



На правах рукописи

Дамаров Игорь Сергеевич

**Полногеномное выявление регуляторных SNPs
человека на основе получения и анализа комплекса
данных ChIP-seq и RNA-seq**

1.5.7. – генетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата биологических наук

Новосибирск - 2025

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН) в лаборатории регуляции экспрессии генов, г.Новосибирск.

Научный руководитель: **Корболина Елена Евгеньевна** к.б.н., н.с. лаборатории регуляции экспрессии генов, Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики СО РАН», г. Новосибирск

Официальные оппоненты: **Миронова Надежда Львовна** д.б.н., в.н.с. лаборатории биохимии нуклеиновых кислот, Федеральное государственное бюджетное учреждение науки Институт химической биологии и фундаментальной медицины СО РАН, г. Новосибирск

Кулаковский Иван Владимирович к.ф.-м.н., д.б.н., в.н.с. группы регуляции биосинтеза белка, Федеральное государственное бюджетное учреждение науки Институт белка РАН, г. Пущино.

Ведущая организация: Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр фундаментальной и трансляционной медицины», г. Новосибирск

Защита диссертации состоится «___»_____20__г. на утреннем заседании Диссертационного совета 24.1.239.01 на базе ФГБНУ «ФИЦ Институт цитологии и генетики СО РАН» в конференц-зале Института по адресу: 630090, г.Новосибирск, проспект ак. Лаврентьева, 10, т.(383)363-49-06, факс(383)333-12-78,

e-mail: dissov@bionet.nsc.ru.

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института <http://www.icgbio.ru/>

Автореферат разослан «___»_____20__г.

Ученый секретарь
диссертационного совета,
доктор биологических наук

Т.М.Хлебодарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Одной из основных задач современной генетики человека является обнаружение вариаций генома, связанных с формированием признаков, и выяснение молекулярных механизмов, за счет которых эти вариации приводят к фенотипическим различиям, включая предрасположенность к заболеваниям и индивидуальную чувствительность к лекарственным препаратам. К настоящему времени с помощью полногеномных исследований ассоциаций (Genome-Wide Association Studies, GWAS) найдено более трехсот тысяч вариаций (в основном однонуклеотидных замен - SNPs), связанных с различными фенотипическими признаками и заболеваниями человека (URL: <https://www.ebi.ac.uk/gwas>; Sollis et al., 2023). Подавляющее число таких SNPs расположено в некодирующей части генома, что предполагает их регуляторную функцию. В частности, нуклеотидные замены в полиморфных позициях, расположенных в промоторных и энхансерных районах генов (rSNPs), могут оказывать влияние на сайты связывания транскрипционных факторов (ТФ), способствуя их разрушению или созданию новых сайтов, а также увеличению или уменьшению эффективности связывания различных ТФ, что приводит к изменению уровня экспрессии соответствующих генов (Degtyareva et al., 2021; Fabo, Khavari, 2023). Однако технология GWAS не дает никакой информации о молекулярных механизмах, опосредующих влияние однонуклеотидной замены на признак. Поэтому в последние годы широко развернулись исследования по функциональной интерпретации данных GWAS как на уровне отдельных SNPs, так и на массовом уровне (Alsheikh et al., 2022; Degtyareva et al., 2021), а также работы по созданию независимых от GWAS масштабных функциональных подходов, основанных на анализе омиксных данных (Aguet et al., 2020; Degtyareva et al., 2021; URL: <https://www.gtexportal.org/home/>).

В частности, развитие методов транскриптомного (RNA-seq) и эпигеномного анализа (DNase-seq, ATAC-seq, ChIP-seq) открыло уникальную возможность количественного определения разницы в представленности двух аллелей (дисбаланс аллелей) каждого гетерозиготного SNP диплоидного организма в соответствующих данных (Degtyareva et al., 2021). Важным преимуществом такого подхода к выявлению потенциальных rSNPs состоит в том, что аллель-специфические события регистрируются для каждого индивидуума отдельно, то есть на одном генетическом фоне и в идентичных условиях окружающей среды, что позволяет получать достоверные данные при исследовании малых выборок, вплоть до одного человека (Harvey et al., 2015). Эти исследования в основном проводятся с использованием культивируемых клеточных линий, транскриптомные и эпигеномные профили которых могут очень сильно отличаться от таковых в исходных клетках и тканях человека. Работ же по выявлению rSNPs с помощью данного функционального подхода на клиническом материале пока еще очень мало, а для русских индивидуумов подобных исследований к настоящему времени не проводилось.

Целью настоящей работы было обнаружение rSNPs в масштабе генома на основании поиска аллель-специфических событий в данных ChIP-seq и RNA-seq, полученных на мононуклеарах периферической крови здоровых доноров, и применение сформированной панели rSNPs для выявления вариаций, вовлеченных в механизмы развития сахарного диабета второго типа и ответа на противодиабетический препарат метформин.

Задачи исследования:

1. Получить экспериментальные данные по профилю экспрессии генов (RNA-seq) и полногеномному профилю распределения меток активного хроматина H3K4me3 и H3K27ac (ChIP-seq) в мононуклеарных клетках периферической крови (PBMCs) условно здоровых людей.

2. Выполнить поиск rSNPs путем выявления аллель-специфических событий в полученных экспериментальных данных с помощью комплекса биоинформатических подходов. Сформировать панель rSNPs, расположенных в промоторных районах генов (± 1000 п.о. от старта транскрипции).

3. Охарактеризовать сформированную панель rSNPs с использованием полногеномных ресурсов; установить связь содержащихся в ней rSNPs с фенотипическими признаками и eQTLs путем сопоставления с имеющимися данными GWAS и консорциума GTEx соответственно.

4. Применить полученную панель к независимым транскриптомным данным для поиска rSNPs, потенциально вовлеченных в механизмы развития сахарного диабета второго типа и диабетической ретинопатии.

5. Применить полученную панель к независимым транскриптомным данным для поиска rSNPs, потенциально связанных с различным ответом пациентов с диабетом второго типа на терапию метформином.

Научная новизна работы. Впервые на клиническом материале – PBMCs российских индивидуумов – были параллельно получены экспериментальные данные по аллель-специфическому связыванию (ChIP-seq с антителами к гистоновым модификациям H3K4me3 и H3K27ac) и аллель-специфической экспрессии (RNA-seq). С помощью примененного биоинформатического алгоритма, выявляющего аллель-специфические события, в полученных данных выполнен поиск rSNPs, по результатам которого составлена панель из 14796 rSNPs, расположенных в промоторных районах 5132 генов. Показано достоверное обогащение данной панели полиморфными сайтами из каталога GWAS и eQTLs из данных консорциума GTEx. Пересечение rSNPs с участками ДНК в пределах ± 1000 п.о. от SNPs, зарегистрированных в GWAS и GTEx, показало связь 38.4% найденных rSNPs с различными фенотипическими признаками и 66.7% — с eQTLs. С использованием независимых транскриптомных данных найдены новые rSNPs, потенциально связанные с развитием диабетической ретинопатии и ответом на антидиабетический препарат метформин.

Теоретическая и научно-практическая ценность. Результаты данной работы расширяют представления об однонуклеотидных заменах, локализованных в регуляторных районах генома человека. Применяемый подход, основанный на анализе аллель-специфических событий в экспериментальных данных ChIP-seq и RNA-seq, успешно использован для полногеномного поиска rSNPs и интерпретации различных омиксных данных. Полученная с помощью этого подхода панель rSNPs вскрывает регуляторное значение большой группы SNPs, идентифицированных в исследованиях GWAS, а также служит независимым подтверждением регуляторной функции значительной группы eQTLs из каталога GTEx. Использование сформированной панели rSNPs для анализа независимых транскриптомных данных с целью выявления вариантов, потенциально связанных с развитием сахарного диабета типа 2 и ответом на метформин, открывает новые перспективы для дальнейшего изучения фундаментальных молекулярных механизмов, лежащих в основе формирования того или иного фенотипа, в том числе индивидуального ответа

на лекарства. Полученные в работе данные могут быть использованы для формирования новых и дополнения имеющихся диагностических панелей.

Положения, выносимые на защиту:

1. Создана панель однонуклеотидных замен (14796) в регуляторных районах генов (rSNP) на основе поиска аллель-специфических событий в экспериментальных данных ChIP-seq и RNA-seq, полученных на мононуклеарных клетках периферической крови условно здоровых людей, применение которой к результатам полногеномного исследования ассоциаций (GWAS) и независимым транскриптомным данным позволяет проводить широкомасштабный поиск rSNP в локусах риска развития различных заболеваний и выявлять новые rSNP, вовлеченные в патогенез заболеваний и ответ на лекарственные препараты, соответственно.

2. rSNP в регуляторных районах генов, белковые продукты которых задействованы в регуляции иммунного ответа (*EIF2AK2*: rs13008681, rs62133148, rs12992610; *STAT2*: rs2066818; *IRF5*: rs3823536, rs11767834, rs2004640) и гомеостаза глюкозы (*TP53*: rs1800899; *AKT1*: rs10136000, rs2498788, rs2498786; *PIK3R2*: rs182197351, rs12052085; *SREBF1*: rs12953299, rs13306736, rs565098931; *RPTOR*: rs8081444), вовлечены в механизмы развития сахарного диабета 2 типа, а rSNP в промоторных районах генов, кодирующих белки, участвующие в сборке комплекса I электрон-транспортной цепи митохондрий (*NDUFA11*: rs1056987; *NDUFBI*: rs12886834, rs7143338, rs34949057) и активации АМФ-зависимой протеинкиназы (*NMT1*: rs62065834, rs62065833, rs6503412; *AK5*: rs2689680) – в ответ на противодиабетический препарат метформин.

Апробация работы. Полученные результаты были представлены и обсуждены на 13-й международной мультиконференции “Bioinformatics of Genome Regulation and Structure/Systems Biology” – BGRS/SB» (Новосибирск, 2022).

Личный вклад автора. Экспериментальная работа – выделение мононуклеарных клеток периферической крови, выделение РНК, иммунопреципитация хроматина, приготовление ДНК и кДНК библиотек для высокопроизводительного секвенирования, была выполнена автором лично. Биоинформатическая работа - предварительная обработка данных и поиск rSNPs по ранее разработанному нашим научным коллективом алгоритму (н.с., к.б.н. Брызгалов Л.О.) с некоторыми изменениями, характеристика полученной панели rSNPs, построение ассоциативных генных сетей и анализ обогащения, также были выполнены автором самостоятельно.

Структура и объём диссертации. Работа состоит из введения, обзора литературы, материалов и методов, результатов, обсуждения, заключения, выводов и списка литературы (362 источника). Объем работы составляет 139 страниц. Работа включает 4 таблицы и 18 рисунков.

СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Обзор литературы

В разделе 1 приведена классификация SNPs на основе их функциональной роли, и для некоторых из них описаны молекулярные механизмы, посредством которых данные однонуклеотидные замены приводят к развитию различных заболеваний. В разделе 2 описаны широкомасштабные методы, применяемые для функциональной интерпретации SNPs, выявленных в GWAS. В разделе 3 описан широкомасштабный подход целенаправленного поиска rSNPs, основанный на

анализе аллель-специфических событий в эпигеномных (ChIP-seq, DNase-seq, ATAC-seq) и/или транскриптомных данных (RNA-seq), и его использование в функциональной интерпретации данных GWAS. В разделе 4 приведены основные механизмы развития сахарного диабета типа 2, изложена роль иммунных клеток в патогенезе данного заболевания, а также приведены результаты по функциональной интерпретации некоторых потенциальных rSNPs, ассоциированных с сахарным диабетом типа 2 по данным GWAS. В разделе 5 описана фармакокинетика и фармакодинамика сахароснижающего препарата метформина, приведены предполагаемые механизмы регуляции глюконеогенеза в печени данным препаратом и изложены некоторые результаты ген-кандидатного анализа и GWAS по поиску SNPs, связанных с индивидуальным ответом на метформин. В разделе 6 приведено краткое заключение, обосновывающее цель и задачи, поставленные в данной работе.

Глава 2. Материалы и методы

Участники исследования. Исследование проведено на группе девяти условно здоровых людей, у которых на момент взятия биологического материала не отмечалось никаких хронических заболеваний в анамнезе. Выборка включала мужчин (N=3) и женщин (N=6) в возрасте 22-47 лет (средний возраст – 32 года).

Выделение РВМС. Забор крови проводился в одной из медицинских клиник г. Новосибирска из локтевой вены натощак. Для экспериментов ChIP-seq и RNA-seq от каждого донора получали 20 мл крови и проводили выделение фракции мононуклеарных клеток на градиенте плотности фиколл-урографин ($\rho=1.077$; «Биолот», Россия) не позднее, чем через 1 час после взятия пробы.

Выделение РНК. Выделение РНК из РВМС проводилось с помощью реагента тризол («Ambion», США). Последующее выделение мРНК проводилось с помощью набора NEBNext Poly (A) mRNA Magnetic Isolation Module («New England Biolabs», США) согласно протоколу производителя.

Иммунопреципитация хроматина с использованием антител к H3K4me3 или H3K27ac. Для иммунопреципитации хроматина использовался метод с фиксацией хроматина 1% формальдегидом («SIGMA», США). Фрагментацию хроматина осуществляли с помощью ультразвукового гомогенизатора («Bandelin electronic GMBH & CO. KG», Германия). Реакцию иммунопреципитации проводили с использованием поликлональных кроличьих антител к H3K4me3 или H3K27ac (ChIP Grade, ab8580/ab4729, «Abcam», Великобритания).

Анализ обогащения ДНК образцов фрагментами ДНК, связанными с H3K4me3. Для образцов, полученных в результате иммунопреципитации хроматина с антителами к H3K4me3, была проведена проверка обогащения целевыми фрагментами ДНК с помощью ПЦР в режиме реального времени со специально подобранными LNA-модифицированными зондами. В качестве «положительного» контроля были подобраны зонды на регуляторные районы генов *KLF6* и *ATXN7L3B*, высоко экспрессирующихся в лейкоцитах (по данным The Human Protein Atlas (URL: <https://www.proteinatlas.org>)) и имеющих высокий уровень триметилирования H3K4 в районе старта транскрипции (по данным ENCODE (URL: <https://www.encodeproject.org>)). Для «отрицательного» контроля подбирались зонды на регуляторный район вблизи старта транскрипции *KRT15*, не экспрессирующегося в лейкоцитах и имеющего низкий уровень триметилирования H3K4 в данном районе.

Подготовка библиотек для секвенирования. Подготовка ДНК и кДНК библиотек выполнялась с помощью наборов: NEBNext Ultra II Directional RNA Library Prep Kit for Illumina («New England Biolabs», США), NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) («New England Biolabs», США) и NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 2) («New England Biolabs», США) согласно протоколам производителя.

Биоинформатический анализ данных ChIP-seq и RNA-seq. Первичная фильтрация данных секвенирования проводилась с помощью Trimmomatic (v. 3.2.2). Выравнивание прочтений на референсный геном сборки GRCh38/hg38 выполнялось с помощью Bowtie 2 (v. 2.5.4). С помощью samtools были отобраны прочтения с показателем качества картирования MAPQ>20. Отбор гетерозиготных позиций проводился с помощью команды bcftools mpileup после объединения данных ChIP-seq и RNA-seq по каждому человеку. Для каждого человека дополнительно создавали индивидуальный альтернативный геном, в котором для всех гетерозиготных позиций референсный аллель заменяли на альтернативный. Далее, все файлы ChIP-seq и RNA-seq для каждого индивидуума выравнивали на индивидуальный альтернативный геном и отбирали прочтения с MAPQ>20. Поиск гетерозиготных SNPs в промоторных районах (± 1000 п.о. от старта транскрипции) из данных ChIP-seq и транскрибируемых районах генов из данных RNA-seq осуществлялся с помощью пакета TxDb.Hsapiens.UCSC.hg38.knownGene (v. 3.18) в среде R. Расчет аллельной асимметрии для отобранных гетерозиготных SNPs проводился между прочтениями с референсным аллелем, выровненными на референсный геном, и прочтениями с альтернативным аллелем, выровненными на альтернативный геном, по биномиальному тесту с последующей поправкой на множественное сравнение по Бенджамини-Хохбергу ($p_{adj} < 0.1$).

Характеристика панели rSNPs. Для характеристики полученной панели rSNPs были привлечены данные консорциума GTEx (The Genotype-Tissue Expression) (URL: <https://www.gtexportal.org/home/>), каталога GWAS (URL: <https://www.ebi.ac.uk/gwas/>) и веб сервера ANANASTRA (ANnotation and enrichment ANalysis of Allele-Specific TRANscription factor binding at SNPs) (URL: <https://ananastra.autosome.org/>).

Поиск и анализ дифференциально экспрессирующихся генов. Поиск дифференциально экспрессирующихся генов (ДЭГ) проводился с помощью пакета DeSeq2 (v. 1.44). Анализ ДЭГ выполнялся с использованием STRING (v. 12.0), CytoNCA (v. 2.1), MCODE (v. 2.0.3), clusterProfiler (v. 4.12.6) и pROC (v. 1.18.5).

Главы 3-4. Результаты и обсуждение

1. Алгоритм поиска rSNPs, связанных с аллель-специфическими событиями в данных ChIP-seq и RNA-seq. Стратегия поиска rSNPs включала 6 основных этапов (Рисунок 1). На первом этапе проведено выделение PBMCs из периферической крови девяти условно здоровых доноров. Каждый образец PBMCs разделяли на две части. Для одной части были получены данные секвенирования активных регуляторных районов генома, маркированных гистоновыми модификациями H3K4me3 и H3K27ac (ChIP-seq). Параллельно для второй части PBMCs выполнено секвенирование транскриптома (RNA-seq). Для поиска гетерозиготных позиций прочтения из данных ChIP-seq и RNA-seq, выровненные на референсный геном, для каждого индивидуума были объединены. Затем, проводилось выравнивание данных ChIP-seq и RNA-seq на индивидуальный

альтернативный геном, в котором во всех найденных гетерозиготных позициях, референсный аллель был заменен на альтернативный. По результатам расчета аллельной асимметрии в данных ChIP-seq было отобрано 37521 SNPs, связанных с событиями аллель-специфического связывания (Allele-Specific Binding, ASB) в промоторных районах в пределах ± 1000 п.о. от старта транскрипции известных генов. По результатам расчета аллельной асимметрии в данных RNA-seq было отобрано 59180 SNPs, связанных с событиями аллель-специфической экспрессии (Allele-Specific Expression, ASE). Гены, в которых был найден только один гетерозиготный SNP, при этом идентифицированный как ASE SNP, считались экспрессируемыми аллель-специфическим образом. В случае обнаружения в пределах гена более одного гетерозиготного SNP, для оценки общей ASE использовался метааналитический подход MBASED (Mayba et al., 2014). На заключительном этапе поиска rSNPs проведено пересечение генов, в промоторных районах которых локализованы ASB SNPs, с генами, экспрессирующимися аллель-специфическим образом. Регуляторными (rSNPs) считались ASB SNPs в промоторах генов, для которых наблюдалась аллель-специфическая экспрессия (ASE). В совокупности, по результатам определения аллель-специфических событий в данных ChIP-seq и RNA-seq нами была сформирована панель из 14796 rSNPs, связанных с экспрессией 5132 генов.

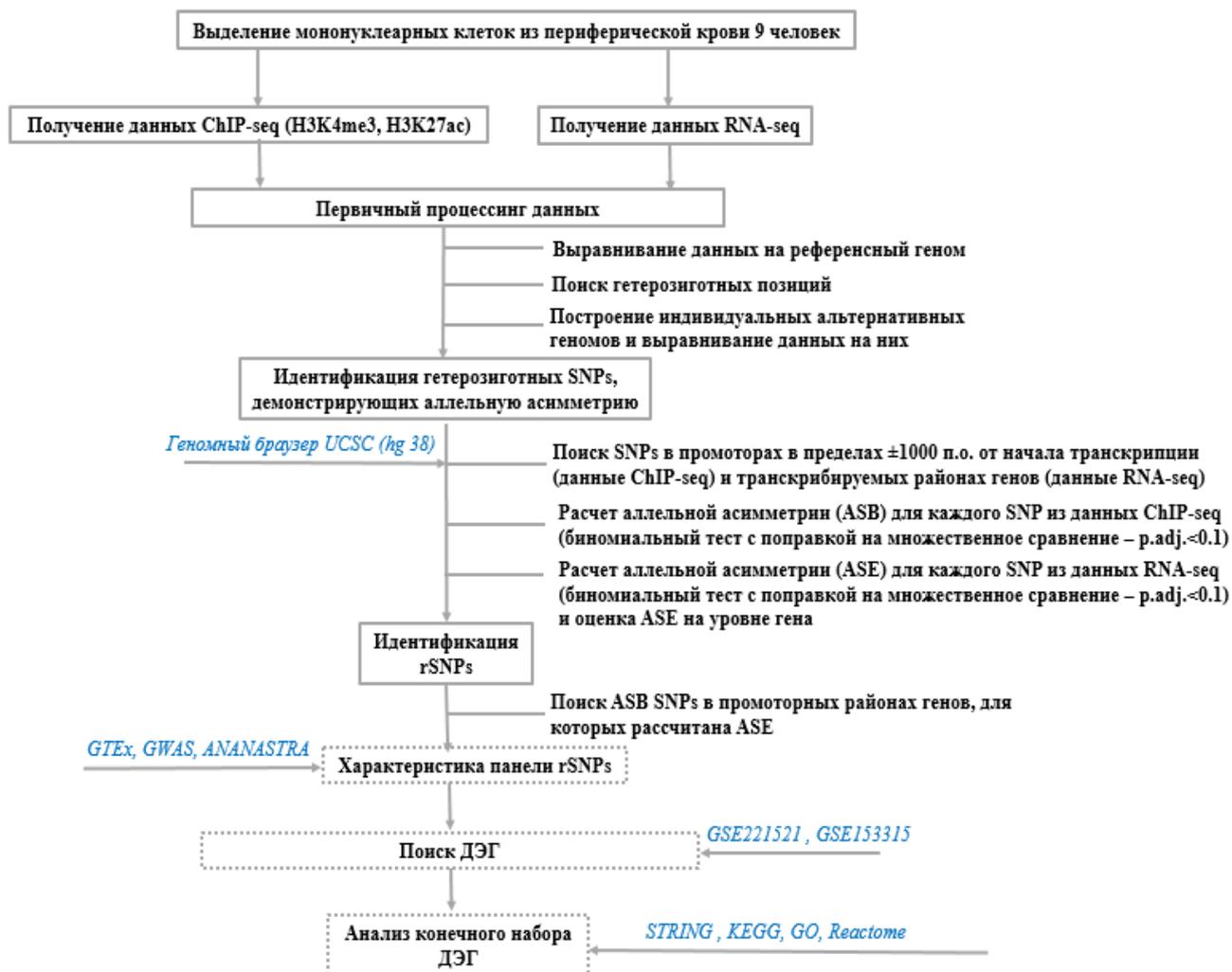


Рисунок 1. Схематическое изображение основных этапов поиска и дальнейшего

анализа rSNPs. Этапы поиска rSNPs выделены сплошным серым контуром. Этапы дальнейшего анализа полученной панели rSNPs выделены пунктирным серым контуром. Голубым курсивом указаны источники получения данных.

2. Характеристика полученной панели rSNPs с использованием данных GWAS, GTEX и ANANASTRA. Пересечение rSNPs с геномными районами в пределах ± 1000 п.о. от SNPs, содержащихся в каталоге GWAS (по состоянию на 03.2023 г.), показало связь 5688 rSNPs (38.4% от всех rSNPs в представленной панели) с различными фенотипическими признаками. При этом, 1107 rSNPs (7.5% от всех rSNPs) были непосредственно представлены в самом каталоге GWAS, а остальные располагались в пределах ± 1000 п.о. от 4483 SNPs из данного каталога. Таким образом, с одной стороны, значительная часть выявленных нами rSNPs (1107) действительно связана с различными фенотипическими признаками, а с другой стороны, полученные нами данные раскрывают молекулярный смысл 4483 SNPs из каталога GWAS, указывая на их регуляторную функцию. Мы обнаружили значительное обогащение панели rSNPs вариантами GWAS по сравнению с набором гетерозиготных нерегуляторных SNPs (критерий χ^2 : $p.value < 2.2 \cdot 10^{-16}$; OR=1.7; 95% CI:1.6-1.8), отобранных по покрытию в наших данных ChIP-seq, что подтверждает существенную роль найденных rSNPs в формировании признаков.

Для rSNPs, непосредственно представленных в GWAS, мы рассчитали обогащение ассоциациями с фенотипами из данного каталога (критерий Фишера с поправкой по Бенджамини-Хохбергу - $p.adj < 0.1$) (Таблица 1). В полученной группе в основном были представлены признаки, связанные с количественными показателями клеток крови (что вполне ожидаемо, поскольку поиск rSNPs был проведен на PBMCs): средний корпускулярный объем, средний корпускулярный гемоглобин, количество лейкоцитов, тромбоцитов, лимфоцитов, моноцитов, эритроцитов, нейтрофилов, эозинофилов, ширина распределения эритроцитов, средний объем тромбоцитов, процент моноцитов в лейкоцитах и др. В совокупности, в нашей панели оказалось 254 rSNPs, связанных с количественными показателями клеток крови. Также выделялась группа признаков, связанная с различными морфометрическими показателями: соотношение талии и бедер с учетом индекса массы тела, окружность бедер с поправкой на индекс массы тела, индекс формы тела, вес при рождении (в совокупности, найдено 49 rSNPs). Достаточно многочисленной была также группа признаков, включающая различные заболевания, которые характеризуются хроническим воспалением: астма (9 rSNPs), ревматоидный артрит (8 rSNPs), рассеянный склероз (8 rSNPs) и диабет типа 2 (23 rSNPs). При этом ревматоидный артрит и рассеянный склероз относились к классу аутоиммунных. Еще одну группу признаков GWAS представляли онкологические заболевания: базальноклеточная карцинома (6 rSNPs), хронический лимфоцитарный лейкоз (4 rSNPs) и плоскоклеточный рак легких (3 rSNPs).

Из всех заболеваний с сахарным диабетом типа 2 было связано наибольшее количество rSNPs из нашей панели. Эти rSNPs были расположены в промоторах генов, участвующих в регуляции процессов воспаления (rs12823740 - *CCDC92* (Ren et al., 2023), rs73069940 - *CTBP1* (Wang et al., 2020a)), ожирения (rs10097617, rs1320164, rs896852 - *TP53INP1* (Seillier et al., 2015)), резистентности к инсулину (*TP53INP1* (Seillier et al., 2015), *CCDC92* (Ren et al., 2023)), секреции инсулина (rs11671664 - *GIPR* (Erfanian et al., 2023), rs893617 - *AP3S2* (Wood et al., 2017)), пролиферации β -клеток поджелудочной железы (rs2066827 - *CDKN1B* (Uchida et al., 2005), митофагии (*TP53INP1* (Seillier et al., 2015)) и транспорта углеводов (rs1570247

- *GBA2* (Körschen et al., 2013)). Также, из признаков стоит отметить уровень гликированного гемоглобина A1c, который является важным маркером сахарного диабета второго типа (связано 14 rSNPs из нашей панели). Кроме того, rSNPs, связанные с вышеупомянутыми морфометрическими показателями, также могут быть маркерами сахарного диабета типа 2. Повышение этих показателей может свидетельствовать об ожирении, которое наблюдается примерно у 90% людей с диабетом второго типа (Whitmore, 2010).

Таблица 1. ТОП 20 признаков, ранжированных по количеству rSNPs, представленных в GWAS

Признак	Общее количество SNPs, включенных в каталог GWAS, для данного признака	Количество rSNPs, напрямую включенных в каталог GWAS, для данного признака	P.adj.	Отношение шансов*
Соотношение талии и бедер с учетом индекса массы тела	3741	36	$1.1 \cdot 10^{-3}$	2.2 (1.5-3.1)
Средний корпускулярный объем	2227	31	$1.1 \cdot 10^{-5}$	3.2 (2.2-4.6)
Количество лейкоцитов	2456	31	$6.1 \cdot 10^{-5}$	2.9 (2-4.2)
Соотношение талии и бедер	2526	28	$1.1 \cdot 10^{-3}$	2.5 (1.7-3.7)
Окружность бедер с поправкой на индекс массы тела	3359	28	$4.4 \cdot 10^{-2}$	1.9 (1.3-2.8)
Количество тромбоцитов	2609	27	$3.9 \cdot 10^{-3}$	2.4 (1.6-3.5)
Средний корпускулярный гемоглобин	2365	25	$4.8 \cdot 10^{-3}$	2.4 (1.6-3.6)
Количество лимфоцитов	1651	24	$9.8 \cdot 10^{-5}$	3.3 (2.1-5)
Количество моноцитов	1767	24	$2.1 \cdot 10^{-4}$	3.1 (2-4.7)
Количество эритроцитов	2487	24	$1.9 \cdot 10^{-2}$	2.2 (1.4-3.3)
Количество нейтрофилов	1559	23	$1.1 \cdot 10^{-4}$	3.4 (2.1-5.1)
Количество эозинофилов	2100	23	$4.8 \cdot 10^{-3}$	2.5 (1.6-3.8)
Сахарный диабет типа 2	2830	23	$7.1 \cdot 10^{-2}$	1.9 (1.2-2.8)
Ширина распределения эритроцитов	1732	20	$5.5 \cdot 10^{-3}$	2.6 (1.6-4.1)
Средний объем тромбоцитов	1402	16	$2.1 \cdot 10^{-2}$	2.6 (1.5-4.3)
Индекс формы тела	1514	16	$3.6 \cdot 10^{-2}$	2.4 (1.4-3.9)
Процентная доля моноцитов в лейкоцитах	738	15	$1.9 \cdot 10^{-4}$	4.6 (2.6-7.7)
Тромбоцитокрит	930	15	$1.6 \cdot 10^{-3}$	3.7 (2-6.1)
Средний корпускулярный объем эритроцитов	821	14	$1.6 \cdot 10^{-3}$	3.9 (2.1-6.6)
Аппендикулярная тощая масса	1569	14	$8.8 \cdot 10^{-2}$	2 (1.1-3.4)

* 95% доверительный интервал указан в скобках

На следующем этапе анализа нашей панели rSNPs, используя данные консорциума GTEx, доступные на 03.2023 г., мы провели поиск rSNPs, связанных с изменением экспрессии генов в различных тканях (поиск eQTLs). Сопоставление нашей панели из 14796 rSNPs с данными анализа eQTLs показало существенно большее совпадение результатов по сравнению с GWAS. Было установлено, что 9871 rSNPs (66.7% от всей панели rSNPs) связаны с eQTLs, при этом 2474 rSNPs

были непосредственно представлены в каталоге GTE_x, а 7397 rSNPs находились в пределах ± 1000 п.о. от 9646 eQTLs из данного каталога. Таким образом, наши данные являются независимым подтверждением функциональности 12120 eQTLs (из которых 2474 eQTLs являются rSNPs согласно полученным нами результатам, а 9646 располагаются в пределах ± 1000 п.о. от rSNPs из нашей панели и предположительно сцеплены с ними). Из всех тканей, данные по которым представлены в GTE_x, наибольшее количество rSNPs ожидаемо было найдено в «цельной крови» - 2464 (результаты для отдельных фракций клеток крови в каталоге не представлены). 4280 rSNPs (28.9% от всех найденных rSNPs) в 1628 генах были одновременно связаны и с изменением экспрессии генов (по данным GTE_x) и с каким-либо фенотипическим признаком (по данным GWAS) (Рисунок 2). Также, по сравнению с гетерозиготными нерегуляторными SNPs, идентифицированными в наших ChIP-seq данных, наша панель rSNPs была значительно обогащена eQTLs вариантами из GTE_x ($p.value < 2.2 \cdot 10^{-16}$; OR=2.49; 95% CI:2.38-2.6).

Значительное совпадение с полученным нами списком 14796 rSNPs было выявлено при пересечении с вариантами, найденными в локусах аллель-специфического связывания (ASB) различных транскрипционных факторов (данные ChIP-seq) ресурса ANANASTRA. Оказалось, что 30.8% (4560, FDR<0.05) наших rSNPs локализируются в сайтах ASB различных ТФ (Рисунок 2). Доля таких rSNPs возрастает до 33,4% (1433, FDR<0.05) при использовании списка из 4280 rSNPs, связанных с eQTLs и различными фенотипическими признаками GWAS. Среди 1433 rSNPs, наибольшее количество было расположено в локусах аллель-специфического связывания ТФ ANDR (N=217), CTCF (N=215), STAT1 (N=167), BRD4 (N=126), ESR1 (N=121), YY1 (N=97), GCR (N=89), SPT5H (N=76), STAG1 (N=74) и ZFX (N=63). Учитывая, что в ANANASTRA представлены ASB SNPs не только в промоторах, но и в других регуляторных районах, а поиск rSNPs в нашем исследовании проводился лишь в пределах ± 1000 п.о. от сайтов старта транскрипции, мы могли бы ожидать большего совпадения результатов с ANANASTRA при более широком охвате регуляторных районов.

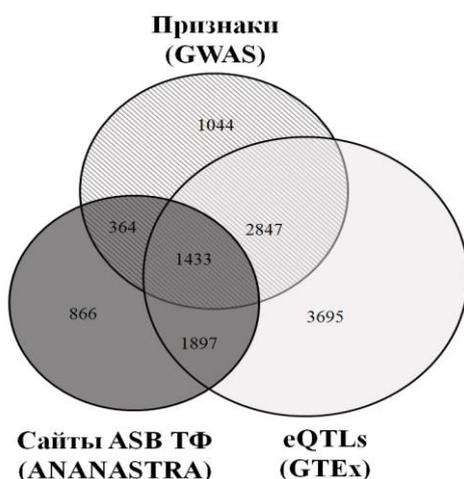


Рисунок 2. Диаграмма Венна, иллюстрирующая количество rSNPs, локализованных в сайтах аллель-специфического связывания ТФ (данные ANANASTRA), связанных с фенотипическими признаками (данные GWAS) и изменением экспрессии генов (данные GTE_x), а также расположенных на пересечении соответствующих групп.

3. Анализ модулей, идентификация хаб генов и путей в ассоциативных генных сетях дифференциально экспрессируемых генов, связанных с сахарным диабетом типа 2. Учитывая обогащение нашей панели

rSNPs вариантами из каталога GWAS, связанными с сахарным диабетом второго типа и рядом сопутствующих признаков, а также тот факт, что из всех заболеваний с диабетом типа 2 связано наибольшее количество rSNPs, мы провели дополнительный поиск rSNPs из нашей панели, которые могут быть вовлечены в механизмы развития диабета типа 2 в независимом наборе транскриптомных данных. Нами была проанализирована самая обширная выборка данных RNA-seq

(по состоянию на 12.2023 г.) из репозитория GEO DataSets, полученных на РВМС лиц с сахарным диабетом второго типа (GSE221521). Эта выборка включала здоровых людей (N=50), людей с диагнозом сахарный диабет типа 2 (N=74) и пациентов с сахарным диабетом типа 2 и дополнительным диагнозом диабетическая ретинопатия (N=69) (Xiang et al., 2023). При сравнении группы здоровых доноров и пациентов с диабетом типа 2 без ретинопатии, мы не выявили дифференциально экспрессирующихся генов (ДЭГ) ($|\log_2FC| > 0.2$; $p_{adj} < 0.05$). Возможно это связано с эффектами получаемой пациентами терапии (к сожалению, авторами статьи не было указано, какой именно) и выбором индивидуумов для контрольной группы, в которую вошли люди со средним показателем HbA1c=5.80%, что по некоторым данным соответствует уровню, характерному для преддиабетического состояния (Wang et al., 2023a). При сравнении группы пациентов с диабетической ретинопатией и здоровых доноров нами было выявлено 4612 ДЭГ, среди которых в промоторных районах 1284 найдено 3810 rSNPs из нашей панели (Рисунок 3). При этом, 2481 rSNPs располагались в промоторах 772 ДЭГ с повышенной экспрессией ($\log_2FC > 0$), а 1329 rSNPs в промоторах 512 ДЭГ с пониженной экспрессией ($\log_2FC < 0$).

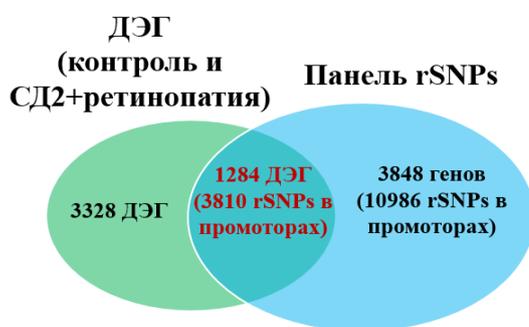


Рисунок 3. Диаграмма Венна, демонстрирующая количество rSNPs из нашей панели, расположенных в промоторах ДЭГ между пациентами с сахарным диабетом типа 2 (СД2) с диабетической ретинопатией и здоровым контролем (GSE221521).

Среди выявленных ДЭГ, в промоторах которых найдены rSNPs из нашей панели, был проведен поиск хаб генов,

под которыми в широком смысле понимают гены, являющиеся наиболее важными регуляторами различных биологических процессов и взаимодействующие с многочисленными генами- или белками-мишенями. Мы исходили из предположения, что rSNPs в промоторах хаб генов, влияя на их экспрессию, могут вносить большой вклад в нарушения регуляции ключевых процессов, связанных с развитием диабета и его осложнений. По результатам построения сети белок-белковых взаимодействий (Protein-Protein Interaction, PPI) между белковыми продуктами ДЭГ с повышенной экспрессией и расчета топологических параметров, мы выделили белковые продукты 31 хаб гена. К хаб генам в основном относились гены, участвующие в регуляции транскрипции, кодирующие: транскрипционные факторы (*TP53*, *SREBF1*, *RXRA*), корепрессор (*NCOR2*), хроматин-связывающий белок (*BRD4*), гистондеацетилазу (*HDAC4*), субъединицу РНК-полимеразы II- *RPB1* (*POLR2A*). Отдельную группу составляли гены, кодирующие серин/треонин киназы (*AKT1*, *MAPK3*, *ATM*, *GAK*, *SMG1*) и клеточные рецепторы (*NOTCH1*, *SCARB1*, *INSR*, *ITGAM*). Многие из найденных хаб генов вовлечены в различные процессы, связанные с развитием диабета типа 2 (Bengal et al., 2020; Saltiel, 2021; Sliwinska et al., 2017), такие как ожирение (*SREBF1*, *H6PD*), воспаление (*TGFBI*), резистентность к инсулину (*RPTOR*, *AKT1*) и т.д. Стоит отметить, что анализ обогащения хаб генов путями KEGG установил Insulin signaling pathway (7 генов) как путь с наибольшим обогащением ($p_{adj} < 5.3 \cdot 10^{-5}$; коррекция по Бенджамини-Хохбергу). В топ 10 путей с наибольшим обогащением также вошел FoxO signaling pathway ($p_{adj} < 1.9 \cdot 10^{-4}$), который ответственен за регуляцию метаболизма глюкозы (Jeon, 2016; Teaney, Cug,

2023) (Рисунок 4 А и Б).

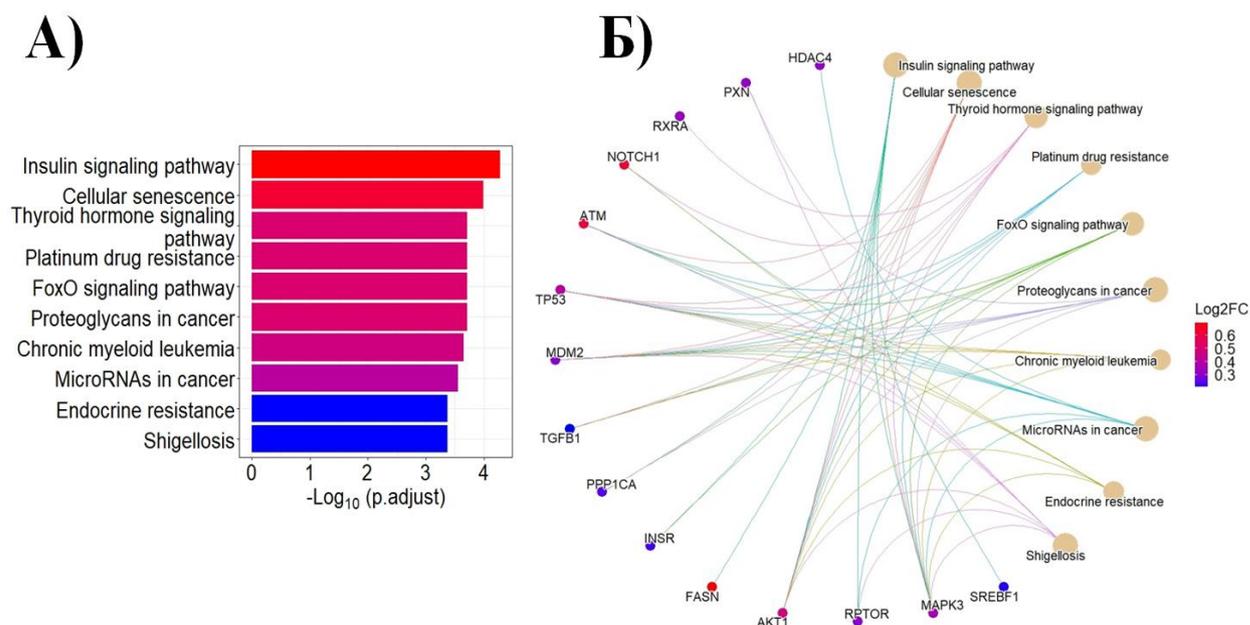


Рисунок 4. Анализ хаб генов с повышенной экспрессией в сети PPI, в промоторах которых найдены rSNPs из нашей панели. **А)** Анализ обогащения хаб генов путями KEGG. Представлены 10 путей с наибольшим обогащением. На горизонтальной оси отложен $-\log_{10}$ от скорректированного значения P-value. Коррекция P-value выполнялась с помощью метода Бенджамини-Хохберга. **Б)** Сетевой график, иллюстрирующий связь между десятью наиболее значимыми путями KEGG, полученными по результатам анализа обогащения (Рисунок 4 А), и хаб генами. Хаб гены выделены цветом в зависимости от \log_2 fold change. Пути KEGG представлены в виде кругов, размер которых отражает количество хаб генов, вовлеченных в эти пути. Пути KEGG и хаб гены, представленные в этих путях, соединены цветными линиями.

С целью выявления процессов и путей, задействованных в развитии сахарного диабета второго типа, и выделения в них функционально связанных групп ДЭГ, в промоторах которых найдены rSNPs, в сети PPI с использованием MCODE (приложение Cytoscape) были выделены значимые модули (Рисунок 5 А-В). Анализ обогащения терминами GO, путями KEGG и Reactome показал, что гены в этих модулях участвуют в процессах метаболизма глюкозы и иммунного ответа. Так, *PIK3R2* и *PIK3CG*, представленные в первом модуле (Рисунок 5 А), кодируют субъединицы PI3K (Phosphoinositide 3-kinases) киназы, функционирующей в качестве вышестоящего регулятора в пути PI3K/АКТ. PI3K опосредует активацию ключевого фермента этого сигнального пути - Akt (АКТ Serine/Threonine Kinase) киназы, которая способствует: перемещению глюкозы внутрь клеток путем опосредованной активации транспортеров глюкозы (GLUT) (Cruz-Pineda et al., 2021); активации гликогеногенеза путем ингибирования киназы гликогенсинтазы-3 (GSK-3) (Manning, Cantley, 2007); негативной регуляции глюконеогенеза путем снижения экспрессии глюкозо-6-фосфатазы (G6PC) (Huang et al., 2018b) и фосфоенолпируваткарбоксикиназы (PCK) (Yu et al., 2021). Примечательно, что нами были обнаружены rSNPs в промоторных районах генов *Akt1* (rs10136000, rs2498788, rs2498786 и rs36214921), *G6PC3* (rs72822657, rs3088347), кодирующего каталитическую субъединицу глюкозо-6-фосфатазы, *GSK3B* (rs334558), ответственного за синтез одной из изоформ киназы-3 гликогенсинтазы и *PCK2*

(rs4982856), обеспечивающего продукцию митохондриальной формы фосфоенолпируваткарбоксикиназы (Рисунок 5 А). Важную роль в воспалении и развитии сахарного диабета типа 2 играет также *TP53* (rs1800899) (Рисунок 5 Б), кодирующий опухолевый супрессор – ТФ p53, который способен напрямую связываться с промоторами *Glut1* и *Glut4* и подавлять их транскрипцию, что способствует развитию инсулинорезистентности (Kung, Murphy, 2016).

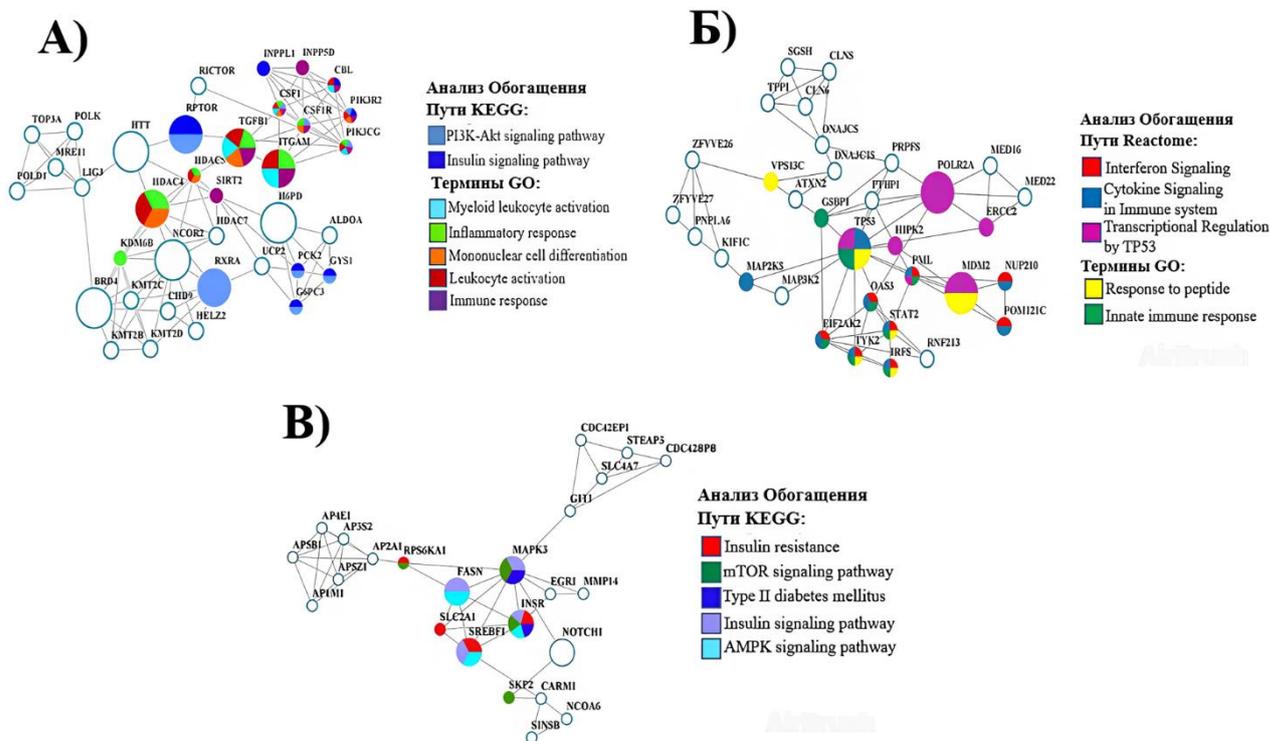


Рисунок 5. Значимые модули в сети PPI, для которых проведен анализ обогащения KEGG, GO и Reactome. **А)** Первый модуль. **Б)** Второй модуль. **В)** Третий модуль. Белки представлены в виде узлов в сети PPI, а взаимодействия между ними обозначены ребрами. Хаб белки представлены более крупными узлами. Белки, аннотированные с помощью KEGG, Reactome и GO, выделены соответствующим цветом как указано в легенде. В легенде указаны статистически значимые термины и пути ($p.adj.<0.05$; коррекция по Бенджамини-Хохбергу), связанные с инсулинорезистентностью и воспалением, которые были выявлены по результатам анализа обогащения.

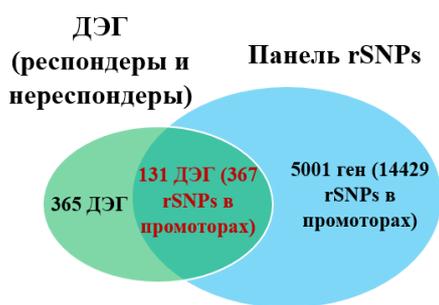
Таким образом, наши данные указывают на то, что rSNPs, влияющие на экспрессию генов в выделенных модулях, могут участвовать в развитии сахарного диабета типа 2 и диабетической ретинопатии, приводя к нарушению регуляции процессов иммунного ответа и метаболизма глюкозы. А rSNPs в промоторах хаб генов могут быть ключевыми регуляторами этих процессов, изменение экспрессии которых может также влиять на их гены/белки-мишени. Дополнительно мы выделили 117 rSNPs, расположенных в промоторах хаб генов, которые на основании уровня их экспрессии позволяют с достаточно высокой точностью разделить людей с диабетом второго типа с диабетической ретинопатией и здоровых лиц в анализируемых данных (ROC анализ; $AUC>0.7$). Такие rSNPs представляют большой интерес для дальнейшего изучения.

4. Поиск rSNPs, потенциально связанных с индивидуальным ответом на противодиабетический препарат метформин. Используя сформированную нами панель rSNPs, мы также провели поиск rSNPs, которые могут быть связаны с ответом

на терапию метформином — противодиабетическим лекарственным средством из группы бигуанидов, который является основным препаратом в лечении и профилактики сахарного диабета 2 типа. Метформин оказывает сахароснижающее действие за счет ингибирования глюконеогенеза и гликогенолиза в печени, задерживания всасывания глюкозы из желудочно-кишечного тракта, повышения чувствительности тканей к инсулину, снижения уровня общего холестерина, липопротеидов низкой плотности и триглицеридов (Foretz et al., 2014; Mkrtumyan et al., 2023).

Для решения поставленной задачи нами были проанализированы данные RNA-seq (GSE153315) клеток крови лиц с сахарным диабетом второго типа, отвечающих на метформин в течение 3-х месячного курса лечения (n=10), и лиц с диабетом второго типа, не отвечающих на метформин при проведении такого курса (n=10) (Vohra et al., 2022). Между группами лиц, отвечающей и не отвечающей на лечение метформином, было найдено 496 ДЭГ. Оказалось, что в промоторных районах 131 из этих генов локализовано 367 rSNPs из нашей панели. 306 rSNPs были расположены в промоторах 118 ДЭГ с повышенной экспрессией, а 61 rSNPs локализовались в промоторах 13 ДЭГ с пониженной экспрессией (Рисунок 6).

Рисунок 6. Диаграмма Венна, демонстрирующая количество rSNPs из нашей панели, расположенных в промоторах ДЭГ между пациентами с диабетом типа 2, отвечающими (респондеры) и не отвечающими (нереспондеры) на метформин в течение трехмесячного курса лечения (GSE153315).



Среди ДЭГ с повышенной экспрессией (лица, отвечающие на метформин, по сравнению с не отвечающими), промоторы которых содержали rSNPs из нашей панели, выделялась группа генов, кодирующих транскрипционные факторы и транскрипционные регуляторы. К ним относятся: 1) *FOXP1* (rs9827299, rs927809457), кодирующий широко известный регулятор

гомеостаза глюкозы (Zou et al., 2015), который также является известным супрессором дифференцировки макрофагов в провоспалительный фенотип (Bartolini et al., 2022); 2) *POU2F2* (rs3826705), продукт которого через активацию сигнального пути АКТ/mTOR играет важную роль в гомеостазе глюкозы (Yang et al., 2021b), а также участвует в клеточном иммунном ответе путем регуляции генов пролиферации и дифференцировки В-клеток (Rahman et al., 2020); 3) *YY-1* (rs113799953, rs760756349), который экспрессируется во всех типах клеток, участвуя в широком спектре регуляторных процессов, в том числе регуляции метаболизма липидов и глюкозы, а также иммунного ответа (Qian, Zhao, 2022a); 4) *KLF6* (rs3812715, rs17135808, rs3829201, rs10795076, rs3812714), продукт которого связан с дифференцировкой моноцитов в макрофаги (Mola et al., 2023); 5) *NFIX* (rs11555274, rs1354007230, rs1293177918, rs1235554616), продукт которого является одним из предполагаемых регуляторов миелопоэза (Jenthro et al., 2021); 6) *GMEB1* (rs1013370834, rs1022902336, rs895890911), продукт которого необходим для нормального функционирования островков эндотелиальных клеток (Xiong et al., 2022); 7) *NCOR1* (rs8076864, rs1331834429, rs178797), кодирующий корепрессорный белок, вовлеченный в механизмы инсулинрезистентности (Paluvai et al., 2023). Также было выявлено значительное число rSNPs в промоторных районах генов

транскрипционных факторов из семейства цинковых пальцев, чьи функции пока еще практически не изучены: *TSHZ2* (rs71354397); *ZNF121* (rs7253981); *ZNF44* (rs386901, rs422563); *ZNF440* (rs1056484229); *ZNF611* (rs1350535255, rs8109324, rs11879101, rs707303, rs10419223, rs1044394923, rs55769230); *ZNF714* (rs7258409, rs182211521, rs61733856); *ZNF793* (rs2291003); *ZNF813* (rs1467168591, rs1157700262, rs2015145, rs4599028, rs112581977, rs1968829, rs181026484).

Отдельно стоит выделить rSNPs, расположенные в генах, которые участвуют в известных механизмах действия метформина (34 rSNPs). Наиболее интересными среди них являются *NDUFA11* и *NDUFB1* (в совокупности найдено 13 rSNPs), кодирующие субъединицы I комплекса цепи переноса электронов в митохондриях, который является широко известной мишенью метформина. Метформин ингибирует I комплекс электрон-транспортной цепи в печени, что приводит к уменьшению концентрации АТФ и увеличению АМФ в клетках и, как следствие, активации АМР-зависимой киназы (АМРК), которая ингибирует глюконеогенез (Рисунок 7 А) (Emelyanova et al., 2021; Fontaine, 2018; LaMoia, Shulman, 2021). *AK5* и *NMT1* являются вышестоящими регуляторами АМРК. *AK5* (rs2689680) кодирует аденилаткиназу, которая катализирует обратимую реакцию между двумя молекулами АДФ с образованием одной молекулы АТФ и АМФ. Связывание молекул АМФ с γ субъединицей АМРК аллостерически активирует данный фермент (Рисунок 7 А) (Ahn et al., 2021). *NMT1* (rs62065834, rs62065833, rs6503412) кодирует миристоилтрансферазу 1, которая добавляет миристоильную группу к N-концевому остатку глицина β субъединицы АМРК, что способствует закреплению АМРК на поверхности лизосомных мембран, где происходит её дальнейшая активация (Рисунок 7 Б). В активированном состоянии АМРК ингибирует комплекс mTORC1, также расположенный на поверхности лизосомных мембран, который активирует пути, связанные с синтезом белков, роста и пролиферации клеток (Herzig, Shaw, 2018; Weyand, Goronzy, 2020).

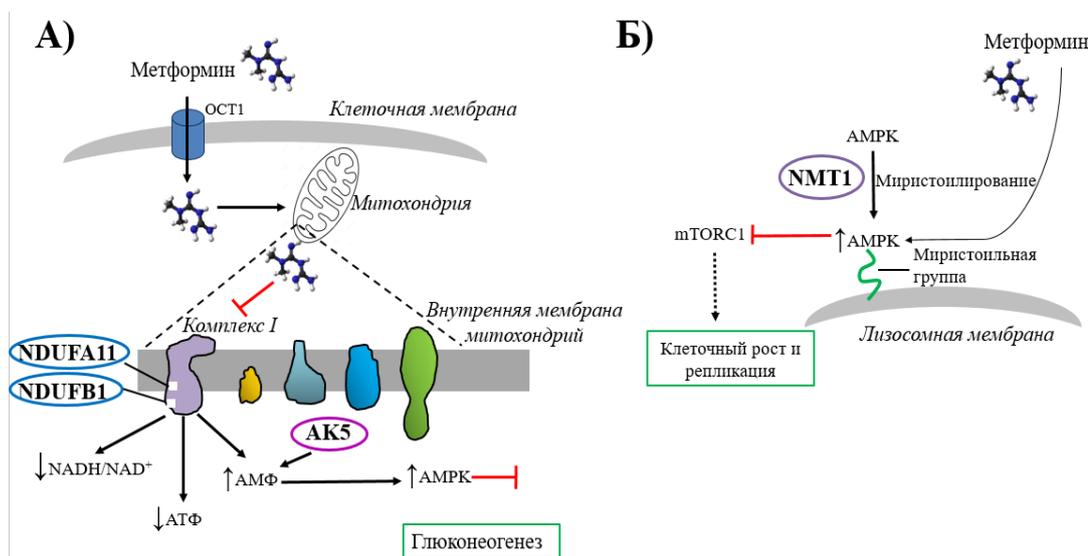


Рисунок 7. Схематическое изображение известных путей реализации эффектов метформина и вовлеченных в них белковых продуктов ДЭГ, в промоторах которых были найдены rSNPs из сформированной нами панели. Основные антидиабетические эффекты метформина реализуются через: **А)** ингибирование комплекса I электрон-транспортной цепи в митохондриях и **Б)** активацию АМРК на поверхности лизосомных мембран. Белковые продукты дифференциально экспрессирующихся генов между лицами с диабетом второго типа, отвечающих и не отвечающих на лечение метформинном, в

промоторах которых найдены rSNPs из нашей панели, выделены цветным овалом.

Важно отметить, что к настоящему времени с помощью исследований ассоциаций (ген-кандидатный анализ, GWAS) было выявлено всего несколько десятков SNPs, связанных с индивидуальным ответом на метформин, значительная часть из которых расположена в генах, кодирующих транспортеры органических катионов (OCTs) и белки, выводящие лекарства и токсины (MATE) (Goswami et al., 2014; Nasykhova et al., 2022; Nasykhova et al., 2020; Xiao et al., 2016; Zhou et al., 2015; Rotroff et al., 2018). Наше исследование значительно расширило список SNPs, которые могут быть связаны с индивидуальным ответом на метформин, и добавило новую информацию о фармакогенетических факторах, влияющих на ответ к данному препарату при сахарном диабете второго типа.

ЗАКЛЮЧЕНИЕ

В настоящей работе на клиническом материале (PBMCs) российских индивидуумов был впервые выполнен расчет аллель-специфических событий по данным экспериментов ChIP-seq с антителами к гистоновым модификациям H3K4me3, H3K27ac и RNA-seq. В результате анализа была сформирована панель из 14796 rSNPs, расположенных в промоторных районах (± 1000 п.о. от старта транскрипции) 5132 генов. Полученная панель rSNPs раскрывает регуляторное значение большой группы SNPs, идентифицированных в исследованиях GWAS, а также служит независимым подтверждением регуляторной функции значительной группы eQTLs из каталога GTEx.

Применение нашей панели к независимым транскриптомным данным с целью поиска rSNPs в промоторах генов, вовлеченных в патогенез сахарного диабета типа 2 с осложнением в виде диабетической ретинопатии, позволило выявить большую группу rSNPs, потенциально связанных с развитием данного заболевания. В результате построения сетей PPI для генов, предположительно связанных с диабетом, были найдены хаб гены, которые являются важными регуляторами биологических процессов. Предположительно, rSNPs в промоторах хаб генов, влияя на их экспрессию, могут вносить значительный вклад в нарушения регуляции ключевых процессов, связанных с развитием диабета и его осложнений. Анализ обогащения построенной сети PPI для генов с повышенной экспрессией (при сравнении пациентов с сахарным диабетом типа 2 и здоровыми лицами) с помощью KEGG/GO/Reactome установил, что гены, в промоторах которых нами обнаружены rSNPs, участвуют в пути передачи инсулинового сигнала и в процессах иммунного ответа.

Применение нашей панели к независимым транскриптомным данным для поиска rSNPs в промоторах ДЭГ, связанных с ответом на сахароснижающий препарат метформин, позволило расширить список rSNPs, которые могут участвовать в индивидуальном ответе на данный антигипергликемический препарат, в том числе rSNPs, кодирующих транскрипционные факторы (*FOXP1*, *POU2F2*, *YY1* и т.д.) и другие регуляторные белки (*TAF3*, *NCOR1*, *SMARCA4*, *PCBD2*). Кроме того, нами были обнаружены rSNPs в промоторных районах ДЭГ, вовлеченных в уже известные механизмы действия метформина, включая гены, которые кодируют потенциальные мишени данного препарата (*NDUFA11*, *NDUFB1*) или являются вышестоящими регуляторами его мишеней (*NMT1*, *AK5*).

ВЫВОДЫ

1) На основании поиска аллель-специфических событий в полученных экспериментальных данных по профилям экспрессии генов (RNA-seq) и меток активного хроматина H3K4me3 и H3K27ac (ChIP-seq) в мононуклеарных клетках периферической крови (PBMCs) девяти здоровых людей сформирована панель из 14796 rSNPs, расположенных в промоторных районах 5132 генов.

2) Установлено, что 38.4% найденных rSNPs связано с различными фенотипическими признаками по данным каталога GWAS, что указывает на регуляторную роль 5590 SNPs из этого каталога. Показано, что 66.7% найденных rSNPs связаны с eQTLs по данным каталога GTEx, что служит независимым подтверждением регуляторной функции 12120 eQTLs из данного каталога.

3) При анализе независимых транскриптомных данных (GSE221521) выявлено 3810 rSNPs, которые расположены в промоторных районах генов, связанных с реализацией метаболических эффектов инсулина (PI3K/Akt/mTOR), а также с процессами активации лейкоцитов, дифференцировки мононуклеарных клеток, формирования резистентности к инсулину, что предполагает важную роль этих rSNPs в механизмах развития сахарного диабета типа 2.

4) При анализе независимых транскриптомных данных (GSE153315) выявлена группа из 367 rSNPs, расположенных в промоторных районах генов, потенциально связанных с молекулярными механизмами действия метформина. В их число вошли гены, кодирующие транскрипционные факторы (*FOXP1*, *POU2F2*, *YY1*) и другие регуляторные белки (*TAF3*, *NCOR1*, *SMARCA4*, *PCBD2*), а также гены, вовлеченные в известные механизмы действия препарата (*NDUFA11*, *NDUFB1*, *NMT1*, *AK5*).

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Damarov I.S.**, Korbolina E.E., Rykova E.Y., Merkulova T.I. Multi-Omics Analysis Revealed the rSNPs Potentially Involved in T2DM Pathogenic Mechanism and Metformin Response // *Int J Mol Sci*. 2024. V. 25. № 17. P. 9297.

2. 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 Genet Selektcii*. 2023. V.27. №6. P. 662-675.

3. Bryzgalov L. O., Korbolina E. E., **Damarov I. S.**, Merkulova T. I.. The functional insight into the genetics of cardiovascular disease: results from the post-GWAS study // *Vavilov Journal of Genetics and Breeding*. 2022. V. 26. №1. P. 65-73.

4. Korbolina 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. V. 22. № 14. P.7240.