, точное определение структуры транскриптов и быстрый анализ качества выравнивания. После выравнивания осуществляются конвертация файлов из формата SAM в формат BAM, сортировка и подготовка к дальнейшему анализу с использованием инструментов SAMtools (Li H. et al., 2009).

Разработанная программа PipeSeq интегрирует три подхода нормализации: DESeq2 (для анализа дифференциальной экспрессии), TPM (абсолютная квантификация) и FPKM (устаревший показатель, включен для обеспечения обратной совместимости). На данном этапе проводятся количественная оценка экспрессии генов методом подсчета числа прочтений, приходящихся на гены и транскрипты, с использованием пакета FeatureCounts (Liao et al., 2014), а также сборка транскриптов и расчет значений FPKM и TPM с помощью инструмента StringTie (Pertea et al., 2015), позволяющего анализировать выравнивания ридов, строить транскрипты и определять экзоны, интроны и сплайс-сайты. В программе PipeSeq реализована возможность отключать функцию построения новых транскриптов для минимизации вероятности получения ложных результатов. Данные этапы выполняются с учетом различных режимов, таких как строгая аннотация (опция «-e» в StringTie) и изменяемая чувствительность (параметр «-c»).

Для статистического анализа дифференциальной экспрессии генов используется пакет на основе отрицательного биномиального распределения DESeq2 (Love et al., 2014), реализованный в библиотеке PyDESeq2 (Muzellec et al., 2023). Пайплайн PipeSeq автоматически готовит входные данные для DESeq2, осуществляет статистический расчет и формирует итоговые таблицы с логарифмическими изменениями уровня экспрессии ($\log_2$FoldChange) и соответствующими значениями скорректированного p-значения – FDR (false discovery rate) по методу Бенджамини–Хохберга. Статистические гипотезы о дифференциальной экспрессии генов проверяются на основе обобщенной линейной модели отрицательного биномиального распределения. После нормализации для счета данных и

оценки дисперсии (с последующим усреднением оценок дисперсии по всем генам с помощью байесовского сглаживания) применяется тест Уолда.

В программе также реализована возможность обработки данных ПЦР-РВ методом ΔΔCt (Livak, Schmittgen, 2001; Schmittgen, Livak, 2008) и проведения сравнительного статистического анализа с тремя используемыми методами нормализации данных РНК-сек (коэффициент линейной корреляции Пирсона, коэффициент ранговой корреляции Спирмена, коэффициент ранговой корреляции Кендалла).

На выходе программа PipeSeq формирует тепловые карты экспрессии, отражающие изменения в уровне экспрессии генов интереса при различных экспериментальных условиях, итоговые таблицы со значениями $\log_2$FoldChange и графики. Пользователю предоставлен широкий выбор настроек отображения данных. Программа доступна для скачивания по ссылке: https://github.com/MarvinMarss/PipeSeq.

Обработка доступных данных секвенирования РНК клеток C. *reinhardtii* с помощью пайплайна PipeSeq

С использованием разработанной программы мы проанализировали три доступных набора данных РНК-сек клеток *C. reinhardtii* штамма дикого типа CC-124 (*wt*, *mt-*) при изменении условий освещения – повышенная освещенность (600 мкмоль/м^2/с) в течение 1 ч, акклиматизация к свету в течение 30 мин, акклиматизация к темноте. Построенная тепловая карта демонстрирует сложную динамику экспрессии генов, кодирующих ФТ GATA, в ответ на изменение световых условий (рис. 2). Большинство полученных значений логарифма изменения уровня экспрессии генов оказались статистически незначимыми.

В условиях повышенной освещенности наблюдается значимое повышение уровня экспрессии генов *GATA-7*, *GATA-9*, *GATA-10*, *GATA-11*, в то время как транскрипция генов *GATA-3*, *GATA-5*, *GATA-8* значимо репрессируется. Вероятно, излишне высокая интенсивность освещения оказывает негативное влияние на жизнедеятельность клеток *C. reinhardtii*.

В ходе акклиматизации к свету в течение 30 мин происходит перенастройка метаболизма клеток, что маркирует значимое повышение уровня транскриптов генов *GATA-3*, *GATA-5*, *GATA-6*, *GATA-7*, *GATA-8*. Экспрессия генов *GATA-2*, *GATA-4*, *GATA-10* в этих условиях достоверно репрессируется.

В темноте достоверно происходят подавление экспрессии генов *GATA-2*, *GATA-3*, *GATA-7*, *GATA-8*, *GATA-11*, *GATA-12* и активная экспрессия гена *GATA-9*.

Доступные наборы данных характеризуются ограниченностью условий эксперимента и скудной охарактеризованностью, а также низким уровнем достоверности изменений. Мы приняли решение воспользоваться методом ОТ-ПЦР-РВ для получения полных данных по экспрессии генов GATA. Анализ данных РНК-сек позволил нам осуществить выбор референс-генов со стабильной экспрессией при изменении условий освещения для проведения собственных экспериментов – были обнаружены и затем подтверждены литературными данными (Liu et al., 2012) гены *RPL19* и *RPL32*, кодирующие рибосомные белки. Экспрессия гена *RBCS*, согласно литературным данным (Sanchez-Tarre, Kiparissides, 2021), изменяется в зависимости от спектра и интенсивности освещения, в связи с чем этот ген выбран нами в качестве контроля изменения условий эксперимента.

Анализ экспрессии генов GATA *C. reinhardtii* методом ОТ-ПЦР-РВ

Полученные методом ОТ-ПЦР-РВ результаты анализа экспрессии генов GATA *C. reinhardtii* представлены на рис. 3.

Вся группа генов GATA отвечает на изменение условий освещения быстро и согласованно. Самый сильный ответ наблюдается в первые 30 мин повышенной освещенности (215 мкмоль/м^2/с) после переноса культуры из стандартных условий (90 мкмоль/м^2/с): почти все гены демонстрируют кратное увеличение уровня экспрессии в ответ на стресс-воздействие. Примечательно, что только в этих условиях происходит значительная активация экспрессии гена *GATA-6*. Повышенная освещенность интенсифицирует метаболизм клеток *C. reinhardtii*, обеспечивая единое направление изменения профиля экспрессии изучаемых генов. Через 2 ч наблюдается стабилизация – экспрессия ряда генов (*GATA-1*, *GATA-6*, *GATA-8*, *GATA-9*, *GATA-10*, *GATA-11*, *GATA-12*) значительно снижается, а генов *GATA-5* и *GATA-7* – усиливается.

В первые 30 мин после переноса культуры из темноты на свет (90 мкмоль/м^2/с) происходит перенастройка метаболизма – наблюдается репрессия транскрипции генов *GATA-1* и *GATA-3* одновременно с активацией генов *GATA- 2*, *GATA-4*, *GATA-5*, *GATA-7*, *GATA-9* и *GATA-10*. К 2 ч экспозиции формируется два профиля, обеспечивающих поддержание процессов роста и развития в оптимальных световых условиях: 1) активно транскрибируемые гены (*GATA-1*, *GATA-4*, *GATA-5*, *GATA-8*, *GATA-9*, *GATA-11*, *GATA-12*); 2) репрессированные гены (*GATA-2*, *GATA-3*, *GATA-6*, *GATA-7*, *GATA-10*).

В первые 30 мин после переноса культуры из света (90 мкмоль/м^2/с) в темноту происходит подавление экспрессии всех генов GATA, за исключением гена *GATA-7*. По-видимому, его продукт участвует в процессах переключения метаболизма клеток *C. reinhardtii* при переходе свет/темнота и темнота/свет. Через 2 ч происходит стабилизация метаболизма клеток с выявлением трех активно транскрибируемых генов, по-видимому, задействованных в обеспечении адаптации к темноте, – *GATA-1*, *GATA-3*, *GATA-9*.

Выявленная динамика экспрессии генов GATA согласуется со стандартной моделью ответа на возникающий стресс: изменение световых условий запускает широкий аварийный каскад (быстрый ответ на раздражитель); длительная экспозиция сужает реакцию до специфических регуляторных модулей. В каждых условиях культивирования наблюдается уникальный профиль экспрессии,

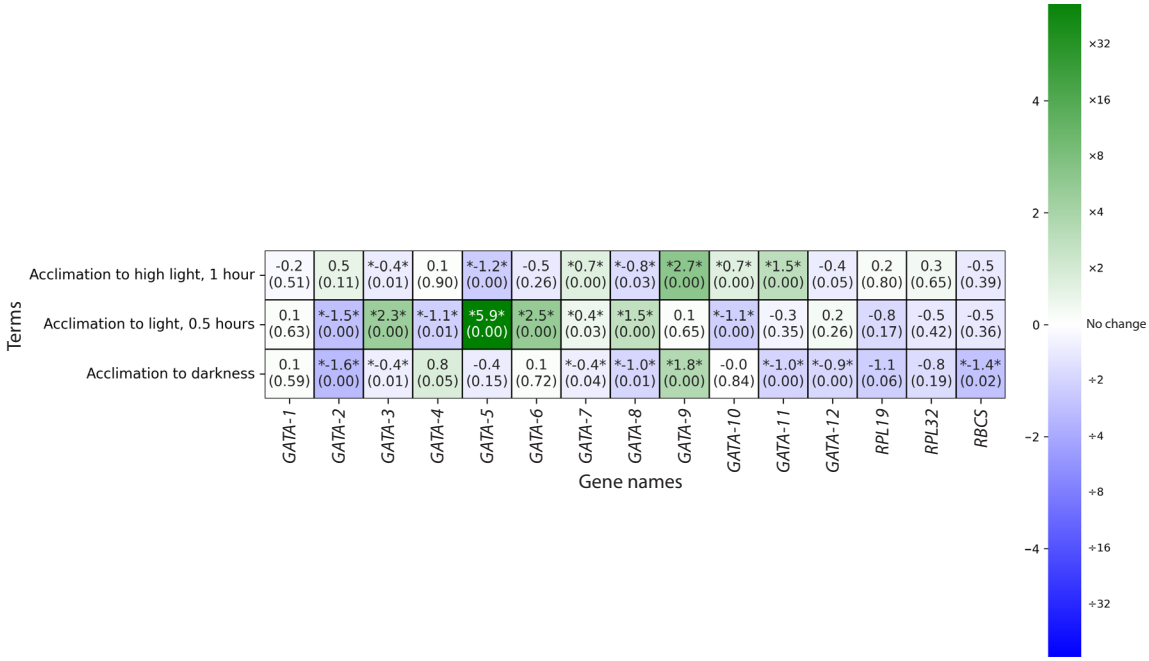

Рис. 2. Тепловая карта экспрессии (на уровне транскрипции) 12 генов, кодирующих факторы транскрипции семейства GATA и трех контрольных генов (*RPL19*, *RPL32*, *RBCS*) *C. reinhardtii* при различных условиях акклиматизации по данным секвенирования РНК.

Слева направо – гены; сверху вниз – экспериментальные условия: акклиматизация к высокому освещению в течение 1 ч (ряд 1), акклиматизация к свету в течение 30 мин (ряд 2), акклиматизация к темноте (ряд 3).
Здесь и на рис. 3: ячейки окрашены в соответствии со значениями $\log_2$FoldChange, где положительные значения (зеленая шкала) отражают индукцию экспрессии, отрицательные (синяя шкала) – репрессию, а нулевые значения (белый цвет) – сохранение уровня экспрессии. Статистическая значимость изменений оценивалась по скорректированным p-значениям с порогом 0.05. Контрольные транскрипты включены для верификации качества нормализации данных. Изображение создано в программе PipeSeq.

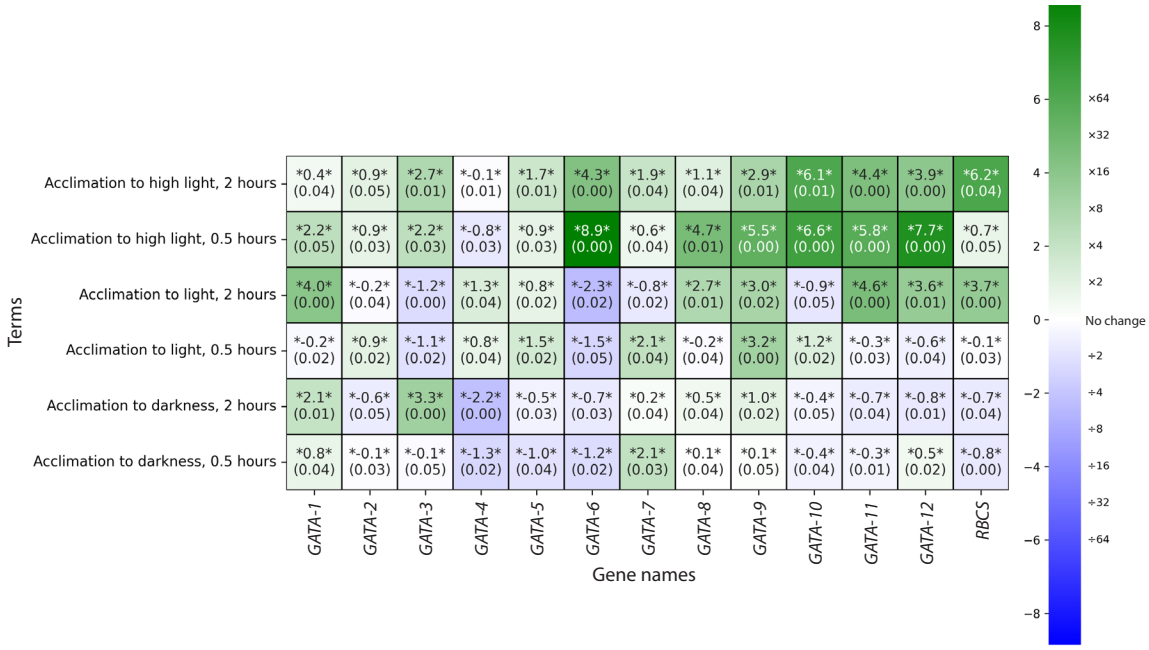

Рис. 3. Тепловая карта экспрессии (на уровне транскрипции) 12 генов, кодирующих факторы транскрипции семейства GATA и гена-контроля изменения условий (*RBCS*) *C. reinhardtii* при различных условиях акклиматизации по данным ОТ-ПЦР-РВ.

Данные нормированы на пару референс-генов *RPL19–RPL32*. Слева направо – гены; сверху вниз – экспериментальные условия: акклиматизация к высокому освещению в течение 2 ч (ряд 1), в течение 30 мин (ряд 2), акклиматизация к свету в течение 2 ч (ряд 3), в течение 30 мин (ряд 4), акклиматизация к темноте в течение 2 ч (ряд 5), в течение 30 мин (ряд 6).

причем часть генов GATA предположительно действует как «переключатель» между световой и темновой программами метаболизма.

Сравнительный анализ полученных результатов

Возможность проведения сравнительного анализа по результатам обработки данных РНК-сек с применением трех методов нормализации, FPKM, TPM, DESeq2, и обработки результатов ОТ-ПЦР-РВ методом ΔΔCt с построением корреляционной матрицы в программе PipeSeq позволяет оценить степень согласованности полученных разными методами и подходами данных (рис. 4).

Межплатформенная согласованность полученных результатов ограничена и зависит от контекста. Несмотря на методологические различия, воспроизводимые совпадения включают индукцию экспрессии генов *GATA-7*, *GATA-9*, *GATA-10* и *GATA-11* при повышенном освещении, репрессию генов *GATA-2*, *GATA-11*, *GATA-12* в темноте и усиление экспрессии гена *GATA-9* в темноте. Ранняя репрессия генов *GATA-3*, *GATA-5*, *GATA-8* при повышенном освещении через 1 ч, по данным РНК-сек, оказалась инвертированной, по данным ОТ-ПЦР-РВ, через 2 ч, что согласуется с потенциальным сдвигом фазы реакции, но для подтверждения этого требуется выборка, согласованная по времени. В темноте оба набора данных подтверждали общую репрессию связанных с фотосинтезом транскриптов генов *RBCS*, *GATA-2*, *GATA-11*, *GATA-12* и повышением *GATA-9*, что указывает на скоординированную программу адаптации к темноте, полученную обоими методами.

Корреляция между методами продемонстрировала сильную внутреннюю согласованность между показателями, полученными на основе данных РНК-сек (DESeq2, FPKM, TPM), и более слабую согласованность с ΔΔCt, что, вероятно, отражает различия во времени отбора проб, поведении референс-генов и разрешающей способности методик. Наибольшая корреляция наблюдается в паре DESeq2–ΔΔCt (см. рис. 4). Результаты сравнительного анализа показывают, что, хотя обе платформы фиксируют перекрывающиеся регуляторные тренды для некоторых генов GATA, каждая из них также обнаруживает специфические для условий и времени изменения в экспрессии. Это подчеркивает ценность интеграции обоих подходов.

Обсуждение

Преимущества разработанной программы PipeSeq

Программа PipeSeq разрабатывалась с учетом минимальной потребности в системном администрировании и ориентирована на запуск в локальной среде Windows. Она предоставляет удобный графический интерфейс, который снижает порог вхождения пользователя, объединяет современные методы анализа данных и подходит для работы с малыми и средними наборами данных на обычном персональном компьютере, что делает ее удобным и практичным решением для индивидуальных исследователей и небольших лабораторий, не имеющих значительных вычислительных мощностей и опыта системного администрирования.

На сегодняшний день существует ряд инструментов и платформ для транскриптомного анализа, таких как Galaxy (Afgan et al., 2018), Nextflow (Di Tommaso et al., 2017), Snakemake (Mölder et al., 2021), а также специализированные связки типа HISAT2-StringTie-Ballgown (Pertea et al., 2016) и Kallisto-Sleuth (Bray et al., 2016). Galaxy представляет собой удобную веб-платформу с графическим интерфейсом, однако требует отдельного администрирования сервера и не всегда подходит для индивидуальных исследователей или небольших лабораторий без ИТ-поддержки (Afgan et al., 2018).

Программные платформы Nextflow и Snakemake предлагают высокую гибкость и масштабируемость за счет возможности параллельного запуска задач и поддержки контейнеризации, обеспечивающей полную воспроизводимость анализа, однако использование данных систем требует от исследователей навыков программирования и администрирования среды Linux, что ограничивает их применение биологами без опыта работы с консолью и скриптами (Di Tommaso et al., 2017; Mölder et al., 2021).

Важным аспектом является поддержка анализа дифференциальной экспрессии генов: современные подходы рекомендуют использование пакетов на основе отрицательного биномиального распределения (например, DESeq2), обеспечивающих высокую точность и контроль ложных открытий (Zhao S. et al., 2020; Zhao Y. et al., 2021; Elahimanesh, Najafi, 2024).

Разработанный нами программный пайплайн PipeSeq (см. рис. 1) обеспечивает использование одних из самых быстрых инструментов для выравнивания (HISAT2) и сборки транскриптов (StringTie), которые значительно превосходят предшествующие методы (TopHat и Cufflinks) по скорости и требовательности к вычислительным ресурсам (Kim et al., 2015, 2019), а интеграция DESeq2 позволяет автоматически получать статистически обоснованные результаты дифференциальной экспрессии с контролем FDR (Love et al., 2014). Это ставит его в один ряд с признанными решениями в области анализа данных РНК-сек, такими как связка HISAT2-StringTie-Ballgown, где применяется аналогичный статистический подход, хотя и менее интегрированный (Pertea et al., 2016). Также преимуществом разработанной программы являются возможность обсчета и визуализации данных ПЦР-РВ методом ΔΔCt и проведение сравнительного статистического анализа (с расчетом коэффициента Пирсона, Спирмена, Кендалла) с тремя методами нормализации данных РНК-сек – FPKM, TPM, DESeq2.

Таким образом, разработанная программа PipeSeq обладает следующими особенностями:

1. Взаимодействие пользователя с программой происходит исключительно через графический интерфейс (пять окон).
2. Все этапы обработки данных объединены в рамках одной программы (не требуется подключение дополнительных пакетов), не требующей подключения к сети Интернет.
3. В программе реализован автоматизированный подбор параметров очистки прочтений (удаление адаптеров, по-

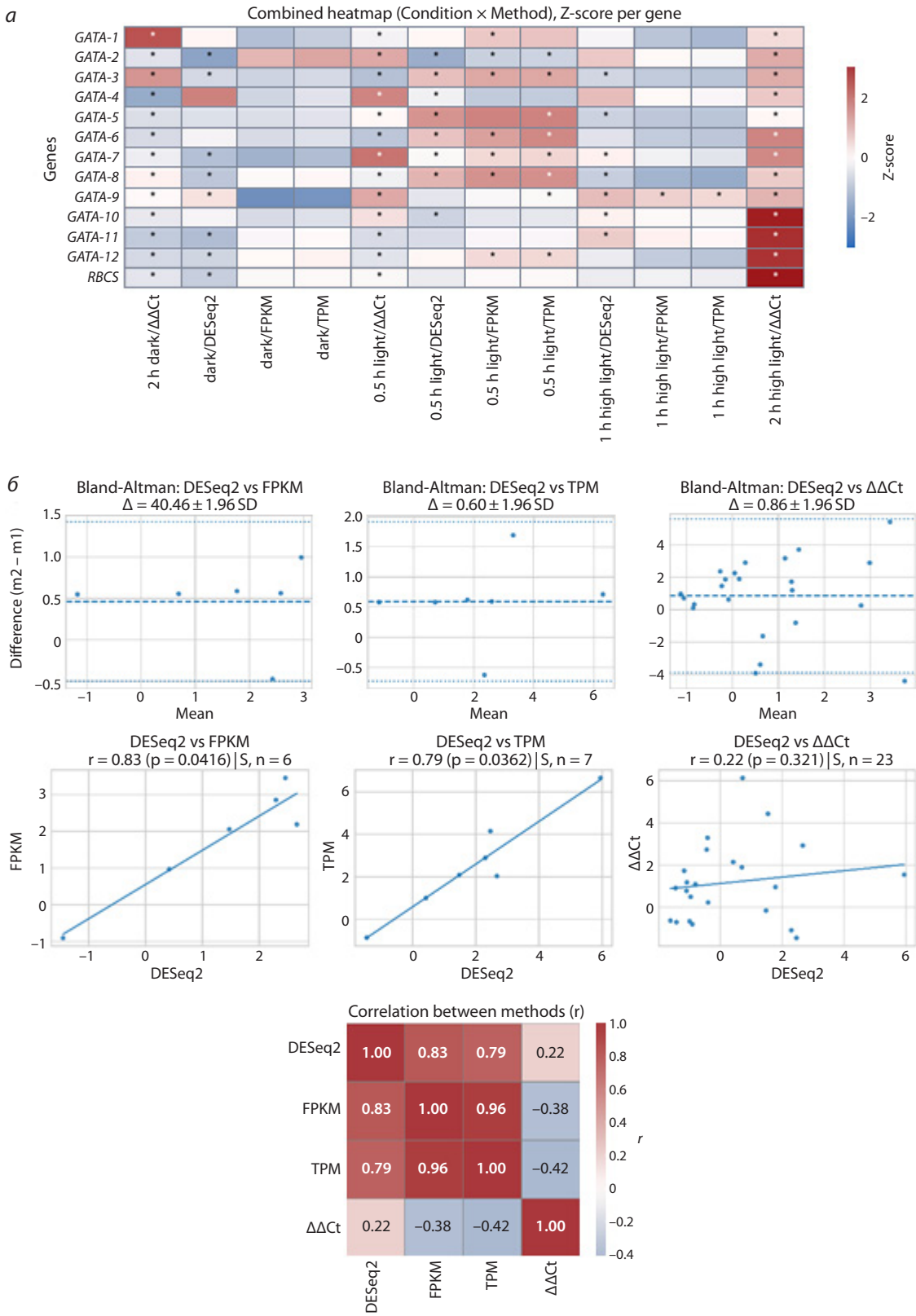

Рис. 4. Сравнительный статистический анализ применения трех методов нормализации данных секвенирования РНК, FPKM, TPM, DESeq2, и метода ΔΔCt данных ОТ-ПЦР-РВ.

а – совмещенная тепловая карта Z-оценок значений логарифма изменения уровня экспрессии (log_2FoldChange) 12 генов, кодирующих факторы транскрипции семейства GATA и гена *RBCS C. reinhardtii* для каждого сочетания условий и метода. Звездочкой отмечены статистически значимые изменения (FDR < 0.05); *б* – диаграммы Бланда–Альтмана, корреляционные диаграммы и матрица статистически достоверных (FDR < 0.05) оценок log_2FoldChange между сравниваемыми методами нормализации данных. Цветовая индикация отражает значение коэффициента Пирсона (r), где красная шкала – положительная корреляция, а синяя – отрицательная. Изображения созданы в программе PipeSeq.

рог качества, минимальная длина прочтения) на основе метрик качества до/после проведения очистки.

4. Встроенный модуль обсчета данных ПЦР-РВ методом ΔΔCt (импорт значений порогового цикла Ct, расчет нормализованной экспрессии, сравнительный статистический анализ по результатам обработки данных РНК-сек и результатов ПЦР-РВ).

Профили экспрессии генов GATA *C. reinhardtii*

Свет – один из основных регуляторов экспрессии генов фотосинтезирующих организмов. По результатам проведенного комплексного исследования мы получили наиболее полные профили экспрессии генов GATA *C. reinhardtii*, дополнив доступные данные РНК-сек собственными экспериментами методом ОТ-ПЦР-РВ. Несмотря на то что в настоящее время результаты транскриптомных исследований в большинстве случаев не требуют дополнительной верификации (Coenye, 2021), мы при поиске интересующих нас данных в открытых базах столкнулись со скудным их количеством (три пакета) и кратким описанием условий выращивания отобранных для работы культур.

Результирующие данные РНК-сек (см. рис. 2) для части генов GATA характеризовались низкой статистической значимостью изменений (FDR > 0.05). Обработка выбранных пакетов данных секвенирования РНК с помощью программы PipeSeq позволила нам обнаружить и подтвердить пару конститутивно экспрессируемых референс-генов *RPL19–RPL32* (Liu et al., 2012), а также профили экспрессии генов интереса в ответ на определенное воздействие с целью дальнейшего применения метода ОТ-ПЦР-РВ для точной количественной оценки этих изменений в расширенном перечне экспериментальных условий.

Полученные нами данные (см. рис. 3) подтверждают, что изменение световых условий является важным фактором, модулирующим экспрессию генов GATA у фотосинтезирующих организмов (Manfield et al., 2007; Naito et al., 2007; Luo et al., 2010; Schröder et al., 2023). Выполненный ранее анализ сети белковых взаимодействий (Virolainen, Chekunova, 2024) выявил три функциональных кластера, в которых предположительно принимают участие 12 факторов GATA *C. reinhardtii*.

Первый функциональный кластер, объединяющий гены *GATA-1*, *GATA-2*, *GATA-10*, связывает фоторецепцию (гены *CHLAMYDOMONAS PHOTOLYASE HOMOLOG 1* (*CPH1*), *SHOC2/SUR8-like LRR* (*CSL*)), циркадную регуляцию (гены семейства *RHYTHM OF CHLOROPLAST* (*ROC*) и метаболизм фосфора (ген *PHOSPHATE STARVATION RESPONSE 1* (*PSR1*)), обеспечивая адаптацию клетки к световому режиму.

Второй кластер (*GATA-9*) связан с функционированием арил-гидрокарбонового рецепторного комплекса и активируется при избытке света, поддерживая детоксикационные процессы в клетке.

Третий кластер (*GATA-3*, *GATA-4*, *GATA-5*, *GATA-6*, *GATA-7*, *GATA-8*, *GATA-10*, *GATA-11*, *GATA-12*) координирует ассимиляцию азота (через связь с геном *NITRATE REDUCTASE* (*NIT1*), ремоделирование хроматина (с генами, кодирующими гистондеацетилазы), репликацию ДНК (с геном, кодирующим хеликазу (*CrRuvBL1*), мембранный транспорт и деление клетки (с генами, кодирующими белок семейства Rab (*RABF1*), кинезин-подобный белок, субъединицу 4 циклосомы), проявляя выраженную светозависимость.

Наблюдаемая динамика экспрессии генов GATA представляет собой скоординированный ответ, интегрирующий ключевые процессы метаболизма клетки в ходе клеточного цикла (Voigt, Münzner, 1987; Müller et al., 2017; Salomé, Merchant, 2019). Ответ всех изученных генов GATA на световой стимул позволяет предположить наличие механизма перекрестной регуляции между тремя функциональными сетями через еще не выявленные или не охарактеризованные гены и белки.

Результаты проведенных исследований позволяют сделать первые значимые предположения относительно функций ФТ GATA у *C. reinhardtii* и зеленых водорослей в целом (Schwechheimer et al., 2022), закладывая необходимую базу для проведения дальнейших исследований.

Сравнительный анализ применения трех методов нормализации данных РНК-сек, FPKM, TPM, DESeq2, и метода ΔΔCt, по данным ОТ-ПЦР-РВ, продемонстрировал наибольшую корреляцию в паре DESeq2–ΔΔCt (см. рис. 4), что подтверждает литературные данные о высокой точности пакетов на основе отрицательного биномиального распределения (Zhao S. et al., 2020; Zhao Y. et al., 2021; Elahimanesh, Najafi, 2024) и демонстрирует надежность метода нормализации ΔΔCt данных ОТ-ПЦР-РВ для точной количественной оценки уровня экспрессии генов интереса (Livak, Schmittgen, 2001; Schmittgen, Livak, 2008; Shi, He, 2014; He et al., 2015; Coenye, 2021; Schröder et al., 2023). В совокупности оба подхода отражают совпадающие тенденции для небольшого количества генов GATA и различия, зависящие от условий, метода и времени, что подчеркивает ценность интеграции массового (РНК-сек) и таргетного (ОТ-ПЦР-РВ) подходов.

Заключение

Разработан и использован программный пайплайн PipeSeq, включающий автоматизированные этапы загрузки, выравнивания и статистического анализа данных РНК-сек, а также позволяющий обсчитывать и визуализировать данные, получаемые методами секвенирования РНК и ПЦР-РВ. Результаты нашего исследования показали, что методы секвенирования РНК и ОТ-ПЦР-РВ дают возможность выявить схожие паттерны изменения экспрессии генов, но демонстрируют различия по оценке степени размера эффекта и чувствительности в детекции изменений.

Полученные данные дают возможность сделать вывод, что ФТ GATA *C. reinhardtii* образуют три функционально специализированные группы (кластеры), согласованная регуляция которых представляет собой ключевой механизм, обеспечивающий корректное прохождение клеточного цикла в изменяющихся условиях внешней среды. Профили экспрессии генов *GATA-2*, *GATA-4*, *GATA-5*, *GATA-6*, *GATA-8*, *GATA-10*, *GATA-11*, *GATA-12* позволяют предположить их участие в регуляции светозависимых процессов метаболизма. Гены *GATA-1*, *GATA-3*, *GATA-7*,

GATA-9 задействованы в переключении метаболизма при переходе свет/темнота и темнота/свет. Дальнейшие исследования ФТ GATA *C. reinhardtii* должны быть направлены на поиск и верификацию генов-мишеней и взаимодействий в регуляторных сетях, а также подтверждение предсказанных функций в ответ на изменение других условий культивирования.

Таким образом, программа PipeSeq продемонстрировала свою эффективность в комплексном исследовании дифференциальной экспрессии генов как инструмент для проведения полного цикла биоинформатического анализа данных РНК-сек с возможностью обработки данных ОТ-ПЦР-РВ и выполнения сравнительного статистического анализа полученных с помощью разных методов результатов. Разработанный пайплайн может быть использован при изучении профилей экспрессии генов любого объекта исследования.

Список литературы / References

Квитко К.В., Борщевская Т.Н., Чунаев А.С., Тугаринов В.В. Петергофская генетическая коллекция штаммов зеленых водорослей (*Chlorella*, *Scenedesmus*, *Chlamydomonas*). В: Культивирование коллекционных штаммов водорослей. Л.: Изд-во Ленингр. ун-та, 1983
[Kvitko K.V., Borschevskaya T.N., Chunaev A.S., Tugarinov V.V. Peterhof genetic collection of green algae strains (*Chlorella*, *Scenedesmus*, *Chlamydomonas*). In: Cultivation of Collection Strains of Algae. Leningrad, 1983 (in Russian)]

Afgan E., Baker D., Batut B., van den Beek M., Bouvier D., Čech M., Chilton J., … Soranzo N., Goecks J., Taylor J., Nekrutenko A., Blankenberg D. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*. 2018;46(W1):W537-W544. doi 10.1093/nar/gky379

Bray N.L., Pimentel H., Melsted P., Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;4(5):525-527. doi 10.1038/nbt.3519

Coenye T. Do results obtained with RNA-sequencing require independent verification? *Biofilm*. 2021;3:100043. doi 10.1016/j.bioflm.2021.100043

Conesa A., Madrigal P., Tarazona S., Gomez-Cabrero D., Cervera A., McPherson A., Szcześniak M.W., Gaffney D.J., Elo L.L., Zhang X., Mortazavi A. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17:13. doi 10.1186/s13059-016-0881-8

Derveaux S., Vandesompele J., Hellemans J. How to do successful gene expression analysis using real-time PCR. *Methods*. 2010;50(4):227-230. doi 10.1016/j.ymeth.2009.11.001

Di Tommaso P., Chatzou M., Floden E.W., Barja P.P., Palumbo E., Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol*. 2017;35(4):316-319. doi 10.1038/nbt.3820

Elahimanesh M., Najafi M. Differentially expressed *genes* of RNA-seq data are suggested on the intersections of normalization techniques. *Biochem Biophys Rep*. 2024;37:101618. doi 10.1016/j.bbrep.2023.101618

Ewels P., Magnusson M., Lundin S., Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048. doi 10.1093/bioinformatics/btw354

Goodstein D.M., Shu S., Howson R., Neupane R., Hayes R.D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., Rokhsar D.S. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(D1):D1178-D1186. doi 10.1093/nar/gkr944

Harris E.H. The *Chlamydomonas* Sourcebook. A Comprehensive Guide to Biology and Laboratory Use. San Diego: Academic Press, 1989

He F., Liu Q., Zheng L., Cui Y., Shen Z., Zheng L. RNA-Seq analysis of rice roots reveals the involvement of post-transcriptional regulation in response to cadmium stress. *Front Plant Sci*. 2015;6:1136. doi 10.3389/fpls.2015.01136

Hunter J.D. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95. doi 10.1109/MCSE.2007.55

Kim D., Langmead B., Salzberg S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-360. doi 10.1038/nmeth.3317

Kim D., Paggi J.M., Park C., Bennett C., Salzberg S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907-915. doi 10.1038/s41587-019-0201-4

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi 10.1093/bioinformatics/btp352

Li X., Wang C.Y. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci*. 2021;13:36. doi 10.1038/s41368-021-00146-0

Liao Y., Smyth G.K., Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi 10.1093/bioinformatics/btt656

Liu C., Wu G., Huang X., Liu S., Cong B. Validation of housekeeping genes for gene expression studies in an ice alga *Chlamydomonas* during freezing acclimation. *Extremophiles*. 2012;16:419-425. doi 10.1007/s00792-012-0441-4

Livak K.J., Schmittgen T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods*. 2001;25(4):402-408. doi 10.1006/meth.2001.1262

Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi 10.1186/s13059-014-0550-8

Luo X.M., Lin W.H., Zhu S., Zhu J.Y., Sun Y., Fan X.Y., Cheng M., … Liu L., Zhang M., Xie Q., Chong K., Wang Z.Y. Integration of light- and brassinosteroid-signaling pathways by a GATA transcription factor in *Arabidopsis*. *Dev Cell*. 2010;19(6):872-883. doi 10.1016/j.devcel.2010.10.023

Manfield I.W., Devlin P.F., Jen C.H., Westhead D.R., Gilmartin P.M. Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the *Arabidopsis* GATA gene family. *Plant Physiol*. 2007;143(2):941-958. doi 10.1104/pp.106.090761

Marioni J.C., Mason C.E., Mane S.M., Stephens M., Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*. 2008;18(9):1509-1517. doi 10.1101/gr.079558.108

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;17(1):10-12. doi 10.14806/ej.17.1.200

McKinney W. pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*. 2011;14(9):1-9.

Merchant S.S., Prochnik S.E., Vallon O., Harris E.H., Karpowicz S.J., Witman G.B., Terry A., … Werner G., Zhou K., Grigoriev I.V., Rokhsar D.S., Grossman A.R. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*. 2007;318(5848):245-250. doi 10.1126/science.1143609

Mölder F., Jablonski K.P., Letcher B., Hall M.B., Tomkins-Tinch C.H., Sochat V., Forster J., … Wilm A., Holtgrewe M., Rahmann S., Nahnsen S., Köster J. Sustainable data analysis with Snakemake. *F1000Res*. 2021;10:33. doi 10.12688/f1000research.29032.2

Mortazavi A., Williams B.A., McCue K., Schaeffer L., Wold B.J. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621-628. doi 10.1038/nmeth.1226

Müller N., Wenzel S., Zou Y., Künzel S., Sasso S., Weiß D., Prager K., Grossman A., Kottke T., Mittag M. A plant cryptochrome controls key features of the *Chlamydomonas* circadian clock and its life cycle. *Plant Physiol*. 2017;174(1):185-201. doi 10.1104/pp.17.00349

Muzellec B., Teleńczuk M., Cabeli V., Andreux M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics*. 2023;39(9):btad547. doi 10.1093/bioinformatics/btad547

Naito T., Kiba T., Koizumi N., Yamashino T., Mizuno T. Characterization of a unique GATA family gene that responds to both light and cytokinin in *Arabidopsis thaliana*. *Biosci Biotechnol Biochem*. 2007;71(6):1557-1560. doi 10.1271/bbb.60692

Pertea M., Pertea G.M., Antonescu C.M., Chang T.C., Mendell J.T., Salzberg S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290-295. doi 10.1038/nbt.3122

Pertea M., Kim D., Pertea G.M., Leek J.T., Salzberg S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650-1667. doi 10.1038/nprot.2016.095

Ren W., Kong L., Jiang S., Ma L., Wang H., Li X., Liu Y., Ma W., Yan X. Genome-wide identification, evolution, and characterization of GATA gene family and GATA gene expression analysis post-MeJA treatment in *Platycodon grandiflorum*. *J Plant Growth Regul*. 2025;44:155-167. doi 10.1007/s00344-024-11468-8

Reyes J.C., Muro-Pastor M.I., Florencio F.J. The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiol*. 2004;134(4):1718-1732. doi 10.1104/pp.103.037788

Riechmann J.L., Heard J., Martin G., Reuber L., Jiang C.Z., Keddie J., Adam L., … Broun P., Zhang J.Z., Ghandehari D., Sherman B.K., Yu G.L. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*. 2000;290(5499):2105-2110. doi 10.1126/science.290.5499.2105

Salomé P.A., Merchant S.S. A series of fortunate events: introducing *Chlamydomonas* as a reference organism. *Plant Cell*. 2019;31(8):1682-1707. doi 10.1105/tpc.18.00952

Sanchez-Tarre V., Kiparissides A. The effects of illumination and trophic strategy on gene expression in *Chlamydomonas reinhardtii*. *Algal Res*. 2021;54:102186. doi 10.1016/j.algal.2021.102186

Schmittgen T., Livak K. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc*. 2008;3:1101-1108. doi 10.1038/nprot.2008.73

Schröder P., Hsu B.Y., Gutsche N., Winkler J.B., Hedtke B., Grimm B., Schwechheimer C. B-GATA factors are required to repress high-light stress responses in *Marchantia polymorpha* and *Arabidopsis thaliana*. *Plant Cell Environ*. 2023;46(8):2376-2390. doi 10.1111/pce.14629

Schwechheimer C., Schröder P.M., Blaby-Haas C.E. Plant GATA factors: their biology, phylogeny, and phylogenomics. *Annu Rev Plant Biol*. 2022;73(1):123-148. doi 10.1146/annurev-arplant-072221-092913

Shi Y., He M. Differential gene expression identified by RNA-Seq and qPCR in two sizes of pearl oyster (*Pinctada fucata*). *Gene*. 2014;538(2):313-322. doi 10.1016/j.gene.2014.01.031

Virolainen P.A., Chekunova E.M. GATA family transcription factors in alga *Chlamydomonas reinhardtii*. *Curr Genet*. 2024;70(1):1. doi 10.1007/s00294-024-01280-y

Voigt J., Münzner P. The *Chlamydomonas* cell cycle is regulated by a light/dark-responsive cell-cycle switch. *Planta*. 1987;172:463-472. doi 10.1007/BF00393861

Wang Z., Gerstein M., Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57-63. doi 10.1038/nrg2484

Wheeler D.L., Barrett T., Benson D.A., Bryant S.H., Canese K., Church D.M., DiCuccio M., … Suzek T.O., Tatusov R., Tatusova T.A., Wagner L., Yaschenko E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2005;33:D39-D45. doi 10.1093/nar/gki062

Zhao S., Ye Z., Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA*. 2020;26(8):903-909. doi 10.1261/rna.074922.120

Zhao Y., Li M.C., Konaté M.M., Chen L., Das B., Karlovich C., Williams P.M., Evrard Y.A., Doroshow J.H., McShane L.M. TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. *J Transl Med*. 2021;19(1):269. doi 10.1186/s12967-021-02936-w

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 12.08.2025. После доработки 21.12.2025. Принята к публикации 22.12.2025.