

Перевод на английский язык <https://vavilov.elpub.ru/jour>

Однонуклеотидные замены в геноме человека, влияющие на связывание факторов транскрипции, и их роль в развитии патологий

Е.В. Антонцева , А.О. Дегтярева, Е.Е. Корболина, И.С. Дамаров, Т.И. Меркулова

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
 antontseva@bionet.nsc.ru

Аннотация. Однонуклеотидные замены, также называемые однонуклеотидными полиморфизмами (single nucleotide polymorphism, SNP), – это наиболее распространенный тип вариаций генома человека. Подавляющая часть выявленных в геноме человека SNP не оказывает какого-либо воздействия на молекулярный фенотип, однако некоторые способны приводить к изменению функции гена или уровня его экспрессии. В то же время большинство SNP, ассоциированных с некими признаками или патологиями, картируются в регуляторных областях генома, изменяют потенциальные сайты связывания транскрипционных факторов и, соответственно, могут влиять на экспрессию генов. В последние десятилетия значительные усилия были направлены на поиск таких регуляторных SNP (rSNP), а также на понимание механизмов, посредством которых они приводят к фенотипическим различиям, в первую очередь к разной предрасположенности к заболеваниям и индивидуальной чувствительности к лекарственным препаратам. Развитие технологии NGS (next generation sequencing) способствовало не только выявлению огромного количества SNP и поиску их ассоциации (genome wide association studies, GWAS) с некими заболеваниями или фенотипическими проявлениями, но и развитию более производительных подходов для их функциональной аннотации. Стоит отметить, что наличие ассоциации не позволяет выделить функциональный, действительно связанный с болезнью вариант последовательности ДНК из множества маркерных, которые выявляются за счет неравновесия по сцеплению. Более того, установление ассоциаций генетических вариантов с заболеванием не дает сведений о функциональности этих вариантов, что необходимо для выяснения молекулярных механизмов развития патологии и разработки эффективных методов ее лечения и профилактики. В связи с этим функциональный анализ SNP, аннотированных в GWAS каталоге, как на полногеномном уровне, так и на уровне отдельных SNP в последние годы стал особенно актуальным. В настоящее время активно развивается полногеномный поиск потенциально регуляторных SNP без каких-либо предварительных знаний об их ассоциации с признаком. Так, картирование локусов количественных признаков экспрессии (eQTL, expression quantitative trait loci) позволяет выявить SNP, для которого в транскриптомах гомозигот по разным его аллелям, а также гетерозигот наблюдаются различия в уровне экспрессии неких генов, причем как близко расположенных, так и на значительном удалении. Для предсказания регуляторных SNP используют также подходы, основанные на поиске аллель-специфических событий в данных RNA-seq, ChIP-seq, DNase-seq, ATAC-seq, MPRA и т.д. Однако для более полной характеристики таких rSNP необходимо устанавливать их ассоциацию с признаком, в частности с предрасположенностью к некой патологии или с чувствительностью к лекарственным препаратам. Таким образом, именно комплексное использование двух основанных на противоположных принципах подходов к поиску значимых для развития признака (патологии) SNP: с одной стороны, исходящего из данных по ассоциации SNP с неким признаком, а с другой стороны, идущего от определения аллель-специфичных изменений на молекулярном уровне (в транскриптоме или регуломе) – существенно обогащает картину наших знаний о роли генетических детерминант в молекулярных механизмах формирования признаков, включая предрасположенность к многофакторным заболеваниям.

Ключевые слова: регуляторный однонуклеотидный полиморфизм; сайты связывания транскрипционных факторов; экспрессия генов; полногеномные исследования.

Для цитирования: Антонцева Е.В., Дегтярева А.О., Корболина Е.Е., Дамаров И.С., Меркулова Т.И. Однонуклеотидные замены в геноме человека, влияющие на связывание факторов транскрипции, и их роль в развитии патологий. *Вавиловский журнал генетики и селекции*. 2023;27(6):662-675. DOI 10.18699/VJGB-23-77

Human-genome single nucleotide polymorphisms affecting transcription factor binding and their role in pathogenesis

E.V. Antontseva , A.O. Degtyareva, E.E. Korbolina, I.S. Damarov, T.I. Merkulova

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
 antontseva@bionet.nsc.ru

Abstract. Single nucleotide polymorphisms (SNPs) are the most common type of variation in the human genome. The vast majority of SNPs identified in the human genome do not have any effect on the phenotype; however, some can lead to changes in the function of a gene or the level of its expression. Most SNPs associated with certain traits or pathologies are

mapped to regulatory regions of the genome and affect gene expression by changing transcription factor binding sites. In recent decades, substantial effort has been invested in searching for such regulatory SNPs (rSNPs) and understanding the mechanisms by which they lead to phenotypic differences, primarily to individual differences in susceptibility to diseases and in sensitivity to drugs. The development of the NGS (next-generation sequencing) technology has contributed not only to the identification of a huge number of SNPs and to the search for their association (genome-wide association studies, GWAS) with certain diseases or phenotypic manifestations, but also to the development of more productive approaches to their functional annotation. It should be noted that the presence of an association does not allow one to identify a functional, truly disease-associated DNA sequence variant among multiple marker SNPs that are detected due to linkage disequilibrium. Moreover, determination of associations of genetic variants with a disease does not provide information about the functionality of these variants, which is necessary to elucidate the molecular mechanisms of the development of pathology and to design effective methods for its treatment and prevention. In this regard, the functional analysis of SNPs annotated in the GWAS catalog, both at the genome-wide level and at the level of individual SNPs, became especially relevant in recent years. A genome-wide search for potential rSNPs is possible without any prior knowledge of their association with a trait. Thus, mapping expression quantitative trait loci (eQTLs) makes it possible to identify an SNP for which – among transcriptomes of homozygotes and heterozygotes for its various alleles – there are differences in the expression level of certain genes, which can be located at various distances from the SNP. To predict rSNPs, approaches based on searches for allele-specific events in RNA-seq, ChIP-seq, DNase-seq, ATAC-seq, MPRA, and other data are also used. Nonetheless, for a more complete functional annotation of such rSNPs, it is necessary to establish their association with a trait, in particular, with a predisposition to a certain pathology or sensitivity to drugs. Thus, approaches to finding SNPs important for the development of a trait can be categorized into two groups: (1) starting from data on an association of SNPs with a certain trait, (2) starting from the determination of allele-specific changes at the molecular level (in a transcriptome or regulome). Only comprehensive use of strategically different approaches can considerably enrich our knowledge about the role of genetic determinants in the molecular mechanisms of trait formation, including predisposition to multifactorial diseases. Key words: regulatory single-nucleotide polymorphism; transcription factor-binding sites; gene expression; genome-wide studies.

For citation: Antontseva E.V., Degtyareva A.O., Korbolina E.E., Damarov I.S., Merkulova T.I. Human-genome single nucleotide polymorphisms affecting transcription factor binding and their role in pathogenesis. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2023;27(6):662-675. DOI 10.18699/VJGB-23-77

Введение

Одной из основных задач генетики человека является выяснение механизмов, посредством которых вариации генома приводят к фенотипическим различиям, в первую очередь к разной предрасположенности к заболеваниям и индивидуальной чувствительности к лекарственным препаратам. Однонуклеотидные замены (SNP) – наиболее распространенный тип вариаций генома (Chanock, 2001). В настоящее время благодаря развитию технологий NGS (next generation sequencing) в базе dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi) (Sherry et al., 2001) зарегистрировано свыше 950 млн SNP в геноме человека, причем более 90 % от общего числа приходится на долю редких, с частотой встречаемости менее 1 %. Представляется маловероятным, что вся масса обнаруженных вариаций может иметь значение для фенотипа, но какая-то их часть, безусловно, является генетической основой фенотипических проявлений, включая предрасположенность к различным заболеваниям. В последние десятилетия на поиск таких SNP были направлены значительные усилия.

Доминирующим подходом к идентификации связанных с признаком однонуклеотидных замен, стартовавшим еще в 1980-х гг., было определение ассоциации с заболеваниями для SNP, обнаруженных в генах-кандидатах (Lander, Schork, 1994; Ring, Kroetz, 2002). Как правило, исследовались лишь SNP, затрагивающие белок-кодирующую часть гена (Cooper, 1998). Несколько позже возник интерес к выяснению функциональности вариантов, расположенных в некодирующих районах генов, т. е. к определению влияния таких вариантов на некий молекулярный фенотип. В частности, было показано, что такие SNP затрагивают сайты связывания транскрипционных факторов (ССТФ), что приводит к изменению экспрессии

соответствующих генов (Ludlow et al., 1996; Piedrafita et al., 1996; Knight et al., 1999; Vasiliev et al., 1999). Однако эти малочисленные функциональные исследования оставались почти незаметными на фоне огромного вала работ по выявлению ассоциаций.

Масштабы обнаружения связанных с заболеваниями SNP резко возросли с внедрением в середине 2000-х гг. технологии GWAS (genome wide association study) (рис. 1, а), которая основана на непредвзятом, не исходящем из каких-либо представлений о формировании признака принципе выявления ассоциированных с ним SNP в масштабе генома (Visscher et al., 2012; Tam et al., 2019). К настоящему времени с помощью GWAS найдено уже более 72 тыс. ассоциаций генетических вариантов с признаками (GWAS Catalog, <https://www.ebi.ac.uk/gwas/>), что позволило выявить много новых генов и генных систем, связанных с предрасположенностью к различным болезням (Buniello et al., 2019; Tam et al., 2019; Claussnitzer et al., 2020).

Однако технология GWAS не дает никакой информации о функциональности обнаруженных вариантов, что сильно затрудняет выяснение молекулярных механизмов, лежащих в основе развития патологии, и, как следствие, разработку эффективных методов ее лечения и профилактики. Помимо того, по результатам GWAS почти невозможно отличить действительно связанный с болезнью вариант от множества маркерных, которые выявляются за счет неравновесия по сцеплению (Lappalainen, 2015; Tam et al., 2019; Zhao et al., 2020). Известно также, что большинство обнаруженных с помощью GWAS SNP располагается в некодирующей части генома и, как правило, в его регуляторных районах (промоторах, энхансерах и т. д.) (Hindorff et al., 2009; Maurano et al., 2012; Bryzgalov

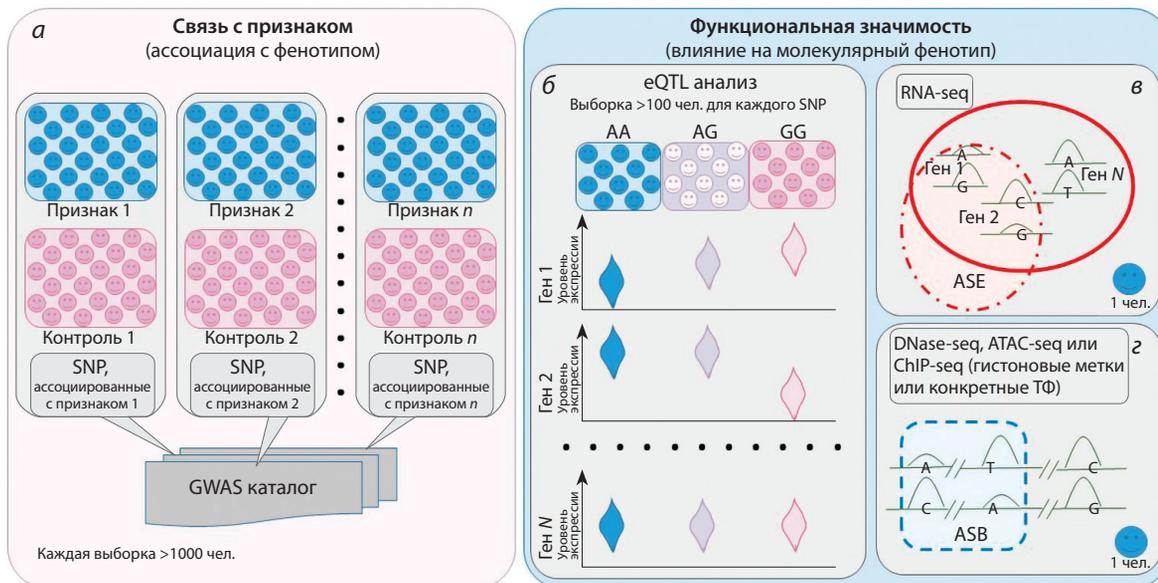


Рис. 1. Полногеномные подходы для анализа SNP.

а – принцип GWAS исследований; б – схема eQTL анализа для каждого SNP; в, з – поиск аллель-специфичных событий: в – экспрессии (ASE), з – связывания (ASB).

et al., 2013; Farh et al., 2015), что предполагает их влияние на связывание транскрипционных факторов (ТФ) и экспрессию генов. Таким образом, стала очевидной необходимость проведения исследований, посвященных функциональной интерпретации данных GWAS как на уровне отдельных потенциально регуляторных SNP (rSNP), так и на массовом уровне. Кроме того, поскольку технология GWAS сильно занижает число выявленных ассоциаций по сравнению с реально существующими в силу необходимости применения жестких порогов при статистической обработке (Tam et al., 2019), возникло осознание необходимости разработки независимых от GWAS масштабных функциональных подходов к поиску rSNP (Westra, Franke, 2014; Maurano et al., 2015; Cavalli et al., 2016b; Korbolina et al., 2018).

Из полногеномных функциональных подходов первым стал (почти одновременно с началом применения технологии GWAS) eQTL анализ (см. рис. 1, б), который, опираясь на данные транскриптомов (ранее использовались данные микрочипового анализа, затем данные РНК-секвенирования, RNA-seq), определяет для каждого SNP разницу в уровне экспрессии отдельных генов у гомозигот по разным аллелям этого SNP, а также его гетерозигот (Westra, Franke, 2014; GTEx Consortium, 2020). Несколько позднее были разработаны методы поиска аллель-специфичных событий как в данных RNA-seq (аллель-специфичные события экспрессии, ASE) (см. рис. 1, в) (Castel et al., 2020; Fan et al., 2020), так и в данных ChIP-seq, DNase-seq и ATAC-seq (аллель-специфичные события связывания, ASB) (см. рис. 1, з) (Maurano et al., 2015; Cavalli et al., 2016a, b; Xu et al., 2020; Korbolina et al., 2021). К числу массовых функциональных подходов относятся также MPRA (massively parallel reporter assay), SNP-seq, and SNP-SELEX (Zhang et al., 2018; Lu et al., 2021; Yan et al., 2021).

В начале настоящего обзора на примерах современных работ рассмотрены rSNP, связанные с развитием патологий по данным GWAS и подробно охарактеризованные как влияющие на взаимодействие с ТФ и экспрессию близлежащих или удаленных генов. Затем описано применение подходов функциональной геномики для интерпретации данных GWAS, а также для независимого от GWAS поиска новых регуляторных вариантов. По ходу изложения показаны сильные и слабые стороны этих подходов и важность комплексного применения GWAS и методов функциональной геномики для выяснения роли SNP в молекулярных нарушениях, лежащих в основе развития патологии.

Функциональная интерпретация данных GWAS на уровне отдельных rSNP

Для функционального изучения отдельных потенциальных rSNP используется широкий спектр экспериментальных методов. На первых этапах обычно применяют классические: метод задержки ДНК-пробы в геле белками ядерного экстракта (electrophoretic mobility shift assay, EMSA) и репортерный анализ (анализ экспрессии репортерного гена под контролем аллельных вариантов района расположения SNP), которые позволяют зарегистрировать факт влияния SNP на связывание некоего ТФ или на экспрессию репортерного гена соответственно (Antontseva et al., 2015; Fang et al., 2017). С целью идентификации ТФ, сайты связывания которых затрагивает замена нуклеотида, проводится EMSA с использованием соответствующих антител или очищенных ТФ (Piedrafita et al., 1996; Knight et al., 1999; Vasiliev et al., 1999; Jiang et al., 2020), иммунопреципитация хроматина с детекцией аллель-асимметрии в продукте ПЦР (ChIP-AS-qPCR) (Gao et al., 2018; Choi et al., 2020; Thynn et al., 2020; Protze et al., 2022) и масс-спектрометрический анализ белков, вы-

деленных из комплексов с олигонуклеотидами, содержащими альтернативные аллели (Fang et al., 2017; Liu D. et al., 2018; Choi et al., 2020).

Влияние нуклеотидной замены в идентифицированных ССТФ на их потенциальные гены-мишени подтверждают в экспериментах с подавлением или искусственным усилением экспрессии генов соответствующих ТФ или же путем изучения эффекта точковых мутаций, внесенных в ССТФ с помощью технологии CRISPR/Cas9 (Prestel et al., 2019; Gutierrez-Arcelus et al., 2020; Pan et al., 2020; Thynn et al., 2020; Wang Y. et al., 2020; Wang X. et al., 2021). В последнее время все более популярным подходом в этом направлении исследований SNP становится также определение аллельного дисбаланса экспрессии для SNP, попадающих в транскрибируемые области генов. Однако стоит отметить, что такой SNP может быть как непосредственно регуляторным (Syddall et al., 2013; Klein et al., 2019), так и маркерным, находящимся в группе сцепления с rSNP, расположенным в нетранскрибируемом районе (Fang et al., 2017; Li X.-X. et al., 2019; Peng et al., 2020).

В современных исследованиях используются различные сочетания перечисленных методов, что хорошо иллюстрируют приведенные ниже примеры.

rs36115365

Хрестоматийным примером хорошо изученного rSNP является rs36115365 (G/C), расположенный в локусе, связанном, по данным GWAS, с различными типами рака (chr5p15.33 – район 2). Корреляционный анализ выявил в этом локусе девять SNP ($r^2 > 0.60$, 1000G EUR population), ассоциированных с развитием рака поджелудочной железы, герминогенной опухоли яичка, рака легких и меланомой. Для скрининга всех девяти SNP на регуляторную активность применялись EMSA и репортерный анализ, которые проводились на восьми клеточных линиях человека (Fang et al., 2017). Использование нескольких клеточных линий – обычная практика подобных исследований (Bryzgalov et al., 2013; Boldes et al., 2020), направленная на максимально возможный охват событий взаимодействия ТФ с их сайтами связывания, что обусловлено существенными различиями в наборах экспрессирующихся ТФ в разных типах клеток (Tobias et al., 2021). Скрининговые исследования идентифицировали в качестве потенциально регуляторного только rs36115365, аллель С которого демонстрировал как гораздо лучшее связывание с неким белком в EMSA, так и большую активацию экспрессии репортерного гена по сравнению с аллелем G (Fang et al., 2017).

rs36115365 расположен на расстоянии ~18 т. п. н. выше начала гена *TERT*, кодирующего обратную транскриптазу теломеразного комплекса, и на расстоянии ~5 т. п. н. ниже конца гена *CLPTMIL*. Согласно данным ChIP-seq (иммунопреципитация хроматина с последующим секвенированием ДНК из преципитатов) из проекта ENCODE (Moore et al., 2020), участок его локализации перекрывается со множеством сайтов связывания ТФ и обогащен гистоновыми метками активного хроматина, что характерно для энхансерных районов. Инактивация этого района с помощью siRNA-опосредованного сайленсинга транскрипции (Malecová, Morris, 2010) привела к снижению

экспрессии только гена *TERT*. Для идентификации ТФ, сайт связывания которого изменяется в результате замены G на С, была проведена инкубация белков экстрактов ядер с олигонуклеотидами, содержащими альтернативные аллели, с последующим масс-спектрометрическим анализом связавшихся белков. По результатам анализа полученных пептидов детектировались четыре ТФ-кандидата, предпочитающих аллель С: ZNF148, VEZF1/ZNF161, ZNF281 и ZNF740. С помощью EMSA и специфических антител к этим ТФ подтвердился только ZNF148, что впоследствии было верифицировано в эксперименте по связыванию с очищенным белком ZNF148. Точку поставил siRNA-опосредованный нокдаун по ZNF148, который привел к снижению экспрессии *TERT* и активности теломеразы, а также к уменьшению длины теломер. Таким образом, аллель С исправляет сайт связывания ZNF148, усиливает экспрессию *TERT* и, как следствие, увеличивает риск канцерогенеза (Fang et al., 2017).

rs174575

Согласно данным GWAS, rs174575 (C/G) связан с повышенным риском развития колоректального рака (Tian et al., 2020). Этот SNP расположен в первом интроне гена дельта-6-десатуразы (*FADS2*) на расстоянии +41.5 т. п. н. от его старта транскрипции (transcription start-site, TSS) и на расстоянии –178.8 т. п. н. от TSS гена длинной некодирующей РНК AP002754.2. Участок локализации rs174575 в хроматине обогащен модификациями гистонов, характерными для активных регуляторных районов (H3K4me1, H3K4me3, H3K27ac), и, по данным DNase-seq (определение участков гиперчувствительности к DNase I), а также ATAC-seq (участков, доступных для транскриптазы, Assay for Transposase-Accessible Chromatin), соответствует открытому хроматину (ENCODE). Предположение о регуляторной роли rs174575 подтверждается данными eQTL анализа, показавшими связь аллеля G с повышенным уровнем экспрессии *FADS2* и AP002754.2.

Компьютерное исследование ДНК-мотивов с использованием ресурсов Cistrome (Zheng et al., 2019) и JASPAR (Fornes et al., 2020) обнаружило, что в случае rs174575 замена G на С приводит к повреждению сайта связывания ТФ E2F1, а данные ChIP-seq из ENCODE, полученные на клетках колоректального рака LoVo, показали, что E2F1 картируется в области локализации данного SNP. Кросс-конкурентный EMSA подтвердил лучшее связывание некоего белка ядерного экстракта с G аллелем, а то, что этот белок – E2F1, показал ChIP-qPCR с соответствующими антителами. Так, в клеточных линиях с различными генотипами: HCT116 (CG), SNU-C1 (CG) и HT115 (CC), наблюдалось более сильное связывание E2F1 в клетках, несущих аллель G. Исследование с использованием репортерных конструкций тоже подтвердило более высокую энхансерную активность фрагмента, содержащего аллель G (Tian et al., 2020).

Непосредственный контакт между областью, содержащей rs174575, и промоторами *FADS2* и AP002754.2 был установлен с помощью метода захвата конформации хромосомы (3C). При этом взаимодействие было гораздо более выраженным в клеточных линиях, содержащих аллель G. Дальнейшие эксперименты показали, что сверх-

экспрессия AP002754.2 резко увеличивает уровень экспрессии *FADS2* в клетках HCT116 и LoVo, а нокадаун AP002754.2 с помощью микроРНК приводит к снижению уровня экспрессии этого гена, что указывает на стимулирующую роль AP002754.2 в регуляции гена *FADS2*. Показано, что сверхэкспрессия *FADS2* и AP002754.2 существенно увеличивает скорость пролиферации клеток HCT116 и LoVo, а нокадаун *FADS2* или AP002754.2 ускоряет рост опухоли в экспериментах *in vivo* на мышах. Известно, что продукт гена *FADS2* – ключевой фермент в биосинтезе полиненасыщенных жирных кислот, включая арахидоновую кислоту, которая, в свою очередь, является предшественником простагландина E2 (PGE2), способствующего росту опухоли и метастазированию. Таким образом, rs174575 действует как аллель-специфический энхансер, стимулирующий транскрипцию *FADS2* и AP002754.2, что приводит к повышенному риску развития колоректального рака в случае аллеля G (Tian et al., 2020).

rs4903064

Согласно данным GWAS, rs4903064 (T/C) связан с почечно-клеточным раком (ПКР), наиболее частым видом рака почки (Scelo et al., 2017). Он расположен в третьем интроне гена *DPF3*, который кодирует белок подсемейства BAF комплекса ремоделирования хроматина SWI/SNF. Показано преобладание аллеля С rs4903064 в образцах опухолевых тканей пациентов с диагнозом «светлоклеточный ПКР» по сравнению с нормальными тканями почки, а также с опухолевой тканью лиц с папиллярным ПКР и хромофобным ПКР. При этом измерение соотношения аллелей в пре-мРНК *DPF3* пациентов со светлоклеточным ПКР, гетерозиготных по rs4903064, подтвердило дисбаланс в сторону аллеля С (Protze et al., 2022).

По данным ATAC-seq, полученным на первичных клетках рака почки, rs4903064 находится в области открытого хроматина, в предполагаемом энхансере. Поскольку замена Т на С (rs4903064) создает потенциальный сайт связывания ТФ HIF, был проведен репортерный анализ на клеточных линиях HeLa и MCF-7. Усиление экспрессии репортерного гена наблюдалось только в случае аллеля С и только при обработке клеток специфическим стабилизатором HIF – диметилноксалилглицином (DMOG). Нокаут по различным изоформам HIF показал, что увеличение репортерной активности при стимуляции DMOG зависит от HIF-1 α . С помощью ChIP-qPCR в первичных клетках почечных канальцев с разными генотипами rs4903064 (ТТ, СТ, СС) подтверждено повышенное связывание HIF-1 α и HIF-1 β с рисковым аллелем С. Для выяснения роли повышенной экспрессии *DPF3* в развитии светлоклеточного ПКР с помощью технологии CRISPR/Cas9 был проведен нокаут *DPF3* в клетках проксимальных почечных канальцев. Продемонстрировано, что клетки с дефектной экспрессией *DPF3* росли медленнее, чем контрольные клоны соответствующих клеток. Это позволяет предположить, что усиление экспрессии *DPF3* в клетках проксимальных канальцев стимулирует пролиферацию (Protze et al., 2022).

rs17114036

rs17114036 (T/C), связанный с ишемической болезнью сердца и ишемическим инсультом, согласно данным GWAS (Dichgans et al., 2014), локализован в пятом интроне гена *PLPP*, кодирующем фосфатазу фосфолипидов 3, которая ингибирует воспаление эндотелия и способствует целостности его монослоя (Panchatcharam et al., 2014; Wu et al., 2015). Эксперименты ATAC-seq и ChIP-seq (H3K27ac, H3K4me2), проведенные на эндотелиальных клетках аорты человека HAEC, идентифицировали область, содержащую rs17114036, как потенциальный энхансер. Энхансерная активность данной области была подтверждена с помощью репортерного анализа, при этом протективный аллель С значительно сильнее увеличивал активность люциферазы при трансфекции векторных конструкций в клетки HAEC по сравнению с аллелем Т. Делеция участка длиной 66 п. н., содержащего rs17114036, с помощью CRISPR/Cas9 значительно снижала экспрессию *PLPP3* по сравнению с неотредактированным геномом и повышала проницаемость монослоя отредактированных HAEC. В результате моделирования процессов гемодинамики было обнаружено увеличение активности исследуемого энхансера, содержащего аллель С, в HAEC при 18-часовом «атерозащитном» потоке по сравнению с «атерогенным» потоком, тогда как никакого эффекта аллеля Т не выявлено. Как выяснилось, замена нуклеотида Т на С приводит к образованию сайта связывания (CACCC) для фактора KLF2, что подтверждено данными ChIP-AS-qPCR анализа в клетках HAEC, гетерозиготных по rs17114036. Эксперименты по котрансфекции с плазмидой, обеспечивающей сверхэкспрессию KLF2, также показали более высокую активность люциферазы в случае тестируемого энхансера, несущего аллель С (Krause et al., 2018).

rs4407214

rs4407214 (T/G) ассоциирован с негативным по эстрогеновому рецептору раком молочной железы. Анализ данных ChIP-seq из проекта ENCODE по расположению меток активного хроматина (H3K4m1, H3K4m2, H3K4m3, H3K9ac, H3K27ac, H3K36m3, H3K79m2, H4K20m1, EZH2 и H2AZ) в клеточной линии HMEC позволил выявить в области rs4407214 регуляторный район, попадающий в интрон 1 гена *WDR43* (белок-кодирующий ген, связанный с процессингом рРНК и рибосомальным биогенезом). С помощью EMSA было показано аллель-специфическое связывание ядерных белков из клеточных линий MCF10A и CAL-51 с ДНК-зондами, воспроизводящими район локализации SNP в геноме. Экспрессия репортерного гена под контролем выявленного регуляторного района на этих же клеточных линиях тоже имела аллель-зависимый характер (Couch et al., 2016; Fachal et al., 2020).

Для идентификации ТФ, сайты связывания которых изменяются из-за замены G на Т (rs4407214), был проведен биоинформатический анализ с использованием базы JASPAR и доступных данных ChIP-seq для этого района из проекта ENCODE. В результате в качестве такого ТФ был идентифицирован USF1. Конкурентный анализ EMSA с ядерными белками, выделенными из клеточных линий CAL-51 и MCF10A, подтвердил аллель-специфи-

ческое связывание USF1 в случае аллеля G. Также было показано, что CRISPR/Cas9 опосредованное удаление предполагаемой регуляторной области, содержащей rs4407214, приводило к снижению уровня экспрессии *PLB1* (фосфолипазы B1), находящегося на расстоянии около 400 т. п. н. от *WDR43*.

Массовая интерпретация данных GWAS с помощью методов функциональной геномики

По результатам GWAS, с признаком (заболеванием), как правило, связано множество локусов (Goldstein, 2009; Boyle et al., 2017), большинство из которых, в свою очередь, содержит множество часто совместно наследуемых SNP (Tak, Farnham, 2015; Schaid et al., 2018), поэтому выбор среди них потенциальных rSNP, вовлеченных в механизм развития патологии, весьма затруднен. К настоящему времени разработан ряд эффективных экспериментальных подходов к решению этой задачи. Прежде всего, это масштабированные варианты подходов, используемых для исследования отдельных rSNP: массовое параллельное исследование репортерных конструкций (massively parallel reporter assay, MPRA), Reel-seq, SNP-seq и SNP-SELEX (Zhao et al., 2020; Lu et al., 2021; Yan et al., 2021).

В частности, MPRA был успешно применен для поиска rSNP, играющих ключевую роль в генетической предрасположенности к красной волчанке (Lu et al., 2021). По данным GWAS, с этим заболеванием ассоциировано 3073 SNP в 91 локусе. Авторы сконструировали баркодированную библиотеку, содержащую 12396 олигонуклеотидов длиной 170 п. н., содержащих в центре все известные варианты для этих 3073 SNP, которые были встроены перед минимальным промотором, присоединенным к гену *eGFP*. На клеточной линии GM12878 было показано влияние альтернативного аллеля на энхансерную активность вставок для 51 SNP из 27 локусов. Небольшое количество выявленных rSNP, вероятнее всего, связано с трансфекцией только одной линии клеток; при использовании нескольких клеточных линий их число стало бы существенно больше из-за увеличения набора вовлеченных в исследование ТФ. Подобным образом с помощью MPRA было определено, что 30 из 832 SNP, связанных, по данным GWAS, с риском развития меланомы, обнаруживают существенную разницу во влиянии альтернативных аллелей на экспрессию репортерного гена в клеточной линии меланомы UACC903 (Choi et al., 2020). Для одного из этих 30 SNP (rs398206), расположенного в интроне 1 гена *MX2*, разница была максимальной. Установлено, что рискованный аллель А rs398206 существенно усиливает связывание ТФ YY1 в условиях *in vitro* (EMSA) и *in vivo* (ChIP-AS-qPCR), что приводит к увеличению уровня экспрессии *MX2*, способствующему развитию меланомы (Choi et al., 2020). Другие примеры можно найти в работах (Ulirsch et al., 2016; Liu S. et al., 2017; Kalita et al., 2018; Klein et al., 2019).

С помощью методически мало различающихся Reel-seq и SNP-seq, основанных на сравнительном изучении связывания ТФ с олигонуклеотидами, содержащими альтернативные аллели, был отобран 521 потенциальный rSNP из 4316 связанных с развитием рака молочной железы по данным GWAS (Zhao et al., 2020) и 403 из 903 ассо-

циированных с раком простаты соответственно (Zhang et al., 2018). Наиболее масштабное исследование с использованием подхода из данной группы опубликовано в 2021 г. (Yan et al., 2021). Подход был назван SNP-SELEX. Для его реализации авторы использовали библиотеку из 383544 олигонуклеотидов длиной 40 п. н., содержащих в центральной части все возможные аллели 95886 SNP. SNP были выбраны по принципу их ассоциации с помощью GWAS с диабетом второго типа или по принципу локализации в окне размером 500 т. п. н., которое содержало бы связанный с этой патологией вариант. Используя 270 рекомбинатных ТФ, авторы провели мультиплексный анализ их связывания с олигонуклеотидами и выявили 11079 SNP, демонстрирующих существенное влияние аллеля на связывание хотя бы с одним ТФ.

Для массовой интерпретации результатов GWAS широко используются также данные функциональной геномики из доступных баз, такие как данные о полногеномных профилях (ChIP-seq) связывания ТФ (Li S. et al., 2020) и гистоновых модификаций (Jones et al., 2020), полногеномных профилях открытого хроматина (ATAC-seq) (Corces et al., 2020), трехмерных хроматиновых контактах (Corces et al., 2020) и расположении энхансерных и суперэнхансерных районов (Gong et al., 2018; Sun W. et al., 2018; Nasser et al., 2021), данные eQTL анализа (Gamazon et al., 2018; Zheng et al., 2019; Barbeira et al., 2021).

Так, например, 8005 SNP, либо непосредственно связанных, по данным GWAS, с большим депрессивным расстройством, либо находящихся в той же группе сцепления, были картированы в пиках ChIP-seq, полученных на тканях мозга или в клетках нейронального происхождения с использованием антител к 34 различным ТФ (Li S. et al., 2020). Затем был произведен поиск сайтов связывания соответствующих ТФ в местах локализации SNP с использованием базы данных, содержащей 7699 весовых матриц (PWM) (Whittington et al., 2016), и обнаружено, что 34 SNP разрушают сайты связывания 15 ТФ. Репортерный анализ подтвердил влияние аллеля на экспрессию для 29 SNP. Один из них, rs3101339, попадал в потенциальный сайт связывания ТФ REST в промоторном районе гена *NEGR1*, а замена А на С существенно повреждала структуру данного сайта, что отражалось в снижении уровня экспрессии репортерного гена под контролем вставки, несущей аллель С. Влияние rs3101339 на экспрессию гена *NEGR1* в условиях *in vivo* было подтверждено путем элиминации соответствующего фрагмента с помощью CRISPR-Cas9 геномного редактирования. Поскольку продукт гена *NEGR1* играет важную роль в поддержании необходимой плотности дендритных шипиков, снижение его экспрессии при замене А на С может вносить существенный вклад в развитие депрессивного состояния (Li S. et al., 2020).

Данные о полногеномных профилях гистоновых меток активного хроматина тоже оказываются весьма информативными для функциональной интерпретации результатов GWAS. Например, для анализа множества SNP, связанных с развитием эпителиального рака яичников, было проведено их картирование в области ChIP-seq пиков для H3K27Ac, полученных авторами при исследовании 26 образцов этого типа рака (Jones et al., 2020). Затем с

помощью инструмента motifbreakR (Coetzee et al., 2015) среди картированных SNP были выбраны 469, в которых нуклеотидная замена существенно меняла сайт связывания какого-либо ТФ. Наиболее частым было изменение последовательности сайта связывания ТФ REST, для которого отмечены как онкосупрессорная, так и онкогенная функции (Jones et al., 2020). Использование массивов ChIP-seq данных о различных гистоновых модификациях из соответствующих баз в сочетании с данными и инструментами ресурса потенциальных регуляторных вариантов (rVarBase) (Guo et al., 2016) позволило также обнаружить 286 и 366 локализованных в суперэнхансерах потенциальных rSNP, связанных с диабетом второго типа (Sun W. et al., 2018) и ишемической болезнью сердца (Gong et al., 2018) соответственно, которые меняли предсказанные ССТФ.

Независимые от GWAS функциональные подходы к выявлению потенциальных rSNP

Современные массовые функциональные подходы для поиска потенциальных rSNP в основном опираются на регистрацию влияния нуклеотидной замены на некий молекулярный фенотип. Это может быть определение в транскриптомах разницы в уровне экспрессии отдельных генов у гомозигот по разным аллелям, а также гетерозигот для каждого SNP (eQTL анализ); идентификация SNP, проявляющих асимметрию представленности в транскриптомных (RNA-seq) – аллель-асимметричные события экспрессии (ASE), или в эпигеномных данных (DNase-seq, ChIP-seq, ATAC-seq) – аллель-асимметричные события связывания (ASB), а также определение эффекта аллеля на экспрессию репортерного гена с помощью массового параллельного репортерного исследования (MPRA).

eQTL анализ

Термин eQTL либо означает факт наличия связи между вариантом (eVariant) и уровнем экспрессии некоего гена/генов (eGene/eGenes) (GTEx Consortium, 2017, 2020), либо относится непосредственно к SNP, аллели которого демонстрируют такую связь, и в этом смысле употребляется гораздо чаще (Fairfax et al., 2014; Fan et al., 2020; Jiang et al., 2020; Werling et al., 2020). Для поиска eQTL пригодны транскриптомные данные, полученные как с использованием микрочипов (Fairfax et al., 2014; Westra, Franke, 2014), так и с помощью РНК-секвенирования (RNA-seq) (GTEx Consortium, 2020). Их вполне достаточно для детекции eQTL, расположенных в транскрибируемых районах (Göring et al., 2007). Для выявления же всего их множества требуются также данные по секвенированию генома (GTEx Consortium, 2020; Werling et al., 2020). В отличие от GWAS, для проведения которых необходимы биологические образцы от многих тысяч индивидуумов (Tam et al., 2019), для eQTL анализа достаточно нескольких сотен участников (Westra et al., 2013; Fairfax et al., 2014; GTEx Consortium, 2020). Однако так же, как и в случае GWAS, для eQTL анализа остается актуальной проблема выделения действительно имеющего отношение к формированию признака SNP от детектируемых за счет неравновесия по сцеплению маркерных вариантов (Zou et al., 2019; Umans et al., 2021).

Наиболее масштабным проектом по получению транскриптомных данных и анализу eQTL является международный консорциум GTEx, в рамках которого получены данные RNA-seq для 15201 образца из 49 тканей, взятых постмортально у 838 доноров, что позволило идентифицировать 4278636 eQTL, связанных с изменением экспрессии 18262 и 5006 генов, кодирующих белки и lincRNA соответственно (GTEx Consortium, 2020).

Существуют и другие массивы eQTL, в том числе полученные не на постмортальном, а на биопсийном материале (Fairfax et al., 2014; Stolze et al., 2020). Например, в работе (Stolze et al., 2020) при анализе транскриптомов (RNA-seq) эндотелия аорты 157 доноров были установлены тысячи eQTL, не зарегистрированных в базе консорциума GTEx. В.Р. Fairfax с коллегами использовали CD14⁺ моноциты, полученные от здоровых индивидуумов, которые в условиях *in vitro* обрабатывали либо интерфероном гамма (в течение 24 ч, 367 индивидуумов), либо липополисахаридом (LPS) бактериальной стенки (2 и 24 ч, 261 и 322 индивидуума соответственно). В качестве контроля использовали CD14⁺ моноциты 414 человек. С помощью микрочипового РНК-профилирования и генотипирования было исследовано 609704 SNP (minor allele frequency, MAF > 0.04) и обнаружена 21516 eQTL, 24.6 % из которых проявляли себя в контрольных клетках, 21.6 % – после двухчасовой обработки LPS, а 25.4 и 28.3 % – после 24-часовой обработки LPS и IFN- γ соответственно. Результаты этой работы демонстрируют важную роль вариаций генома в характере транскриптомного ответа на препарат (Fairfax et al., 2014). В заключение надо отметить, что для поиска eQTL может быть взят любой массив транскриптомных данных, полученных для сотни и более индивидуумов, как, например, было сделано нами (Korbolina et al., 2021) с использованием данных RNA-seq для постмортальных образцов мозга 96 человек (Ramaker et al., 2017).

В настоящее время eQTL анализ преимущественно применяется для идентификации групп генов, вовлеченных в формирование признака (Hormozdiari et al., 2016; Morrow et al., 2018; Gamazon et al., 2019; Ratnapriya et al., 2019; Jaffe et al., 2020). Также его результаты часто используют для приоритизации обнаруженных с помощью GWAS однонуклеотидных замен с целью их последующего тщательного экспериментального изучения. В качестве примера можно привести rs13239597, расположенный в промоторе гена *TNPO3* и, согласно GWAS, ассоциированный с красной волчанкой и рассеянным склерозом. eQTL анализ транскриптомов лимфобластоидных клеточных линий, полученных от 373 индивидуумов, не выявил никакого влияния rs13239597 на экспрессию гена *TNPO3*, зато обнаружил достоверную связь аллеля А этого SNP с увеличенной экспрессией гена *IRF5*, расположенного на расстоянии 118 т. п. н. (Thynn et al., 2020), что совпадало с данными консорциума GTEx (GTEx Consortium, 2017). Анализ доступных результатов Hi-C позволил установить, что *IRF5* является одним из 12 генов, непосредственно контактирующих с участком локализации rs13239597. Компьютерное исследование мотивов, потенциально меняющих сродство к ТФ в результате нуклеотидной замены (Coetzee et al., 2015), выявило четыре таких ТФ: EVI1,

ERF, GATA1 и TAL1. С помощью ChIP-AS-qPCR было показано, что EVI1 гораздо лучше связывается с районом rs13239597 в случае А аллеля по сравнению с аллелем С (Thynn et al., 2020). Подобные примеры можно найти также в статьях (Roca-Ayats et al., 2019; Jiang et al., 2020; Tian et al., 2020).

Широкомасштабный поиск аллель-специфичных событий экспрессии (ASE) и связывания (ASB)

Развитие основанных на NGS методов транскриптомного (RNA-seq) и эпигеномного анализа (ChIP-seq, DNase-seq, ATAC-seq) открыло уникальную возможность количественного определения разницы в представленности двух аллелей (дисбаланс аллелей) каждого гетерозиготного полиморфного сайта диплоидного организма в соответствующих данных (Maurano et al., 2015; Cavalli et al., 2016a, b; Castel et al., 2020; Fan et al., 2020; Xu et al., 2020; Korbolina et al., 2021). Важная особенность этих подходов к выявлению потенциальных rSNP в сравнении с eQTL анализом и тем более с GWAS, в которых проводятся исследования SNP множества индивидуумов на фоне различного геномного контекста и условий жизни, состоит в том, что аллель-асимметричные события регистрируются для каждого индивидуума и фон для них один и тот же. Это позволяет получать достоверные данные при исследовании очень малого числа индивидуумов, вплоть до одного (Harvey et al., 2015). Увеличение выборок требуется лишь для вовлечения в анализ большего числа SNP, находящихся в гетерозиготном состоянии. Так, проведенные расчеты показывают, что использование данных, полученных для 20 индивидуумов, теоретически позволяет определить ASE или ASB события для 65–70 % SNP, имеющих частоту встречаемости в популяции ≥ 5 % (Cavalli et al., 2016a).

Благодаря этому обстоятельству существенно расширяются возможности проведения фармакогенетических и фармакогеномных исследований. Самым ярким примером такого исследования является одновременное изучение аллель-специфичных эффектов 50 различных веществ (стероидных и пептидных гормонов, нутриентов, широко применяемых лекарств и ряда загрязнителей окружающей среды) с использованием первичных культур пяти типов клеток (LCL, PBMC, HUVEC, SMC и меланоцитов), каждый из которых был представлен образцами от трех индивидуумов (Moyerbrailean et al., 2016). Анализ полученных транскриптомных данных позволил обнаружить более 300 SNP, дисбаланс в представленности аллелей которых в транскриптомах возникал или существенно возрастал в ответ на обработку тем или иным препаратом. С помощью такого же подхода были выявлены зависимые от индуктора (LPS бактериальной стенки) ASE в 19 генах иммунного ответа при анализе транскриптомов мононуклеарных клеток крови восьми индивидуумов (Edsgård et al., 2016), а также 561 ASE, реагирующих на обработку CD4⁺ Т клеток (от 24 генотипированных человек) иммунизованными антителами к CD3/CD28 (Gutierrez-Arcelus et al., 2020). Получение таких результатов открывает новый, не связанный с какой-либо первоначальной гипотезой путь к выяснению механизмов индивидуальной чувствительности к лекарственным препаратам.

Самый крупный к настоящему времени ресурс ASE, содержащий 431 миллион таких событий, получен на основе данных RNA-seq и полногеномного секвенирования (WGS) консорциума GTEx (GTEx Consortium, 2020) и опубликован в статье (Castel et al., 2020). Этот ресурс может служить источником для массового выявления rSNP, например, при сопоставлении с данными GWAS или eQTL. Важно отметить, что в случае отсутствия WGS соответствующая информация может быть получена с помощью более сложных методов биоинформатического поиска аллель-специфичных событий непосредственно в данных RNA-seq (Harvey et al., 2015; Moyerbrailean et al., 2016; Fan et al., 2020; Korbolina et al., 2021).

ChIP-seq эксперименты с использованием антител к различным ТФ позволяют напрямую регистрировать события аллель-асимметричного взаимодействия этих белков со своими сайтами в случае гетерозиготного состояния SNP в этих сайтах. В пионерной работе, выполненной в лаборатории Клауса Ваделиуса, были проанализированы все доступные на тот момент профили связывания ТФ из проекта ENCODE в линиях клеток GM12878 (В клетки), H1-hESC, K562 и SK-N-SH, что позволило обнаружить 9962 SNP, демонстрирующих аллельный дисбаланс в связывании ТФ (ASB) (Cavalli et al., 2016b). С помощью того же подхода было найдено 3713 SNP с аллельным дисбалансом в связывании ТФ в линиях клеток HepG2 и HeLa-S3. Тестирование 39 из них в репортерной системе люциферазы подтвердило влияние аллеля на экспрессию репортерного гена для 27 SNP (Cavalli et al., 2016a). Детальный анализ одного из них, rs953413, показал, что аллель А разрушает сайт связывания ТФ FOXA, в результате чего снижается связывание не только этого ТФ, но и кооперативно взаимодействующего с ним ТФ HNF4 α , что в итоге приводит к снижению экспрессии гена *ELOVL2* и может быть одним из факторов патогенеза неалкогольной жировой болезни печени (Pan et al., 2020).

Поскольку аллель-асимметричные изменения в профилях гистоновых модификаций и открытого хроматина могут являться отражением вызванных SNP изменений в связывании ТФ (Kar et al., 2014; Hatayama, Aruga, 2018; Huang et al., 2018; Yi et al., 2020), эти данные тоже широко применяются для поиска ASB. Так, в работе (Maurano et al., 2015) было проанализировано 493 профиля открытого хроматина (DNase-seq), полученного на различных линиях клеток, в которых обнаружено 64 599 SNP с ASB. Биоинформатический анализ с использованием весовых матриц для 2203 мотивов ССТФ из разных источников показал, что большинство выявленных SNP может влиять на связывание ТФ и, как следствие, на доступность соответствующих районов к DNase I (Maurano et al., 2015). Более новый метод обнаружения открытого хроматина ATAC-seq, основанный на способности гиперактивной мутантной формы транспозазы Tn5 опознавать открытые участки ДНК в составе хроматина (Marinov, Shipony, 2021), также используется для поиска ASB событий. В частности, с его помощью выявлено 53 rSNP в клеточной линии рака молочной железы MCF-7 и 125 rSNP в линии мезенхимальных стволовых клеток MSC человека. Тридцать процентов rSNP, найденных в MCF-7, и 43 % найденных в MSC идентифицируются как eQTL в данных

консорциума GTEx, что указывает на их влияние на экспрессию генов (Xu et al., 2020). Примеры использования данных ChIP-seq с антителами к гистоновым меткам для регистрации событий ASB можно найти в статьях (Sun J. et al., 2016; D'Oliveira Albanus et al., 2021; Li M. et al., 2021).

В качестве наиболее продуктивного подхода к поиску rSNP можно рассматривать сочетание поиска событий ASE и ASB. К примеру, в нашей работе (Korbolina et al., 2018) вначале были идентифицированы события ASB в данных ChIP-seq из проекта ENCODE для гистоновых модификаций H3K27ac, H3K4me1, H3K4me2, H3K4me3 и H3K27me3, а также для 456 ТФ и связанных с ними белков в линиях клеток человека K562, MCF-7 и HCT-116. Затем с использованием данных RNA-seq, полученных на тех же линиях, были выделены SNP (rSNP), связанные с изменением уровня экспрессии генов. По данным GWAS, из 1633 найденных таким образом rSNP 27 имеют связь с онкологическими заболеваниями (Korbolina et al., 2018), а 14 – с когнитивными расстройствами (Bryzgalov et al., 2018). Еще 30 rSNP удалось связать с колоректальным раком (КРР), используя данные ICGC (International Cancer Genome Consortium) (Seshagiri et al., 2012). Генотипирование больных КРР и здоровых индивидуумов по шести из них показало связь rs590352, rs4796672 и rs2072580 с данным заболеванием (Leberfarb et al., 2020). Для rs2072580 была также найдена связь с раком молочной железы (Degtyareva et al., 2020). Позднее применение этого подхода позволило выявить 14266 rSNP при обработке полученных в работе (Reyes-Palomares et al., 2020) данных профилирования гистоновой метки H3K4me3 (ChIP-seq) и RNA-seq для образцов эпителия легочной артерии 19 индивидуумов (Korbolina et al., 2021).

Массовое параллельное изучение репортерных конструкций (MPRA)

Анализ влияния аллелей полиморфных сайтов на экспрессию репортерного гена при одновременной трансфекции сотен и тысяч баркодированных плазмидных конструкций в эукариотические клетки с последующим секвенированием транскриптомов также является информативным подходом для поиска rSNP (Vockley et al., 2015; Tewhey et al., 2016; Movva et al., 2019). Наиболее масштабное исследование с помощью этого подхода было проведено в лаборатории Баса Ван Стинсела (van Arensbergen et al., 2019). С использованием беспромоторной плазмиды и фрагментированных геномов (длина фрагмента 150–500 п. н.) четырех индивидуумов, принадлежащих разным этническим группам, для каждого индивидуума были сконструированы две баркодированные библиотеки, где вставки должны были играть роль промотора. При этом авторы, основываясь на данных об инициации транскрипции в энхансерных районах (Natoli, Andrau, 2012; van Arensbergen et al., 2017), рассчитывали регистрировать не только промоторы, но и энхансеры. Использование ДНК людей генетически далеких этносов позволяло надеяться на охват максимального числа полиморфных сайтов, находящихся в гомозиготе по разным аллелям хотя бы у двух из них.

После трансфекции полученными библиотеками K562 и HepG2 клеток было найдено 19 и 14 тысяч потенциаль-

ных rSNP соответственно, большая часть из которых не перекрывалась, что еще раз свидетельствует о тканеспецифичности надгеномной (белковой) регуляторной машины. Выявленные SNP демонстрировали существенное обогащение в регуляторных районах генома. При этом обогащение было втрое выше в промоторных районах (приблизительно в 15 раз), чем в энхансерных (приблизительно в 5 раз), что, очевидно, объясняется дизайном репортерных конструкций. Для ряда rSNP с помощью масс-спектрометрического анализа белков, взаимодействующих с олигонуклеотидами, содержащими альтернативные аллели, авторам удалось идентифицировать ТФ, сайты связывания которых меняет нуклеотидная замена. В частности, показано, что аллель А rs623853 нарушает связывание ТФ семейства ELF, а аллель С rs554591 снижает связывание ZNF787, одновременно повышая связывание KLF и SP (van Arensbergen et al., 2019).

Заключение

Программы координированного включения, выключения и изменения уровня экспрессии различных генов, лежащие в основе онтогенетических событий, существования множества типов дифференцированных клеток и способности клеток реагировать на различные факторы внешней и внутренней среды, обеспечиваются регуляторной частью генома многоклеточных организмов. Информация, закодированная в регуляторных районах, транслируется в необходимый паттерн генной экспрессии прежде всего за счет связывания ТФ со специфическими последовательностями в регуляторных районах (промоторами, энхансерами, сайленсерами и т. д.) (Lan et al., 2012; Меркулова и др., 2013; Dubois-Chevalier et al., 2018; Chen, Pugh, 2021; Tobias et al., 2021). По современным представлениям, SNP в регуляторных районах генов, затрагивающие сайты связывания ТФ и изменяющие уровень генной экспрессии, играют центральную роль в вариабельности фенотипических проявлений, включая предрасположенность/устойчивость к многофакторным заболеваниям (Maurano et al., 2015; Deplancke et al., 2016; Carrasco Pro et al., 2020). С этим связан огромный интерес как к функциональной интерпретации SNP, демонстрирующих ассоциацию с различными заболеваниями (в первую очередь по данным GWAS), так и к разработке массовых функциональных подходов к поиску rSNP. Интерпретация данных GWAS проводится либо на уровне отдельных SNP, либо на массовом уровне с использованием разнообразных методов функциональной геномики (рис. 2). Параллельно те же методы функциональной геномики применяются для самостоятельного поиска rSNP, но при этом возникает необходимость решения обратной задачи – установления связи найденных rSNP с признаком (заболеванием). Самым популярным способом решения этой задачи является сопоставление полученных результатов с имеющимися данными GWAS (см. рис. 2). Однако таким образом обычно удается связать с различными признаками лишь 1.5–3 % найденных rSNP (Cavalli et al., 2016b, 2019; Korbolina et al., 2021). В этой связи весьма перспективным представляется eQTL анализ, в результате которого можно установить влияние многих конкретных rSNP на экспрессию довольно больших групп генов, последующий

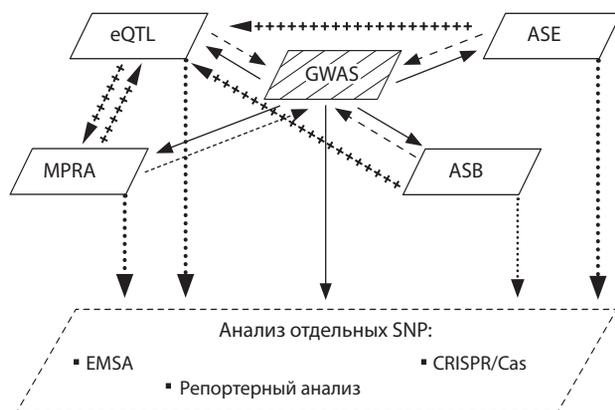


Рис. 2. Схема комплексного подхода для поиска функционально значимых для развития признака (патологии) SNP, основанного на использовании двух противоположных принципов: от связи с признаком к его функции (сплошные стрелки) и, наоборот, от функции к признаку (штриховые стрелки). Стрелки из крестиков показывают пути усиления доказательной базы функциональной значимости массивов SNP, а пунктирные – пути детального изучения отдельных SNP.

анализ которых с помощью современных инструментов функциональной аннотации (GO, Kegg и др.) позволяет «выходить» на признак (Korboлина et al., 2021).

Подводя итоги, следует заключить, что комплексное использование двух основанных на противоположных принципах подходов к поиску значимых для развития признака (патологии) SNP – с одной стороны, исходящего из данных по ассоциации SNP с неким признаком, а с другой стороны, идущего от определения аллель-специфичных изменений на молекулярном уровне (в транскриптом или регуломе) – существенно обогащает картину наших знаний о роли генетических детерминант в молекулярных механизмах формирования признаков, включая предрасположенность к многофакторным заболеваниям.

Список литературы / References

Меркулова Т.И., Ананько Е.А., Игнатъева 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.]
Antonseva E.V., Matveeva M.Y., Bondar N.P., Kashina E.V., Leberfarb E.Y., Bryzgalov L.O., Gervas P.A., Ponomareva A.A., Cherdynseva N.V., Orlov Y.L., Merkulova T.I. Regulatory single nucleotide polymorphisms at the beginning of intron 2 of the human *KRAS* gene. *J. Biosci.* 2015;40(5):873-883. DOI 10.1007/s12038-015-9567-8.
Barbeira A.N., Bonazzola R., Gamazon E.R., Liang Y., Park Y., Kim-Hellmuth S., Wang G., Jiang Z., Zhou D., Hormozdiari F., Liu B., Rao A., Hamel A.R., Pividori M.D., Aguet F., Bastarache L., Jordan D.M., Verbanck M., Do R., Stephens M., Ardlie K., McCarthy M., Montgomery S.B., Segrè A.V., Brown C.D., Lappalainen T., Wen X., Im H.K. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* 2021;22(1):49. DOI 10.1186/s13059-020-02252-4.
Boldes T., Merenbakh-Lamin K., Journo S., Shachar E., Lipson D., Yeheskel A., Pasmanik-Chor M., Rubinek T., Wolf I. R269C variant of *ESR1*: high prevalence and differential function in a subset of pancreatic cancers. *BMC Cancer*. 2020;20(1):531. DOI 10.1186/s12885-020-07005-x.

Boyle E.A., Li Y.I., Pritchard J.K. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169(7):1177-1186. DOI 10.1016/j.cell.2017.05.038.
Bryzgalov L.O., Antonseva E.V., Matveeva M.Y., Shilov A.G., Kashina E.V., Mordvinov V.A., Merkulova T.I. Detection of regulatory SNPs in human genome using ChIP-seq ENCODE data. *PLoS One*. 2013;8(10):e78833. DOI 10.1371/journal.pone.0078833.
Bryzgalov L.O., Korboлина E.E., Brusentsov I.I., Leberfarb E.Y., Bondar N.P., Merkulova T.I. Novel functional variants at the GWAS-implicated loci might confer risk to major depressive disorder, bipolar affective disorder and schizophrenia. *BMC Neurosci.* 2018; 19(S1):22. DOI 10.1186/s12868-018-0414-3.
Buniello A., MacArthur J.A.L., Cerezo M., Harris L.W., Hayhurst J., Malangone C., McMahon A., Morales J., Mountjoy E., Solliis E., Suveges D., Vrousou O., Whetzel P.L., Amode R., Guillen J.A., Riat H.S., Trevanion S.J., Hall P., Junkins H., Flicek P., Burdett T., Hindorf L.A., Cunningham F., Parkinson H. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019; 47(D1):D1005-D1012. DOI 10.1093/nar/gky1120.
Carrasco Pro S., Bulekova K., Gregor B., Labadorf A., Fuxman Bass J.I. Prediction of genome-wide effects of single nucleotide variants on transcription factor binding. *Sci. Rep.* 2020;10(1):17632. DOI 10.1038/s41598-020-74793-4.
Castel S.E., Aguet F., Mohammadi P., Aguet F., Anand S., Ardlie K.G., Gabriel S., Getz G.A., Graubert A., Hadley K., ... Moore H.M., Nierras C.R., Rao A.K., Vaught J.B., Volpi S., Ardlie K.G., Lappalainen T. A vast resource of allelic expression data spanning human tissues. *Genome Biol.* 2020;21(1):234. DOI 10.1186/s13059-020-02122-z.
Cavalli M., Pan G., Nord H., Wallén Arzt E., Wallerman O., Wadelius C. Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals. *Genomics*. 2016a;107(6):248-254. DOI 10.1016/j.ygeno.2016.04.006.
Cavalli M., Pan G., Nord H., Wallerman O., Wallén Arzt E., Berggren O., Elvers I., Eloranta M.-L., Rönnblom L., Lindblad Toh K., Wadelius C. Allele-specific transcription factor binding to common and rare variants associated with disease and gene expression. *Hum. Genet.* 2016b;135(5):485-497. DOI 10.1007/s00439-016-1654-x.
Cavalli M., Baltzer N., Umer H.M., Grau J., Lemnian I., Pan G., Wallerman O., Spalinskas R., Sahlén P., Grosse I., Komorowski J., Wadelius C. Allele specific chromatin signals, 3D interactions, and motif predictions for immune and B cell related diseases. *Sci. Rep.* 2019;9(1):2695. DOI 10.1038/s41598-019-39633-0.
Chanock S. Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis. Markers*. 2001;17(2): 89-98. DOI 10.1155/2001/858760.
Chen H., Pugh B.F. What do transcription factors interact with? *J. Mol. Biol.* 2021;433(14):166883. DOI 10.1016/j.jmb.2021.166883.
Choi J., Zhang T., Vu A., Ablain J., Makowski M.M., Colli L.M., Xu M., Hennessey R.C., Yin J., Rothschild H., Gräwe C., Kovacs M.A., Funderburk K.M., Brossard M., Taylor J., Pasaniuc B., Chari R., Chanock S.J., Hoggart C.J., Demenais F., Barrett J.H., Law M.H., Iles M.M., Yu K., Vermeulen M., Zon L.I., Brown K.M. Massively parallel reporter assays of melanoma risk variants identify *MX2* as a gene promoting melanoma. *Nat. Commun.* 2020;11(1):2718. DOI 10.1038/s41467-020-16590-1.
Claussnitzer M., Cho J.H., Collins R., Cox N.J., Dermitzakis E.T., Hurles M.E., Kathiresan S., Kenny E.E., Lindgren C.M., MacArthur D.G., North K.N., Plon S.E., Rehm H.L., Risch N., Rotimi C.N., Shendure J., Soranzo N., McCarthy M.I. A brief history of human disease genetics. *Nature*. 2020;577(7789):179-189. DOI 10.1038/s41586-019-1879-7.
Coetzee S.G., Coetzee G.A., Hazelett D.J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics*. 2015;31(23):3847-3849. DOI 10.1093/bioinformatics/btv470.
Cooper D. The human gene mutation database. *Nucleic Acids Res.* 1998;26(1):285-287. DOI 10.1093/nar/26.1.285.

- Corces M.R., Shcherbina A., Kundu S., Gloude-mans M.J., Frésard L., Granja J.M., Louie B.H., Eulalio T., Shams S., Bagdatli S.T., Mumbach M.R., Liu B., Montine K.S., Greenleaf W.J., Kundaje A., Montgomery S.B., Chang H.Y., Montine T.J. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* 2020;52(11):1158-1168. DOI 10.1038/s41588-020-00721-x.
- Couch F.J., Kuchenbaecker K.B., Michailidou K., Mendoza-Fandino G.A., Nord S., Lilyquist J., Olswold C., Hallberg E., Agata S., Ahsan H., ... Slager S., Chenevix-Trench G., Pharoah P.D.P., Monteiro A.A.N., Garcia-Closas M., Easton D.F., Antoniou A.C. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. *Nat. Commun.* 2016;7(1):11375. DOI 10.1038/ncomms11375.
- Degtyareva A.O., Leberfarb E.Y., Efimova E.G., Brusentsov I.I., Usova A.V., Lushnikova E.L., Merkulova T.I. *rs2072580T>A* polymorphism in the overlapping promoter regions of the *SART3* and *ISCU* genes associated with the risk of breast cancer. *Bull. Exp. Biol. Med.* 2020;169(1):81-84. DOI 10.1007/s10517-020-04829-2.
- Deplancke B., Alpern D., Gardeux V. The genetics of transcription factor DNA binding variation. *Cell.* 2016;166(3):538-554. DOI 10.1016/j.cell.2016.07.012.
- Dichgans M., Malik R., König I.R., Rosand J., Clarke R., Gretarsdottir S., Thorleifsson G., Mitchell B.D., Assimes T.L., Levi C., ... Willenborg C., Laaksonen R., Voight B.F., Stewart A.F.R., Rader D.J., Hall A.S., Kooner J.S. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke.* 2014;45(1):24-36. DOI 10.1161/STROKEAHA.113.002707.
- D'Oliveira Albanus R., Kyono Y., Hensley J., Varshney A., Orchard P., Kitzman J.O., Parker S.C.J. Chromatin information content landscapes inform transcription factor and DNA interactions. *Nat. Commun.* 2021;12(1):1307. DOI 10.1038/s41467-021-21534-4.
- Dubois-Chevalier J., Mazrooei P., Lupien M., Staels B., Lefebvre P., Eeckhoutte J. Organizing combinatorial transcription factor recruitment at cis-regulatory modules. *Transcription.* 2018;9(4):233-239. DOI 10.1080/21541264.2017.1394424.
- Edsgård D., Iglesias M.J., Reilly S.-J., Hamsten A., Tornvall P., Odeberg J., Emanuelsson O. GeneiASE: detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Sci. Rep.* 2016;6(1):21134. DOI 10.1038/srep21134.
- Fachal L., Aschard H., Beesley J., Barnes D.R., Allen J., Kar S., Pooley K.A., Dennis J., Michailidou K., Turman C., ... Edwards S.L., Antoniou A.C., Chenevix-Trench G., Simard J., Easton D.F., Kraft P., Dunning A.M. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat. Genet.* 2020;52(1):56-73. DOI 10.1038/s41588-019-0537-1.
- Fairfax B.P., Humburg P., Makino S., Naranbhai V., Wong D., Lau E., Jostins L., Plant K., Andrews R., McGee C., Knight J.C. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science.* 2014;343(6175):1246949. DOI 10.1126/science.1246949.
- Fan J., Hu J., Xue C., Zhang H., Susztak K., Reilly M.P., Xiao R., Li M. ASEPe: gene-based detection of allele-specific expression across individuals in a population by RNA sequencing. *PLoS Genet.* 2020;16(5):e1008786. DOI 10.1371/journal.pgen.1008786.
- Fang J., Jia J., Makowski M., Xu M., Wang Z., Zhang T., Hoskins J.W., Choi J., Han Y., Zhang M., ... Amos C.I., Iles M.M., Nathanson K.L., Landi M.T., Vermeulen M., Brown K.M., Amundadottir L.T. Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of *TERT* by ZNF148. *Nat. Commun.* 2017;8(1):15034. DOI 10.1038/ncomms15034.
- Farh K.K.-H., Marson A., Zhu J., Kleinewietfeld M., Housley W.J., Beik S., Shores N., Whitton H., Ryan R.J.H., Shishkin A.A., Hatan M., Carrasco-Alfonso M.J., Mayer D., Luckey C.J., Patsopoulos N.A., De Jager P.L., Kuchroo V.K., Epstein C.B., Daly M.J., Hafler D.A., Bernstein B.E. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature.* 2015;518(7539):337-343. DOI 10.1038/nature13835.
- Fornes O., Castro-Mondragon J.A., Khan A., van der Lee R., Zhang X., Richmond P.A., Modi B.P., Correard S., Gheorghe M., Baranašić D., Santana-Garcia W., Tan G., Chèneby J., Ballester B., Parcy F., Sandelin A., Lenhard B., Wasserman W.W., Mathelier A. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2020;48(D1):D87-D92. DOI 10.1093/nar/gkz1001.
- Gamazon E.R., Segre A.V., van de Bunt M., Wen X., Xi H.S., Hormozdiari F., Ongen H., Konkashbaev A., Derks E.M., Aguet F., Quan J., Nicolae D.L., Eskin E., Kellis M., Getz G., McCarthy M.I., Dermitzakis E.T., Cox N.J., Ardlie K.G. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* 2018;50(7):956-967. DOI 10.1038/s41588-018-0154-4.
- Gamazon E.R., Zwinderman A.H., Cox N.J., Denys D., Derks E.M. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat. Genet.* 2019;51(6):933-940. DOI 10.1038/s41588-019-0409-8.
- Gao P., Xia J.-H., Sipeky C., Dong X.-M., Zhang Q., Yang Y., Zhang P., Cruz S.P., Zhang K., Zhu J., Lee H.-M., Suleman S., Giannareas N., Liu S., Tammela T.L.J., Auvinen A., Wang X., Huang Q., Wang Liguang, Manninen A., Vaarala M.H., Wang Liang, Schleutker J., Wei G.-H. Biology and clinical implications of the 19q13 aggressive prostate cancer susceptibility locus. *Cell.* 2018;174(3):576-589.e18. DOI 10.1016/j.cell.2018.06.003.
- Goldstein D.B. Common genetic variation and human traits. *N. Engl. J. Med.* 2009;360(17):1696-1698. DOI 10.1056/NEJMp0806284.
- Gong J., Qiu C., Huang D., Zhang Y., Yu S., Zeng C. Integrative functional analysis of super enhancer SNPs for coronary artery disease. *J. Hum. Genet.* 2018;63(5):627-638. DOI 10.1038/s10038-018-0422-2.
- Göring H.H.H., Curran J.E., Johnson M.P., Dyer T.D., Charlesworth J., Cole S.A., Jowett J.B.M., Abraham L.J., Rainwater D.L., Comuzie A.G., Mahaney M.C., Almasy L., MacCluer J.W., Kissebah A.H., Collier G.R., Moses E.K., Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* 2007;39(10):1208-1216. DOI 10.1038/ng2119.
- GTEX Consortium. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204-213. DOI 10.1038/nature24277.
- 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 L., Du Y., Qu S., Wang J. rVarBase: an updated database for regulatory features of human variants. *Nucleic Acids Res.* 2016;44(D1):D888-D893. DOI 10.1093/nar/gkv1107.
- Gutierrez-Arcelus M., Baglaenko Y., Arora J., Hannes S., Luo Y., Amariuta T., Teslovich N., Rao D.A., Ermann J., Jonsson A.H., Navarrete C., Rich S.S., Taylor K.D., Rotter J.I., Gregersen P.K., Esko T., Brenner M.B., Raychaudhuri S. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *Nat. Genet.* 2020;52(3):247-253. DOI 10.1038/s41588-020-0579-4.
- Harvey C.T., Moyerbrailean G.A., Davis G.O., Wen X., Luca F., Pique-Regi R. QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics.* 2015;31(8):1235-1242. DOI 10.1093/bioinformatics/btu802.
- Hatayama M., Aruga J. Role of Zic family proteins in transcriptional regulation and chromatin remodeling. In: Aruga J. (Ed.) Zic Family. Advances in Experimental Medicine and Biology. Vol. 1046. Singapore: Springer, 2018;353-380. DOI 10.1007/978-981-10-7311-3_18.
- Hindorf L.A., Sethupathy P., Junkins H.A., Ramos E.M., Mehta J.P., Collins F.S., Manolio T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA.* 2009;106(23):9362-9367. DOI 10.1073/pnas.0903103106.

- Hormozdiari F., van de Bunt M., Segrè A.V., Li X., Joo J.W.J., Bilow M., Sul J.H., Sankararaman S., Pasaniuc B., Eskin E. Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* 2016;99(6):1245-1260. DOI 10.1016/j.ajhg.2016.10.003.
- Huang Q., Ma C., Chen L., Luo D., Chen R., Liang F. Mechanistic insights into the interaction between transcription factors and epigenetic modifications and the contribution to the development of obesity. *Front. Endocrinol. (Lausanne)*. 2018;9:370. DOI 10.3389/fendo.2018.00370.
- Jaffe A.E., Hoepfner D.J., Saito T., Blanpain L., Ukaigwe J., Burke E.E., Collado-Torres L., Tao R., Tajinda K., Maynard K.R., Tran M.N., Martinowich K., Deep-Soboslay A., Shin J.H., Kleinman J.E., Weinberger D.R., Matsumoto M., Hyde T.M. Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. *Nat. Neurosci.* 2020;23(4):510-519. DOI 10.1038/s41593-020-0604-z.
- Jiang Z., Huang Y., Zhang P., Han C., Lu Y., Mo Z., Zhang Z., Li X., Zhao S., Cai F., Huang L., Chen C., Shi Z., Zhang Y., Ling F. Characterization of a pathogenic variant in GBA for Parkinson's disease with mild cognitive impairment patients. *Mol. Brain.* 2020;13(1):102. DOI 10.1186/s13041-020-00637-x.
- Jones M.R., Peng P.-C., Coetzee S.G., Tyrer J., Reyes A.L.P., Corona R.I., Davis B., Chen S., Dezem F., Seo J.-H., Kar S., Dareng E., Berman B.P., Freedman M.L., Plummer J.T., Lawrenson K., Pharoah P., Hazelett D.J., Gayther S.A. Ovarian cancer risk variants are enriched in histotype-specific enhancers and disrupt transcription factor binding sites. *Am. J. Hum. Genet.* 2020;107(4):622-635. DOI 10.1016/j.ajhg.2020.08.021.
- Kalita C.A., Brown C.D., Freiman A., Isherwood J., Wen X., Pique-Regi R., Luca F. High-throughput characterization of genetic effects on DNA-protein binding and gene transcription. *Genome Res.* 2018;28(11):1701-1708. DOI 10.1101/gr.237354.118.
- Kar S., Parbin S., Deb M., Shilpi A., Sengupta D., Rath S.K., Rakshit M., Patra A., Patra S.K. Epigenetic choreography of stem cells: the DNA demethylation episode of development. *Cell. Mol. Life Sci.* 2014;71(6):1017-1032. DOI 10.1007/s00018-013-1482-2.
- Klein J.C., Keith A., Rice S.J., Shepherd C., Agarwal V., Loughlin J., Shendure J. Functional testing of thousands of osteoarthritis-associated variants for regulatory activity. *Nat. Commun.* 2019;10(1):2434. DOI 10.1038/s41467-019-10439-y.
- Knight J.C., Udalova I., Hill A.V.S., Greenwood B.M., Peshu N., Marsh K., Kwiatkowski D. A polymorphism that affects OCT-1 binding to the *TNF* promoter region is associated with severe malaria. *Nat. Genet.* 1999;22(2):145-150. DOI 10.1038/9649.
- Korboлина E.E., Brusentsov I.I., Bryzgalov L.O., Leberfarb E.Y., Degtyareva A.O., Merkulova T.I. Novel approach to functional SNPs discovery from genome-wide data reveals promising variants for colon cancer risk. *Hum. Mutat.* 2018;39(6):851-859. DOI 10.1002/humu.23425.
- Korboлина E.E., Bryzgalov L.O., Ustrokhanova D.Z., Postovalov S.N., Poverin D.V., Damarov I.S., Merkulova T.I. A panel of rSNPs demonstrating allelic asymmetry in both ChIP-seq and RNA-seq data and the search for their phenotypic outcomes through analysis of DEGs. *Int. J. Mol. Sci.* 2021;22(14):7240. DOI 10.3390/ijms22147240.
- Krause M.D., Huang R.-T., Wu D., Shentu T.-P., Harrison D.L., Whalen M.B., Stolze L.K., Di Rienzo A., Moskowitz I.P., Civelek M., Romanoski C.E., Fang Y. Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. *Proc. Natl. Acad. Sci. USA.* 2018;115(48):E11349-E11358. DOI 10.1073/pnas.1810568115.
- Lan X., Farnham P.J., Jin V.X. Uncovering transcription factor modules using one- and three-dimensional analyses. *J. Biol. Chem.* 2012;287(37):30914-30921. DOI 10.1074/jbc.R111.309229.
- Lander E.S., Schork N.J. Genetic dissection of complex traits. *Science.* 1994;265(5181):2037-2048. DOI 10.1126/science.8091226.
- Lappalainen T. Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Res.* 2015;25(10):1427-1431. DOI 10.1101/gr.190983.115.
- Leberfarb E.Y., Degtyareva A.O., Brusentsov I.I., Maximov V.N., Voevodova M.I., Autenshlus A.I., Morozov D.V., Sokolov A.V., Merkulova T.I. Potential regulatory SNPs in the *ATXN7L3B* and *KRT15* genes are associated with gender-specific colorectal cancer risk. *Per. Med.* 2020;17(1):43-54. DOI 10.2217/pme-2019-0059.
- Li M., Huang H., Li L., He C., Zhu L., Guo H., Wang L., Liu Jiafeng, Wu S., Liu Jingxin, Xu T., Mao Z., Cao N., Zhang K., Lan F., Ding J., Yuan J., Liu Y., Ouyang H. Core transcription regulatory circuitry orchestrates corneal epithelial homeostasis. *Nat. Commun.* 2021;12(1):420. DOI 10.1038/s41467-020-20713-z.
- Li S., Li Y., Li X., Liu J., Huo Y., Wang J., Liu Z., Li M., Luo X.-J. Regulatory mechanisms of major depressive disorder risk variants. *Mol. Psychiatry.* 2020;25(9):1926-1945. DOI 10.1038/s41380-020-0715-7.
- Li X.-X., Peng T., Gao J., Feng J.-G., Wu D.-D., Yang T., Zhong L., Fu W.-P., Sun C. Allele-specific expression identified rs2509956 as a novel long-distance *cis*-regulatory SNP for *SCGB1A1*, an important gene for multiple pulmonary diseases. *Am. J. Physiol. Lung. Cell. Mol. Physiol.* 2019;317(4):L456-L463. DOI 10.1152/ajplung.00275.2018.
- Liu D., Qin S., Ray B., Kalari K.R., Wang L., Weinsilboum R.M. Single nucleotide polymorphisms (SNPs) distant from xenobiotic response elements can modulate Aryl hydrocarbon receptor function: SNP-dependent CYP1A1 induction. *Drug Metab. Dispos.* 2018;46(9):1372-1381. DOI 10.1124/dmd.118.082164.
- Liu S., Liu Y., Zhang Q., Wu J., Liang J., Yu S., Wei G.-H., White K.P., Wang X. Systematic identification of regulatory variants associated with cancer risk. *Genome Biol.* 2017;18(1):194. DOI 10.1186/s13059-017-1322-z.
- Lu X., Chen X., Forney C., Donmez O., Miller D., Parameswaran S., Hong T., Huang Y., Pujato M., Cazares T., Miraldi E.R., Ray J.P., de Boer C.G., Harley J.B., Weirauch M.T., Kottyan L.C. Global discovery of lupus genetic risk variant allelic enhancer activity. *Nat. Commun.* 2021;12(1):1611. DOI 10.1038/s41467-021-21854-5.
- Ludlow L.B., Schick B.P., Budarf M.L., Driscoll D.A., Zackai E.H., Cohen A., Konkle B.A. Identification of a mutation in a GATA binding site of the platelet glycoprotein Ibb promoter resulting in the Bernard-Soulier syndrome. *J. Biol. Chem.* 1996;271(36):22076-22080. DOI 10.1074/jbc.271.36.22076.
- Malecová B., Morris K.V. Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs. *Curr. Opin. Mol. Ther.* 2010;12(2):214-222.
- Marinov G.K., Shipony Z. Interrogating the accessible chromatin landscape of eukaryote genomes using ATAC-seq. In: Shomron N. (Ed.) *Deep Sequencing Data Analysis. Methods in Molecular Biology.* Vol. 2243. New York: Humana, 2021;183-226. DOI 10.1007/978-1-0716-1103-6_10.
- Maurano M.T., Humbert R., Rynes E., Thurman R.E., Haugen E., Wang H., Reynolds A.P., Sandstrom R., Qu H., Brody J., Shafer A., Neri F., Lee K., Kutayavin T., Stehling-Sun S., Johnson A.K., Canfield T.K., Giste E., Diegel M., Bates D., Hansen R.S., Neph S., Sabo P.J., Heimfeld S., Raubitschek A., Ziegler S., Cotsapas C., Sotoodehnia N., Glass I., Sunyaev S.R., Kaul R., Stamatoypoulos J.A. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337(6099):1190-1195. DOI 10.1126/science.1222794.
- Maurano M.T., Haugen E., Sandstrom R., Vierstra J., Shafer A., Kaul R., Stamatoypoulos J.A. Large-scale identification of sequence variants influencing human transcription factor occupancy *in vivo*. *Nat. Genet.* 2015;47(12):1393-1401. DOI 10.1038/ng.3432.
- Moore J.E., Purcaro M.J., Pratt H.E., Epstein C.B., Shores N., Adrian J., Kawli T., Davis C.A., Dobin A., Kaul R., ... Snyder M.P., Bernstein B.E., Wold B., Hardison R.C., Gingeras T.R., Stamatoypoulos J.A., Weng Z. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature.* 2020;583(7818):699-710. DOI 10.1038/s41586-020-2493-4.
- Morrow J.D., Cho M.H., Platig J., Zhou X., DeMeo D.L., Qiu W., Celli B., Marchetti N., Criner G.J., Bueno R., Washko G.R., Glass K., Quackenbush J., Silverman E.K., Hersh C.P. Ensemble genomic

- analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Hum. Genomics*. 2018;12(1):1. DOI 10.1186/s40246-018-0132-z.
- Movva R., Greenside P., Marinov G.K., Nair S., Shrikumar A., Kundaje A. Deciphering regulatory DNA sequences and noncoding genetic variants using neural network models of massively parallel reporter assays. *PLoS One*. 2019;14(6):e0218073. DOI 10.1371/journal.pone.0218073.
- Moyerbrailean G.A., Richards A.L., Kurtz D., Kalita C.A., Davis G.O., Harvey C.T., Alazizi A., Watzka D., Sorokin Y., Hauff N., Zhou X., Wen X., Pique-Regi R., Luca F. High-throughput allele-specific expression across 250 environmental conditions. *Genome Res*. 2016;26(12):1627-1638. DOI 10.1101/gr.209759.116.
- Nasser J., Bergman D.T., Fulco C.P., Guckelberger P., Doughty B.R., Patwardhan T.A., Jones T.R., Nguyen T.H., Ulirsch J.C., Lekschas F., Muallim K., Natri H.M., Weeks E.M., Munson G., Kane M., Kang H.Y., Cui A., Ray J.P., Eisenhaure T.M., Collins R.L., Dey K., Pfister H., Price A.L., Epstein C.B., Kundaje A., Xavier R.J., Daly M.J., Huang H., Finucane H.K., Hacohen N., Lander E.S., Engreitz J.M. Genome-wide enhancer maps link risk variants to disease genes. *Nature*. 2021;593(7858):238-243. DOI 10.1038/s41586-021-03446-x.
- Natoli G., Andrau J.-C. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* 2012;46(1):1-19. DOI 10.1146/annurev-genet-110711-155459.
- Pan G., Cavalli M., Carlsson B., Skrtic S., Kumar C., Wadelius C. rs953413 regulates polyunsaturated fatty acid metabolism by modulating *ELOVL2* expression. *iScience*. 2020;23(2):100808. DOI 10.1016/j.isci.2019.100808.
- Panchatcharam M., Salous A.K., Brandon J., Miriyala S., Wheeler J., Patil P., Sunkara M., Morris A.J., Escalante-Alcalde D., Smyth S.S. Mice with targeted inactivation of *Ppap2b* in endothelial and hematopoietic cells display enhanced vascular inflammation and permeability. *Arterioscler. Thromb. Vasc. Biol.* 2014;34(4):837-845. DOI 10.1161/ATVBAHA.113.302335.
- Peng T., Zhong L., Gao J., Wan Z., Fu W.-P., Sun C. Identification of rs11615992 as a novel regulatory SNP for human *P2RX7* by allele-specific expression. *Mol. Genet. Genomics*. 2020;295(1):23-30. DOI 10.1007/s00438-019-01598-0.
- Piedrafita F.J., Molander R.B., Vansant G., Orlova E.A., Pfahl M., Reynolds W.F. An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element. *J. Biol. Chem.* 1996;271(24):14412-14420. DOI 10.1074/jbc.271.24.14412.
- Prestel M., Prell-Schicker C., Webb T., Malik R., Lindner B., Ziesch N., Rex-Haffner M., Röh S., Vitorawong T., Lehm M., Mokry M., den Ruijter H., Haitjema S., Asare Y., Söllner F., Najafabadi M.G., Aherrahrou R., Civelek M., Samani N.J., Mann M., Haffner C., Dichgans M. The atherosclerosis risk variant rs2107595 mediates allele-specific transcriptional regulation of *HDAC9* via E2F3 and Rb1. *Stroke*. 2019;50(10):2651-2660. DOI 10.1161/STROKEAHA.119.026112.
- Protze J., Naas S., Krüger R., Stöhr C., Kraus A., Grampp S., Wiesener M., Schiffer M., Hartmann A., Wullich B., Schödel J. The renal cancer risk allele at 14q24.2 activates a novel hypoxia-inducible transcription factor-binding enhancer of *DPF3* expression. *J. Biol. Chem.* 2022;298(3):101699. DOI 10.1016/j.jbc.2022.101699.
- Ramaker R.C., Bowling K.M., Lasseigne B.N., Hagenauer M.H., Hardigan A.A., Davis N.S., Gertz J., Cartagena P.M., Walsh D.M., Vawter M.P., Jones E.G., Schatzberg A.F., Barchas J.D., Watson S.J., Bunney B.G., Akil H., Bunney W.E., Li J.Z., Cooper S.J., Myers R.M. Post-mortem molecular profiling of three psychiatric disorders. *Genome Med.* 2017;9(1):72. DOI 10.1186/s13073-017-0458-5.
- Ratnapriya R., Sosina O.A., Starostik M.R., Kwicklis M., Kaphahn R.J., Fritsche L.G., Walton A., Arvanitis M., Gieser L., Pietraszkiewicz A., Montezuma S.R., Chew E.Y., Battle A., Abecasis G.R., Ferrington D.A., Chatterjee N., Swaroop A. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* 2019;51(4):606-610. DOI 10.1038/s41588-019-0351-9.
- Reyes-Palomares A., Gu M., Grubert F., Berest I., Sa S., Kasowski M., Arnold C., Shuai M., Srivas R., Miao S., Li D., Snyder M.P., Rabonovitch M., Zaugg J.B. Remodeling of active endothelial enhancers is associated with aberrant gene-regulatory networks in pulmonary arterial hypertension. *Nat. Commun.* 2020;11(1):1673. DOI 10.1038/s41467-020-15463-x.
- Ring H.Z., Kroetz D.L. Candidate gene approach for pharmacogenetic studies. *Pharmacogenomics*. 2002;3(1):47-56. DOI 10.1517/14622416.3.1.47.
- Roca-Ayats N., Martínez-Gil N., Cozar M., Gerousi M., Garcia-Giralt N., Ovejero D., Mellibovsky L., Nogués X., Díez-Pérez A., Grinberg D., Balcells S. Functional characterization of the *C7ORF76* genomic region, a prominent GWAS signal for osteoporosis in 7q21.3. *Bone*. 2019;123:39-47. DOI 10.1016/j.bone.2019.03.014.
- Scelo G., Purdie M.P., Brown K.M., Johansson M., Wang Z., Eckel-Passow J.E., Ye Y., Hofmann J.N., Choi J., Foll M., ... Deleuze J.-F., McKay J.D., Parker A.S., Wu X., Houlston R.S., Brennan P., Chagnock S.J. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. *Nat. Commun.* 2017;8(1):15724. DOI 10.1038/ncomms15724.
- Schaid D.J., Chen W., Larson N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* 2018;19(8):491-504. DOI 10.1038/s41576-018-0016-z.
- Seshagiri S., Stawiski E.W., Durinck S., Modrusan Z., Storm E.E., Conboy C.B., Chaudhuri S., Guan Y., Janakiraman V., Jaiswal B.S., Guillory J., Ha C., Dijkgraaf G.J.P., Stinson J., Gnad F., Huntley M.A., Degenhardt J.D., Haverty P.M., Bourgon R., Wang W., Koepfen H., Gentleman R., Starr T.K., Zhang Z., Largaespada D.A., Wu T.D., de Sauvage F.J. Recurrent R-spondin fusions in colon cancer. *Nature*. 2012;488(7413):660-664. DOI 10.1038/nature11282.
- Sherry S.T., Ward M.H., Kholodov M., Baker J., Phan L., Smigielski E.M., Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308-311. DOI 10.1093/nar/29.1.308.
- Stolze L.K., Conklin A.C., Whalen M.B., López Rodríguez M., Öunap K., Selvarajan I., Toropainen A., Örd T., Li J., Eshghi A., Solomon A.E., Fang Y., Kaikkonen M.U., Romanoski C.E. Systems genetics in human endothelial cells identifies non-coding variants modifying enhancers, expression, and complex disease traits. *Am. J. Hum. Genet.* 2020;106(6):748-763. DOI 10.1016/j.ajhg.2020.04.008.
- Sun J., Zhao Y., McGreal R., Cohen-Tayar Y., Rockowitz S., Wilczek C., Ashery-Padan R., Shechter D., Zheng D., Cvekl A. Pax6 associates with H3K4-specific histone methyltransferases Mll1, Mll2, and Set1a and regulates H3K4 methylation at promoters and enhancers. *Epigenetics Chromatin*. 2016;9(1):37. DOI 10.1186/s13072-016-0087-z.
- Sun W., Yao S., Tang J., Liu S., Chen J., Deng D., Zeng C. Integrative analysis of super enhancer SNPs for type 2 diabetes. *PLoS One*. 2018;13(1):e0192105. DOI 10.1371/journal.pone.0192105.
- Syddall C.M., Reynard L.N., Young D.A., Loughlin J. The identification of *trans*-acting factors that regulate the expression of *GDF5* via the osteoarthritis susceptibility SNP rs143383. *PLoS Genet.* 2013;9(6):e1003557. DOI 10.1371/journal.pgen.1003557.
- Tak Y.G., Farnham P.J. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin*. 2015;8:57. DOI 10.1186/s13072-015-0050-4.
- Tam V., Patel N., Turcotte M., Bossé Y., Paré G., Meyre D. Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 2019;20(8):467-484. DOI 10.1038/s41576-019-0127-1.
- Tewhey R., Kotliar D., Park D.S., Liu B., Winnicki S., Reilly S.K., Andersen K.G., Mikkelsen T.S., Lander E.S., Schaffner S.F., Sabeti P.C. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*. 2016;165(6):1519-1529. DOI 10.1016/j.cell.2016.04.027.
- Thynn H.N., Chen X.-F., Hu W.-X., Duan Y.-Y., Zhu D.-L., Chen H., Wang N.-N., Chen H.-H., Rong Y., Lu B.-J., Yang M., Jiang F.,

- Dong S.-S., Guo Y., Yang T.-L. An allele-specific functional SNP associated with two systemic autoimmune diseases modulates *IRF5* expression by long-range chromatin loop formation. *J. Invest. Dermatol.* 2020;140(2):348-360.e11. DOI 10.1016/j.jid.2019.06.147.
- Tian J., Lou J., Cai Y., Rao M., Lu Z., Zhu Y., Zou D., Peng X., Wang H., Zhang M., Niu S., Li Y., Zhong R., Chang J., Miao X. Risk SNP-mediated enhancer-promoter interaction drives colorectal cancer through both *FADS2* and *AP002754.2*. *Cancer Res.* 2020; 80(9):1804-1818. DOI 10.1158/0008-5472.CAN-19-2389.
- Tobias I.C., Abatti L.E., Moorthy S.D., Mullany S., Taylor T., Khader N., Filice M.A., Mitchell J.A. Transcriptional enhancers: from prediction to functional assessment on a genome-wide scale. *Genome.* 2021;64(4):426-448. DOI 10.1139/gen-2020-0104.
- Ulirsch J.C., Nandakumar S.K., Wang L., Giani F.C., Zhang X., Rogov P., Melnikov A., McDonel P., Do R., Mikkelsen T.S., Sankaran V.G. Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell.* 2016;165(6):1530-1545. DOI 10.1016/j.cell.2016.04.048.
- Umans B.D., Battle A., Gilad Y. Where are the disease-associated eQTLs? *Trends Genet.* 2021;37(2):109-124. DOI 10.1016/j.tig.2020.08.009.
- van Arensbergen J., FitzPatrick V.D., de Haas M., Pagie L., Sluimer J., Bussemaker H.J., van Steensel B. Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol.* 2017;35(2):145-153. DOI 10.1038/nbt.3754.
- van Arensbergen J., Pagie L., FitzPatrick V.D., de Haas M., Baltissen M.P., Comoglio F., van der Weide R.H., Teunissen H., Vösa U., Franke L., de Wit E., Vermeulen M., Bussemaker H.J., van Steensel B. High-throughput identification of human SNPs affecting regulatory element activity. *Nat. Genet.* 2019;51(7):1160-1169. DOI 10.1038/s41588-019-0455-2.
- Vasiliev G.V., Merkulov V.M., Kobzev V.F., Merkulova T.I., Ponomarenko M.P., Kolchanov N.A. Point mutations within 663–666 bp of intron 6 of the human *TD02* gene, associated with a number of psychiatric disorders, damage the YY-1 transcription factor binding site. *FEBS Lett.* 1999;462(1-2):85-88. DOI 10.1016/S0014-5793(99)01513-6.
- Visscher P.M., Brown M.A., McCarthy M.I., Yang J. Five years of GWAS discovery. *Am. J. Hum. Genet.* 2012;90(1):7-24. DOI 10.1016/j.ajhg.2011.11.029.
- Vockley C.M., Guo C., Majoros W.H., Nodzenski M., Scholtens D.M., Hayes M.G., Lowe W.L., Reddy T.E. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. *Genome Res.* 2015;25(8):1206-1214. DOI 10.1101/gr.190090.115.
- Wang X., Hayes J.E., Xu X., Gao X., Mehta D., Lilja H.G., Klein R.J. Validation of prostate cancer risk variants rs10993994 and rs7098889 by CRISPR/Cas9 mediated genome editing. *Gene.* 2021; 768:145265. DOI 10.1016/j.gene.2020.145265.
- Wang Y., Ma R., Liu B., Kong J., Lin H., Yu X., Wang R., Li L., Gao M., Zhou B., Mohan M., Yu H., Hou Z., Shen H., Qian B. SNP rs17079281 decreases lung cancer risk through creating an YY1-binding site to suppress *DCBLD1* expression. *Oncogene.* 2020; 39(20):4092-4102. DOI 10.1038/s41388-020-1278-4.
- Werling D.M., Pochareddy S., Choi J., An J.-Y., Sheppard B., Peng M., Li Z., Dastmalchi C., Santpere G., Sousa A.M.M., Tebbenkamp A.T.N., Kaur N., Gulden F.O., Breen M.S., Liang L., Gilson M.C., Zhao X., Dong S., Klei L., Cicek A.E., Buxbaum J.D., Adle-Biassette H., Thomas J.-L., Aldinger K.A., O'Day D.R., Glass I.A., Zaitlen N.A., Talkowski M.E., Roeder K., State M.W., Devlin B., Sanders S.J., Sestan N. Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex. *Cell Rep.* 2020;31(1):107489. DOI 10.1016/j.celrep.2020.03.053.
- Westra H.-J., Franke L. From genome to function by studying eQTLs. *Biochim. Biophys. Acta.* 2014;1842(10):1896-1902. DOI 10.1016/j.bbadis.2014.04.024.
- Westra H.-J., Peters M.J., Esko T., Yaghootkar H., Schurmann C., Kettunen J., Christiansen M.W., Fairfax B.P., Schramm K., Powell J.E., ... Psaty B.M., Ripatti S., Teumer A., Frayling T.M., Metspalu A., van Meurs J.B.J., Franke L. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 2013;45(10):1238-1243. DOI 10.1038/ng.2756.
- Whittington T., Gao P., Song W., Ross-Adams H., Lamb A.D., Yang Y., Svezia I., Klevebring D., Mills I.G., Karlsson R., Halim S., Dunning M.J., Egevad L., Warren A.Y., Neal D.E., Grönberg H., Lindberg J., Wei G.-H., Wiklund F. Gene regulatory mechanisms underpinning prostate cancer susceptibility. *Nat. Genet.* 2016;48(4): 387-397. DOI 10.1038/ng.3523.
- Wu C., Huang R.-T., Kuo C.-H., Kumar S., Kim C.W., Lin Y.-C., Chen Y.-J., Birukova A., Birukov K.G., Dulin N.O., Civelek M., Lusic A.J., Loyer X., Tedgui A., Dai G., Jo H., Fang Y. Mechano-sensitive PPAP2B regulates endothelial responses to atherorelevant hemodynamic forces. *Circ. Res.* 2015;117(4):e41-e53. DOI 10.1161/CIRCRESAHA.117.306457.
- Xu S., Feng W., Lu Z., Yu C.Y., Shao W., Nakshatri H., Reiter J.L., Gao H., Chu X., Wang Y., Liu Y. regSNPs-ASB: a computational framework for identifying allele-specific transcription factor binding from ATAC-seq data. *Front. Bioeng. Biotechnol.* 2020;8:886. DOI 10.3389/fbioe.2020.00886.
- Yan J., Qiu Y., Ribeiro dos Santos A.M., Yin Y., Li Y.E., Vinckier N., Nariai N., Benaglio P., Raman A., Li X., Fan S., Chiou J., Chen F., Frazer K.A., Gaulton K.J., Sander M., Taipale J., Ren B. Systematic analysis of binding of transcription factors to noncoding variants. *Nature.* 2021;591(7848):147-151. DOI 10.1038/s41586-021-03211-0.
- Yi M., Tan Y., Wang L., Cai J., Li Xiaoling, Zeng Z., Xiong W., Li G., Li Xiayu, Tan P., Xiang B. TP63 links chromatin remodeling and enhancer reprogramming to epidermal differentiation and squamous cell carcinoma development. *Cell. Mol. Life Sci.* 2020;77(21):4325-4346. DOI 10.1007/s00018-020-03539-2.
- Zhang P., Xia J.-H., Zhu J., Gao P., Tian Y.-J., Du M., Guo Y.-C., Suleman S., Zhang Q., Kohli M., Tillmans L.S., Thibodeau S.N., French A.J., Cerhan J.R., Wang L.-D., Wei G.-H., Wang L. High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. *Nat. Commun.* 2018;9(1):2022. DOI 10.1038/s41467-018-04451-x.
- Zhao Y., Wu D., Jiang D., Zhang X., Wu T., Cui J., Qian M., Zhao J., Oesterreich S., Sun W., Finkel T., Li G. A sequential methodology for the rapid identification and characterization of breast cancer-associated functional SNPs. *Nat. Commun.* 2020;11(1):3340. DOI 10.1038/s41467-020-17159-8.
- Zheng R., Wan C., Mei S., Qin Q., Wu Q., Sun H., Chen C.-H., Brown M., Zhang X., Meyer C.A., Liu X.S. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* 2019;47(D1):D729-D735. DOI 10.1093/nar/gky1094.
- Zou J., Hormozdiari F., Jew B., Castel S.E., Lappalainen T., Ernst J., Sul J.H., Eskin E. Leveraging allelic imbalance to refine fine-mapping for eQTL studies. *PLoS Genet.* 2019;15(12):e1008481. DOI 10.1371/journal.pgen.1008481.

ORCID ID

E.V. Antontseva orcid.org/0000-0002-4214-7153
A.O. Degtyareva orcid.org/0000-0001-8586-2256

E.E. Korbolina orcid.org/0000-0002-7460-5892
I.S. Damarov orcid.org/0000-0002-3883-3054
T.I. Merkulova orcid.org/0000-0002-2707-0127

Благодарности. Работа выполнена при финансовой поддержке Российского научного фонда, грант № 23-15-00113.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 17.01.2023. После доработки 24.03.2023. Принята к публикации 30.03.2023.