

Сайк Ольга Владимировна

**ВЫЯВЛЕНИЕ АССОЦИАТИВНЫХ
СВЯЗЕЙ МЕЖДУ ЗАБОЛЕВАНИЯМИ И
ГЕНЕТИЧЕСКОЙ РЕГУЛЯЦИЕЙ
МЕТАБОЛИЧЕСКИХ ПРОЦЕССОВ НА
ОСНОВЕ АВТОМАТИЧЕСКОГО
АНАЛИЗА ТЕКСТОВ И БАЗ ДАННЫХ**

03.01.09 Математическая биология, биоинформатика

**Автореферат
диссертации на соискание ученой степени
кандидата биологических наук**

Новосибирск – 2018

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» в лаборатории компьютерной протеомики, г. Новосибирск, Россия.

Научный руководитель: **Иванисенко Владимир Александрович**
кандидат биологических наук, доцент, в.н.с.,
зав. лаб. компьютерной протеомики ФГБНУ ФИЦ
«Институт цитологии и генетики СО РАН»
(ИЦИГ СО РАН), г. Новосибирск

Официальные оппоненты: **Доронин Борис Матвеевич**
доктор медицинских наук, профессор,
заведующий кафедрой неврологии
Федерального государственного бюджетного
образовательного учреждения высшего образования
«Новосибирского государственного медицинского
университета» Министерства здравоохранения РФ
(ФГБОУ ВО НГМУ Минздрава России), г. Новосибирск

Марков Олег Владимирович
кандидат биологических наук,
н.с. лаборатории биохимии нуклеиновых кислот
ФГБНУ Институт химической биологии и
фундаментальной медицины СО РАН
(ИХБФМ СО РАН), г. Новосибирск

Ведущее учреждение: **ФГБУН Институт общей генетики
им. Н.И. Вавилова РАН, г. Москва**

Защита диссертации состоится «___» _____ 20__ г. на утреннем заседании диссертационного совета Д 003.011.01 на базе Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦИГ СО РАН), в конференц-зале Института по адресу: 630090, г. Новосибирск, проспект ак. Лаврентьева 10, тел. (383)363-49-06, e-mail: dissov@bionet.nsc.ru. С диссертацией можно ознакомиться в библиотеке ИЦИГ СО РАН и на сайте Института: www.bionet.nsc.ru

Автореферат разослан «___» _____ 2018 г.

Ученый секретарь
диссертационного совета,
доктор биологических наук



Т.М. Хлебодарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. В настоящее время согласно Международной Классификации Болезней (МКБ) выделяют около 20 тысяч различных заболеваний (Gersenovic, 1995), которые подразделяются на моно-факторные и мульти-факториальные. Большинство моно-факторных заболеваний связано именно с нарушением метаболических процессов. Однако, нарушение метаболических процессов также имеет огромное значение в патогенезе мульти-факториальных заболеваний (DeBerardinis, Thompson, 2012).

Реконструкция и анализ молекулярно-генетических сетей, описывающих патогенез заболеваний, является биоинформатической основой развития методов диагностики и лечения заболеваний. Ассоциативные генные сети представляют собой комплексные молекулярно-генетические сети, ассоциированные с какими-либо биологическими процессами, фенотипическими признаками или заболеваниями. По сравнению с генными сетями, оперирующими только молекулярно-генетическими объектами, в ассоциативных генных сетях используется расширенный состав объектов, в том числе молекулярно-генетические объекты (гены, микроРНК, белки, метаболиты и т.д.), а также биологические процессы, метаболические и транспортные пути, пути передачи сигналов, фенотипические признаки, заболевания, мутации, внутренние и внешние факторы. Такое свойство ассоциативных генных сетей позволяет наиболее полно представлять информацию о механизмах формирования сложных фенотипических признаков, а также о процессах, приводящих к развитию патологий.

На данный момент в научных публикациях и базах данных накоплено огромное количество фактов о молекулярно-генетических взаимодействиях и их связях с заболеваниями. Так система PubMed включает более 28 миллионов публикаций по биомедицинской тематике. Интернет-ресурс журнала Nucleic Acids Research, посвященный базам данных, на 2018 год перечисляет 1737 молекулярно-биологических баз данных, более ста из которых имеют непосредственное отношение к биомедицине (Rigden, Fernández, 2017). В настоящее время, для анализа такого большого объема информации активно используются методы автоматической экстракции знаний (Bundschuh *et al.*, 2008; Furlong, 2013).

Ассоциативные генные сети представляют собой эффективное средство для интеграции разнородных данных, описывающих механизмы развития заболеваний, представленных в статьях и базах данных. Анализ ассоциативных генных сетей позволяет проводить приоритизацию генов для решения задач генотипирования, выявлять потенциальные лекарственные мишени, предсказывать потенциальные биомаркеры заболеваний и т.д. На сегодняшний день существует множество методов и компьютерных программ, предназначенных для разностороннего анализа генных сетей, однако, анализ особенностей генетической регуляции метаболических процессов в генных сетях остается плохо разработанной областью.

Цель и задачи исследования. Целью настоящей работы являлось построение и анализ фреймовых моделей, представляющих собой паттерны в ассоциативных

генных сетях, описывающих связь заболеваний с генетической регуляцией метаболических процессов, выявленную на основе автоматического анализа текстов научных публикаций и баз данных.

Для достижения цели, были поставлены следующие задачи:

1. Создание базы данных Promedia, содержащей фреймовые модели, описывающие связь заболеваний с генетической регуляцией метаболических процессов на уровне взаимодействий между транскрипционными факторами, ферментами, белками-регуляторами, микроРНК и метаболитами, выявленную на основе автоматического анализа текстов научных публикаций и баз данных.

2. Анализ транзитивных ассоциаций между участниками фреймовых моделей, направленный на предсказание потенциальных биомаркеров заболеваний и перспективных кандидатов для генотипирования.

3. Анализ фреймовых моделей из базы данных Promedia, направленный на выявление молекулярно-генетических механизмов патогенеза ряда социально-значимых заболеваний, включая первичную открыто-угольную глаукому, лимфедему, преэклампсию, гипертонию, астму, болезнь Паркинсона и гепатит С.

4. Предсказание метаболитов – потенциальных летучих биомаркеров ревматоидного артрита с использованием ассоциативных фреймовых моделей.

5. Поиск генов-кандидатов из числа участников фреймовых моделей, важных для развития коморбидного состояния астмы и гипертонии.

Научная новизна работы. Новизна данной работы состоит в том, что впервые на основе масштабного автоматического анализа научных публикаций и баз данных была создана база данных Promedia (PROtein METabolite Disease Associations), содержащая фреймовые модели, описывающие связь генетической регуляции метаболических процессов с заболеваниями. Впервые на основе анализа более 2 млн. фреймовых моделей из базы Promedia показана статистическая значимость транзитивности ассоциаций между участниками фреймовых моделей. Найденные закономерности могут быть использованы для предсказания потенциальных биомаркеров заболеваний и перспективных кандидатов для генотипирования. Впервые на основе анализа фреймовых моделей предсказаны 17 метаболитов – потенциальных летучих биомаркеров ревматоидного артрита, включая ацетон, экспериментальная проверка которого имела хорошее согласие с предсказанием. Полученные результаты могут быть использованы для планирования экспериментов по выявлению неинвазивных биомаркеров для диагностики ревматоидного артрита. Комплексный анализ ассоциативных генных сетей позволил впервые предложить гены-кандидаты для генотипирования при коморбидных состояниях астмы/гипертонии. В качестве наиболее перспективных генов-кандидатов были предложены гены *IL10*, *TLR4* и *CAT*.

Теоретическая и практическая значимость работы. База данных фреймовых моделей Promedia, разработанная в диссертационной работе, может быть использована для предсказания потенциальных биомаркеров заболеваний и перспективных кандидатов для генотипирования. Построенные фреймовые модели

дополняют существующие сведения о механизмах, лежащих в основе различных заболеваний, что дает возможность для развития более совершенных средств диагностики и лечения заболеваний. Белки, гены, метаболиты и микроРНК, предложенные в работе в качестве наиболее перспективных кандидатов, могут быть использованы для экспериментальной проверки их связи с заболеваниями. Результаты, полученные в диссертационной работе, могут быть использованы для преподавания теории генных сетей.

Основные положения, выносимые на защиту.

- 1) Фреймовые модели из базы данных Promedia описывают генетическую регуляцию метаболических процессов, ассоциированных с заболеваниями, и позволяют проводить приоритизацию потенциальных биомаркеров заболеваний и кандидатов для генотипирования.
- 2) Оксид азота, ацетальдегид, метанол и ацетон имеют наибольший приоритет в качестве потенциальных биомаркеров для неинвазивной диагностики ревматоидного артрита.
- 3) Гены *IL10*, *TLR4* и *CAT* имеют наибольший приоритет в качестве генов-кандидатов для генотипирования при коморбидных состояниях астмы/гипертонии.

Публикации. По теме диссертации было опубликовано 19 работ, в том числе 8 статей в рецензируемых журналах из списка ВАК.

Апробация работы. Материалы диссертации были представлены на XVII российском национальном конгрессе "Человек и лекарство" (Москва, 2010), 10 всероссийской научно-пр. конференции «Интеллектуальный потенциал ученых России» (Барнаул, 2010), II всероссийской научной конференции с международным участием "Научное творчество XXI века" (Красноярск, 2010), конференции "Russian - German center for integrative biology and computation" (Новосибирск, 2011), школе молодых ученых «Биоинформатика и системная биология» (Новосибирск, 2011), форуме "INTERRA", секция биомедицина (Новосибирск, 2011), II Российском конгрессе с международным участием «молекулярные основы клинической медицины – возможное и реальное» (Москва, 2012), 8-ой международной конференции БГРС (Новосибирск, 2012), международной научно-практической конференции "Медицина XXI века" (Смоленск, 2013), школе молодых ученых "Системная биология и биоинформатика" (Новосибирск, 2013), международной конференции "Высокопроизводительное секвенирование в геномике" (Новосибирск, 2013), 9-ой международной конференции БГРС (Новосибирск, 2014), 8-ой всероссийской научно-практической конференции с международным участием «Молекулярная диагностика 2014» (Москва, 2014), 7-ом Российском симпозиуме «Белки и пептиды» (Новосибирск, 2015), 10-ой международной конференции БГРС (Новосибирск, 2016), международной конференции «Беляевские чтения» (Новосибирск, 2017), XVI Международной научно-практической конференции «Экспериментальные и теоретические исследования в современной науке» (Новосибирск, 2018).

Объём и структура работы. Материал диссертации изложен на 185

страницах, содержит 47 рисунков и 14 таблиц. Список литературы включает 435 источников. Диссертация состоит из введения, обзора литературы, описания материалов и методов, описания результатов работы в четырех главах, заключения, выводов, списка литературных источников.

Личный вклад автора. Основные результаты получены автором самостоятельно. Разработка структуры и создание интерфейса базы данных проводились совместно с к.т.н. Деменковым П.С. Анализ молекулярно-генетических сетей, ассоциированных с лимфедемой, проводился совместно с д.м.н. Нимаевым В.В. и Усмановым Д.Б. Анализ путей метаболизма ацетона проводился совместно с д.б.н. Мошкиным М.П.

Благодарности. Выражаю свою искреннюю благодарность научному руководителю диссертации к.б.н. Иванисенко В.А. за помощь на всех этапах выполнения диссертации. Благодарю к.т.н. Деменкова П.С. за консультации по вопросам создания базы данных. Выражаю благодарность академику, д.б.н. Н.А. Колчанову за ценные замечания и контроль за ходом работы над диссертацией. Искренне благодарю д.б.н. Мошкина М.П. за всестороннюю поддержку раздела работы, посвященного изучению метаболитов-летучих биомаркеров. Выражаю благодарность д.б.н. Мошкину М.П., академику, д.м.н. Козлову В.А., Саморокову С.Н., Балдину М.Н., член-корреспонденту РАН Грузнову В.М. и профессору д.м.н. Досенко В.Е. за предоставленные экспериментальные данные. Благодарю д.м.н. Нимаева В.В. и Усманова Д.Б. за поддержку раздела работы, посвященного изучению лимфедемы. Выражаю признательность Вяткину Ю.В., к.б.н. Антонцу Д.В. и к.ф.-м.н. Штокало Д.Н. за консультации по вопросам использования биоинформатического программного обеспечения, пакетов языков программирования Python и R. Выражаю благодарность д.б.н. Лихошваю В.А., д.б.н. Пономаренко М.П., Подколотному Н.Л. за ценные замечания и продуктивные научные дискуссии. Благодарю к.б.н. Долгову Е.В. за ценные замечания по тексту диссертации. Особую благодарность выражаю к.б.н. Игнатъевой Е.В. и к.б.н. Стручалину М.В. за рецензирование данной работы.

СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Обзор литературы

В мире существуют хорошо разработанные методы анализа научных публикаций и баз данных, которые применяются для реконструкции ассоциативных генных сетей, описывающих механизмы развития патологий (Jenssen *et al.*, 2001; Cheung *et al.*, 2012; Ivanisenko *et al.*, 2015). Ассоциативные генные сети широко используются для решения таких задач, как поиск биомаркеров, лекарственных мишеней, изучения молекулярных механизмов развития как отдельно взятых заболеваний, так и их комбинаций, определяющих коморбидные состояния патологий (Goh *et al.*, 2007; Rzhetsky *et al.*, 2007; Wang *et al.*, 2017). Новым перспективным направлением в неинвазивной диагностике заболеваний является поиск летучих химических соединений (Corradi, Mutti, 2005). В частности, такие соединения в качестве биомаркеров активно исследуются для разработки средств

диагностики нарушений сердечной деятельности (Phillips *et al.*, 2003; Pabst *et al.*, 2007), респираторных заболеваний (Corradi *et al.*, 2003a; Abba, 2009), сахарного диабета (Novak *et al.*, 2007), шизофрении (Brass, 1970), заболеваний желудочно-кишечного тракта (Probert *et al.*, 2004; Garner *et al.*, 2007) и др.

В настоящее время поиск биомаркеров в основном базируется на использовании экспериментальных подходов. В свою очередь анализ генных сетей открывает широкие возможности для предсказания биомаркеров (Childs, Valle, 2000; Jimenez-Sanchez *et al.*, 2001). Однако, существующие исследования по этой тематике проводились в основном с использованием сетей белок-белок взаимодействий (Goehler *et al.*, 2004; Rual *et al.*, 2005; Stelzl *et al.*, 2005; Lim *et al.*, 2006; Oldham *et al.*, 2006; Pujana *et al.*, 2007). Учет в генных сетях, описывающих молекулярные механизмы заболеваний, генетической регуляции метаболических процессов, в которых могут быть задействованы метаболиты, ферменты, транскрипционные факторы и микроРНК, остается до сих пор недостаточно разработанной областью, что подтверждает актуальность настоящей диссертационной работы.

Глава 2. Материалы и методы

2.1. Создание базы данных Promedia

При создании базы данных Promedia была использована реляционная система управления базами данных MySQL. Информация для заполнения базы данных Promedia была получена методом автоматического анализа текстов научных публикаций, реализованном в системе ANDSystem (Ivanisenko *et al.*, 2015), а также экстрагирована из баз данных. В работе использовались следующие базы данных, доступные для скачивания на январь-декабрь 2017 года: HMDB (Wishart *et al.*, 2017), SMPDB (Frolkis *et al.*, 2009), KEGG (Kanehisa *et al.*, 2016), ChEBI (Degtyarenko *et al.*, 2007), TRRUST (Han *et al.*, 2015), UniProt (UniProt Consortium, 2016), Brenda (Schomburg *et al.*, 2004), NCBI Gene, ClinVar (Landrum *et al.*, 2015), DisGeNET (Piñero *et al.*, 2015), miR2Disease (Jiang *et al.*, 2008), miRTarBase (Chou *et al.*, 2017), HMDD (Li *et al.*, 2013) и база данных ANDCell системы ANDSystem (Ivanisenko *et al.*, 2015). Экстракция необходимой информации из перечисленных выше источников и наполнение базы данных Promedia проводилось с помощью программ, написанных на языке программирования Python. Пользовательский интерфейс разрабатывался с использованием языка программирования PHP и языка гипертекстовой разметки HTML.

2.2. Анализ транзитивных ассоциаций заболеваний с другими объектами во фреймовых моделях

Анализ проводился для фреймовых моделей трех типов: (1) Фермент (участник 1) ↔ Метаболит (участник 2); (2) Транскрипционный фактор (участник 1) → Фермент (участник 2); (3) МикроРНК (участник 1) → Фермент (участник 2). Рассматривались только фреймовые модели, для которых в базе данных Promedia имелась информация об ассоциациях каждого из их участников хотя бы с одним заболеванием. Для описания алгоритма проверки значимости транзитивных ассоциаций введем следующие обозначения: набор заболеваний, ассоциированных

с участником 1 будем называть набором 1, а набор заболеваний, ассоциированных с участником 2 – набор 2. Статистическая проверка, того что доля общих заболеваний в наборах 1 и 2 для фреймовых моделей трех перечисленных типов из базы данных Promedia, превышает такое значение, ожидаемое по случайным причинам, проводилась с помощью генерации псевдослучайных фреймовых моделей. Псевдослучайные фреймовые модели строились следующим образом: (1) для каждой рассмотренной фреймовой модели из базы данных Promedia набор 1 заменялся на набор такого же размера заболеваний, случайно выбранных из базы данных Promedia. При этом в число случайно выбранных заболеваний могли попасть заболевания из набора 1. Дублирование заболеваний внутри набора 1 не допускалось; (2) рассчитывалась доля заболеваний общих для наборов 1 и 2. Для этого число заболеваний общих для наборов 1 и 2 было разделено на число заболеваний в наборе 1. Такая процедура проводилась как для фреймовых моделей из базы данных Promedia, так и для соответствующих им псевдослучайных фреймовых моделей; (3) с помощью *t*-критерия Стьюдента анализировалась статистическая значимость отличия среднего значения доли общих заболеваний для фреймовых моделей из базы данных Promedia от такого значения для псевдослучайных фреймовых моделей. Для расчета *t*-критерия Стьюдента использовалась функция `stats.ttest_ind` пакета `scipy`, реализованного на языке программирования Python (Oliphant, 2007). Дополнительно сравнение распределений было проведено с помощью непараметрического *U*-критерия Манна-Уитни, реализованного в пакете `scipy` (функция `stats.mannwhitneyu`) на языке программирования Python (Oliphant, 2007); (4) шаги 1-3 применялись для случая, когда в качестве случайного набора заболеваний брался набор 2; (5) проводился анализ различий распределений, полученных при выполнении шагов 1-3 и шага 4 независимо друг от друга. При статистической значимости отличия (превышения) доли общих заболеваний в наборах 1 и 2 для фреймовых моделей из базы данных Promedia, от такого значения для псевдослучайных фреймовых моделей, с $p\text{-value} < 0,01$ свойство транзитивности ассоциаций считалось выполненным.

2.3. Выявление участников фреймовых моделей, потенциально вовлеченных в патогенез ряда социально-значимых заболеваний

На первом шаге анализа из базы данных Promedia были экстрагированы фреймовые модели, включающие 4 участника. Отбирались такие фреймовые модели, в которых три участника имели ассоциативные связи с рассматриваемым заболеванием, согласно базе данных Promedia, а четвертый участник (кандидат) не имел таких связей. На втором шаге для каждого *i*-ого участника анализируемой фреймовой модели подсчитывалось число N_{fi} фреймовых моделей, ассоциированных с рассматриваемым заболеванием, в которых данный участник присутствовал. На третьем шаге рассчитывался показатель приоритета фреймовой модели P_j , как сумма значений N_{fi} для всех участников рассматриваемой фреймовой модели. На следующем шаге проводилось ранжирование фреймовых моделей согласно убыванию показателя P_j . Наибольшим приоритетом обладали

фреймовые модели с максимальным значением показателя P_j . Участники наиболее приоритетных моделей, для которых ранее не было показано наличие ассоциативных связей с изучаемым заболеванием, были предложены как кандидаты для дальнейшей экспериментальной проверки.

2.4. Поиск летучих метаболитов – потенциальных биомаркеров ревматоидного артрита

На основе информации из разработанной базы данных Promedia были построены фреймовые модели, участники которых были ассоциированы с ревматоидным артритом. Среди метаболитов, участвующих в построенных фреймовых моделях, были отобраны потенциально летучие. Потенциально летучими считали метаболиты с молекулярной массой менее 200 г/моль (Rowan, 2011; Abdullah et al., 2015; Куценогий и др., 2017) и температурой плавления ниже 0°C (температура плавления льда). Далее на основании информации из базы данных HMDB (Wishart et al., 2017) были выявлены метаболиты, обнаруживаемые в плазме крови. Потенциально летучие метаболиты, присутствовавшие в крови, считали наиболее перспективными кандидатами для исследования возможности их использования в качестве неинвазивных летучих биомаркеров ревматоидного артрита.

2.5. Выявление генов-кандидатов для генотипирования, потенциально вовлеченных в развитие коморбидного состояния астмы и гипертонии

Реконструкция ассоциативных генных сетей астмы и гипертонии проводилась с помощью системы ANDSystem (Ivanisenko et al., 2015). Сверхпредставленные биологические процессы Gene Ontology были выявлены с помощью сервиса DAVID 6.8 (Huang et al., 2008). **Критерий 1** рассчитывался с помощью системы для приоритизации генов Endeavour (Tranchevent et al., 2016) версии 3.71. На вход в качестве тестового и обучающего набора подавался список генов из полной сети астма/гипертония. **Критерий 2** рассчитывался с помощью системы для приоритизации генов TopGene (Chen et al., 2009). На вход по очереди подавались гены из полной сети астма/гипертония, а в качестве обучающего набора подавался список генов из полной сети астма/гипертония, из которого был исключен анализируемый ген. **Критерий 3** рассчитывался как суммарное количество сверхпредставленных биологических процессов из базы Gene Ontology, в которые анализируемый ген был вовлечен. Значения **критерия 4** рассчитывались для генов, как средний показатель центральности по степени (degree centrality), центральности по близости (closeness centrality) и центральности по посредничеству (betweenness centrality) для полной сети астма/гипертония. Для оценки центральности вершин в графах генных сетей использовались функции `nx.degree_centrality` для расчета центральности по степени, `nx.closeness_centrality` для расчета центральности по близости и `nx.betweenness_centrality` для расчета центральности по посредничеству из пакета `networkx`, реализованного на языке программирования Python (Hagberg et al., 2008). **Критерий 5** рассчитывался аналогично критерию 4 для генетической регуляторной сети астма/гипертония. Для расчета показателей **критерия 6** на первом шаге для астмы и гипертонии был

построен список биологических процессов, которые были связаны с астмой и гипертонией в системе ANDSystem. Рассматривались следующие типы связей: «association» и «regulation». Далее все биологические процессы, присутствующие в системе ANDSystem, были разбиты на 2 группы — тестовую, включающую 118 биологических процессов, ассоциированных одновременно с астмой и гипертонией, и контрольную: все остальные 13538 биологических процессов из ANDSystem. Для каждого из 205 генов/белков, ассоциированных одновременно и с астмой, и с гипертонией, была найдена связь с биологическими процессами с помощью ANDSystem. С помощью т-критерия Стьюдента оценивалась специфичность связи гена/белка с тестовой группой биологических процессов. Для расчета т-критерия Стьюдента использовалась функция `stats.ttest_ind` с параметром `equal_var=False` из пакета `scipy.stats`, реализованного на языке программирования Python (Millman, Aivazis, 2011). Поправка Бонферрони на множественное сравнение проводилась функцией `p.adjust(Y, "bonferroni")` пакета `stats`, реализованного на языке программирования R (Shaffer, 1995). Для расчета **критерия 7** с помощью разработанной базы данных Promedia были построены фреймовые модели, причем в каждой модели одновременно присутствовали участники, ассоциированные с астмой, и участники, ассоциированные с гипертонией. Значения критерия рассчитывались как число фреймовых моделей, в которых присутствовал анализируемый ген-кандидат. **Критерий 8** показывает наличие в гене полиморфизмов, ассоциированных либо с астмой, либо с гипертонией (значение критерия считалось равным 1). При отсутствии таких полиморфизмов значение критерия было равно 250. **Критерий 9** показывает наличие полиморфизмов, ассоциированных с какими-либо заболеваниями, коморбидными либо к астме, либо к гипертонии. По литературе вручную был составлен список заболеваний, коморбидных к астме, и список заболеваний, коморбидных к гипертонии. Для этого были просмотрены 196 публикаций из системы PubMed, найденных по запросу «asthma comorbid diseases», и 622 публикации, найденные по запросу «hypertension comorbid diseases», отфильтрованные по параметру «Free full text». Значение критерия было равно 1, если в анализируемом гене-кандидате присутствуют данные полиморфизмы, если нет – значение было равно 250. **Критерий 10** показывает наличие полиморфизмов, ассоциированных с каким-либо заболеванием, не включая заболевания из пунктов (критерий 8) и (критерий 9). Значение критерия было равно 1, если в анализируемом кандидате присутствуют данные полиморфизмы, если нет – значение было равно 250. Для расчета **критериев 8-10** использовалась информация об ассоциациях полиморфизмов с заболеваниями из баз данных SNPedia (Cariaso et al., 2011), miRdSNP (Bruno et al., 2012), GWAS catalog (MacArthur et al., 2016) и DisGeNET (Piñero et al., 2015, 2016). Считалось, что полиморфизм является ассоциированным с заболеванием, если такая информация была представлена хотя бы в одной из перечисленных баз данных. Для каждого гена **окончательный показатель приоритета** рассчитывался, как среднее значение рангов, построенных по критериям 1-10.

Глава 3. База данных Promedia

В данной работе для описания сложных систем метаболизма и его регуляции предлагается использование фреймовых моделей, представляющих собой паттерны в ассоциативных генных сетях, участники которых (гены, белки, микроРНК и метаболиты) имеют связи с заболеваниями. Схематичное представление фреймовых моделей показано на рисунке 1. В представленном примере транскрипционный фактор связан с набором заболеваний 1, а также с геном, кодирующим фермент, который в свою очередь связан с набором заболеваний 2. Фермент катализирует превращение метаболита-субстрата в метаболит-продукт. Каждый из метаболитов связан с собственным набором заболеваний. Для учета регуляции трансляции на рисунке показан объект микроРНК, который связан с набором заболеваний 3. Заболевания в рассмотренных наборах могут пересекаться.

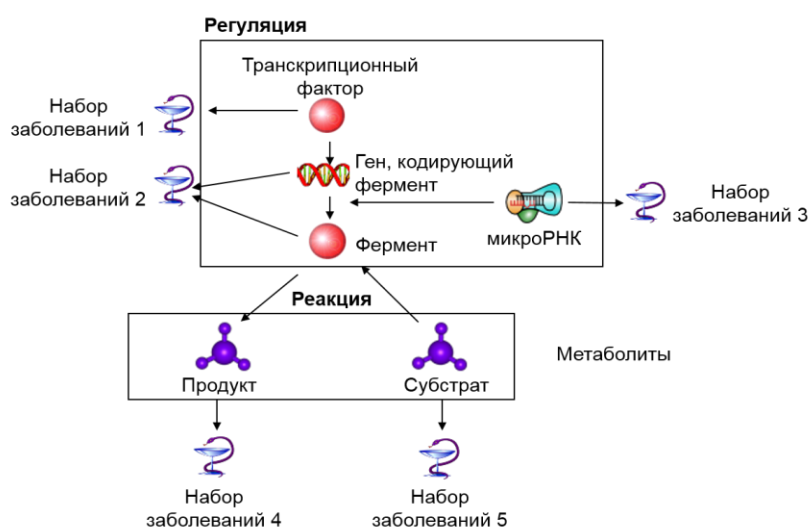


Рисунок 1. Схематичное представление ассоциативной связи генетической регуляции ферментативной реакции с заболеваниями с помощью фреймовых моделей.

Была разработана база данных Promedia, которая содержит фреймовые модели, описывающие ассоциации генетической регуляции метаболических процессов с заболеваниями (Сайк и др., 2010, 2011). Promedia интегрирует информацию, извлеченную из различных источников, включая фактографические базы данных, а также первоисточники в виде текстов научных публикаций (рисунок 2).

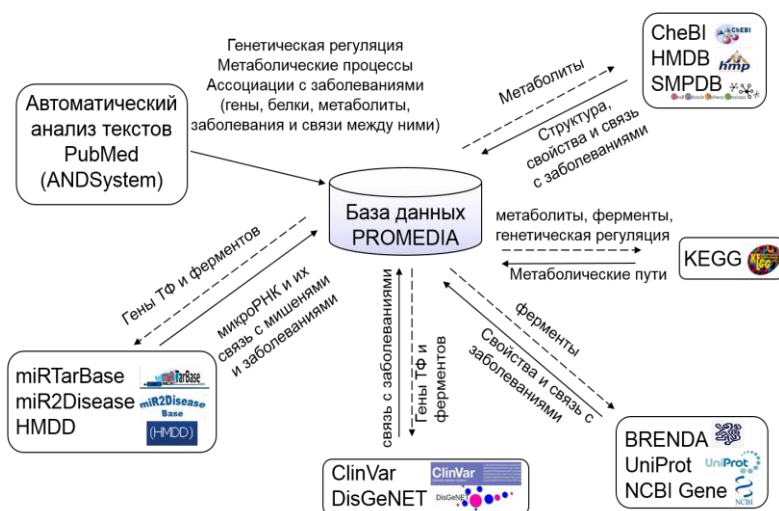


Рисунок 2. Схема извлечения знаний об ассоциативных связях заболеваний с генами, белками, микроРНК и метаболитами. Заполнение базы данных Promedia. Пунктирными стрелками обозначены запросы к базам данных. Сплошными стрелками обозначен процесс экстракции данных из баз.

Реляционная модель базы данных Promedia содержит 28 таблиц и включает разделы, посвященные метаболитам (133378 записей), ферментам (4229 записей), белкам-регуляторам (8391 запись), транскрипционным факторам (847 записей), микроРНК (2599 записей), а также заболеваниям (16266 записей). База данных Promedia включает более 700 млн фреймовых моделей для более 16 тысяч заболеваний. Анализ информации из базы данных Promedia показал, что наибольшее количество транскрипционных факторов, ферментов, метаболитов и микроРНК ассоциировано с раковыми заболеваниями, что согласуется с литературными данными (Duckett, Belldegrun, 1992; Duffy, 2013; Friedman et al., 2013; Pappalardo et al., 2013; Rebucci, Michiels, 2013). Пользовательский интерфейс базы данных Promedia доступен по адресу «<http://www-bionet.sccc.ru/promedia/>».

Глава 4. Анализ потенциальных транзитивных ассоциаций заболеваний с другими объектами во фреймовых моделях

Транзитивность подразумевает для элементов a , b , c из множества X выполнение отношений R , таких, что если aRb и bRc , то выполняется aRc (Schmidt, 2010). На основании этого транзитивной ассоциацией с заболеванием считается потенциальная связь между одним молекулярно-генетическим объектом и данным заболеванием, в случае, когда известна ассоциативная связь соседей этого объекта в генной сети с рассматриваемым заболеванием (Swanson, Smalheiser, 1997; Narayanasamy et al., 2004). В данной работе были сформулированы гипотезы о потенциальной транзитивности ассоциаций участников фреймовых моделей с заболеваниями:

1. **(А)** Если некое заболевание D ассоциировано с транскрипционным фактором TF и TF ассоциирован с ферментом EN , тогда D также ассоциировано с EN . **(Б)** И наоборот, если некое заболевание D ассоциировано с ферментом EN и EN ассоциирован с транскрипционным фактором TF , тогда D также ассоциировано с TF .
2. **(А)** Если некое заболевание D ассоциировано с микроРНК Mir и Mir ассоциирована с ферментом EN , тогда D также ассоциировано с EN . **(Б)** И наоборот, если некое заболевание D ассоциировано с ферментом EN и EN ассоциирован с микроРНК Mir , тогда D также ассоциировано с Mir .
3. **(А)** Если некое заболевание D ассоциировано с ферментом EN и EN ассоциирован с метаболитом Met , тогда D также ассоциировано с Met . **(Б)** И наоборот, если некое заболевание D ассоциировано с метаболитом Met и Met ассоциирован с ферментом EN , тогда D также ассоциировано с EN .

Анализ показал статистическую значимость (таблица 1) выполнения во фреймовых моделях транзитивных ассоциаций с заболеваниями таких объектов, как транскрипционные факторы, микроРНК, ферменты и метаболиты (Сайк и др., 2018).

Таблица 1. Статистическая значимость выполнения транзитивных ассоциаций во фреймовых моделях.

Гипотеза	Число проанализированных фреймовых моделей	Среднее значение доли общих заболеваний для фреймовых моделей Promedia (%)	Среднее значение доли общих заболеваний для псевдослучайных фреймовых моделей	t-критерий Стьюдента, p-value	U-критерий Манна-Уитни, p-value
1. (А)	844	34.58	5.21	10^{-148}	10^{-149}
1. (Б)	844	20.81	3.08	10^{-105}	10^{-107}
2. (А)	69852	1.12	0.04	10^{-230}	10^{-233}
2. (Б)	69852	2.77	1.29	10^{-196}	10^{-198}
3. (А)	28512	3.88	0.49	10^{-100}	10^{-102}
3. (Б)	28512	13.26	1.29	10^{-94}	10^{-96}

Свойство транзитивности ассоциаций было использовано для предсказания потенциальных биомаркеров заболеваний среди метаболитов, ферментов, микроРНК и транскрипционных факторов.

5. Анализ ассоциативных фреймовых моделей на примере конкретных социально-значимых заболеваний

С использованием данных базы Promedia были выявлены фреймовые модели, участники которых имели наибольшее количество ассоциаций с такими заболеваниями, как (1) первичная открыто-угольная глаукома (ПОУГ), (2) лимфедема, (3) преэклампсия, (4) гипертония, (5) астма, (6) болезнь Паркинсона и (7) гепатит С (Glotov et al., 2015; Saik et al., 2015, 2016, 2017, 2018; Сайк и др., 2016; Усманов и др., 2017; Янкина и др., 2018). На основе анализа фреймовых моделей с использованием свойства транзитивных ассоциаций были предложены новые гены/белки и метаболиты, которые могут быть потенциально ассоциированы с рассматриваемыми заболеваниями.

Например, для ПОУГ наиболее приоритетной оказалась фреймовая модель, представленная на рисунке 3, включающая транскрипционный фактор p53. Гибель клеток зрительного нерва путем апоптоза является одним из ключевых процессов патогенеза ПОУГ (Wang et al., 2015a). Транскрипционный фактор p53 является одним из основных регуляторов процесса апоптоза (Amaral et al., 2010), а полиморфизмы в данном гене ассоциированы с ПОУГ (Blanco-Marchite et al., 2011). Известно, что транскрипционный фактор p53 способен активировать экспрессию гена рецептора эпидермального фактора роста *EGFR* (Deb et al., 1994). При активации *EGFR* в экспериментальной модели глаукомы было показано снижение уровня деградации ганглиозных клеток сетчатки (Liu et al., 2006). Активация *EGFR* приводит к повышению уровня активности фермента циклооксигеназы-2 *COX-2* (Merchant et al., 2005). Показано, что при ПОУГ наблюдается снижение уровня экспрессии *COX-2* в эпителиальных клетках глаза (Barraza et al., 2010). Линолевая кислота в организме переводится в липидные гидропероксиды под действием

фермента COX-2 (Blair, 2001). В литературе не приводятся сведения об ассоциациях линолевой кислоты с ПОУГ, однако, в публикации Rong et al., 2017 говорится о связи этого метаболита с первичной закрыто-угольной глаукомой (Rong et al., 2017). Построенная нами фреймовая модель позволяет выдвинуть предположение о возможности существования связи между метаболитом «линолевая кислота» и заболеванием «первичная открыто-угольная глаукома».

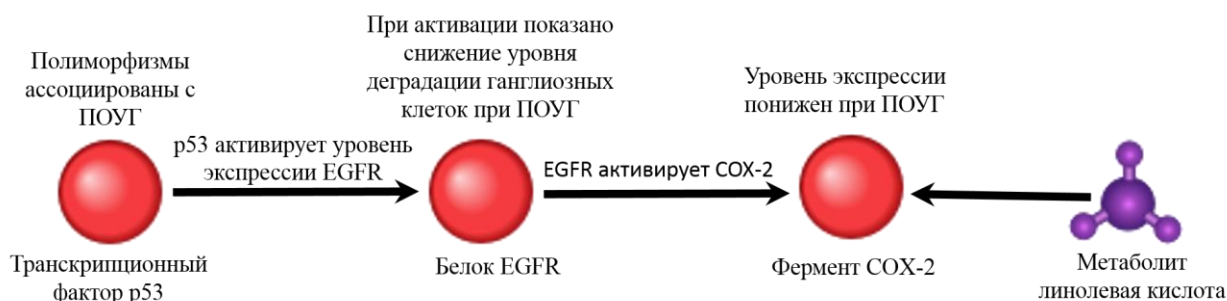


Рисунок 3. Фреймовая модель, связанная с ПОУГ, включающая транскрипционный фактор p53, белок EGFR, фермент COX-2 и метаболит линолевая кислота.

Таким образом, на основе анализа фреймовых моделей из базы данных Promedia были предсказаны потенциальные ассоциативные связи между заболеваниями и другими объектами ассоциативных генных сетей: (1) открыто-угольной глаукомой и метаболитом «линолевая кислота» (рисунок 3); (2) лимфедемой и микроРНК-451; (3) преэклампсией и транскрипционным фактором IRF3; (4) гипертонией и транскрипционными факторами FOS, HIF1A и CREB3L1; (5) астмой и микроРНК-125a; (6) болезнью Паркинсона и микроРНК-320; (7) гепатитом С и ферментом PLA2G4C.

6. Применение фреймовых моделей для планирования экспериментов по профилированию метаболитов и генотипированию

6.1. Поиск летучих метаболитов – потенциальных биомаркеров ревматоидного артрита

Ревматоидный артрит является хроническим заболеванием, вызванным аутоиммунными процессами в соединительной ткани, вследствие которых развивается полиартрит (множественное воспаление) мелких суставов (Kornø et al., 2010). На основе информации из базы данных Promedia было реконструировано 20320013 фреймовых моделей, ассоциированных с ревматоидным артритом. Всего в данных моделях участвовало 18507 метаболитов. Из них потенциально летучими с молекулярной массой менее 200 г/моль (Rowan, 2011; Abdullah et al., 2015; Куценогий и др., 2017) были 885. Согласно информации из базы данных HMDB (Wishart et al., 2017), 198 метаболитов могут быть обнаружены в плазме крови. Чем ниже температура плавления вещества, тем с большей вероятностью оно является летучим (Mullin et al., 1968; Kragl et al., 2002). В связи с этим, в качестве наиболее перспективных кандидатов, среди 198 выявленных метаболитов, были отобраны 17 метаболитов, температура плавления которых ниже 0°C (оксид азота, ацетальдегид,

метанол, ацетон, метиламин, диметиламин, формальдегид, D-лимонен, сероводород, аммиак, пропиленгликоль, углекислый газ, угарный газ, пропионовая кислота, хлорацетальдегид, 1,3-диаминопропан, пероксид водорода). В работе Thornadtsson et al., 2018 показано, что уровень оксида азота является летучим биомаркером воспаления и значительно снижен в выдыхаемом воздухе пациентов с ревматоидным артритом (Thornadtsson et al., 2018). Исследование образцов выдыхаемого воздуха пациентов с ревматоидным артритом, выполненное Мошкиным М.П. и соавт. показало статистически значимое снижение концентрации ацетона в выдыхаемом воздухе в ходе лечения (Сайк и др., 2011). В исследованиях Zabek et al., 2016 и Cheng et al., 2017 года была подтверждена возможность использования ацетона в качестве биомаркера при ревматоидном артрите. Было показано, что уровень ацетона у пациентов, страдающих данным заболеванием, повышен (Zabek et al., 2016; Cheng et al., 2017).

Таким образом, результаты анализа ассоциативных фреймовых моделей регуляции метаболизма ацетона при ревматоидном артрите согласуются с экспериментальными данными.

6.2 Поиск генов – потенциальных кандидатов для генотипирования, ассоциированных с коморбидным состоянием астмы и гипертонии

Заболевания считаются коморбидными, если они наблюдаются у одного пациента одновременно значительно чаще, чем можно ожидать по случайным причинам (Пузырев, 2015). Астма - хроническое воспалительное заболевание дыхательных путей, главными характеристиками которого являются гиперчувствительность дыхательных путей к различным стимулам и обратимая обструкция воздушного потока (Xu et al., 2017). В патогенезе гипертонии воспаление также играет важную роль (Montecucco et al., 2011; Pietri et al., 2015). Нарушения общих молекулярно-генетических путей, по-видимому, обуславливает коморбидность астмы и гипертонии (Xu et al., 2017). В работе Su et al., 2016 показано, что совместная встречаемость этих двух заболеваний у пациентов значительно выше, чем могло ожидать по случайным причинам, OR=1.66 [1.47, 1.88], p-value<0.00001 (Su et al., 2016). Учитывая сложность механизмов коморбидности, в работе применялся комплексный подход, включающий анализ фреймовых моделей и различные критерии приоритизации (всего 10 критериев, в том числе критерии, вычисленные с помощью программ Endeavour и TopGene).

Среди наиболее приоритетных генов выявлены: (1) *IL-10*, кодирующий противовоспалительный цитокин интерлейкин-10; (2) *TLR4*, кодирующий толл-подобный рецептор 4, участвующий во врожденном иммунитете; (3) *CAT*, кодирующий фермент каталазу (Saik et al., 2018). Известно, в генах *IL10*, *TLR4* и *CAT* были обнаружены полиморфизмы, ассоциированные с астмой, и полиморфизмы, ассоциированные с гипертонией (Lyon et al., 2004; Zhou et al., 2005; Mak et al., 2006; Timasheva et al., 2008; Zhang et al., 2011; Schneider et al., 2015).

В качестве примера рассмотрим связи *IL-10* с заболеваниями астма и гипертония, выявленные на основе анализа данных из базы Promedia. Известно, что

при астме наблюдается снижение уровня продукции IL-10 альвеолярными макрофагами (Chung, 2001). Кроме того, IL-10 может подавлять образование оксида азота (рисунок 4), который играет важную роль в воспалении дыхательных путей (Ogawa et al., 2008). Было показано, что повышение уровня IL-10 нормализует артериальное давление и эндотелиальную функцию при гипертонии (Tinsley et al., 2010; Harmon et al., 2015). Пример фреймовой модели, центральным объектом которой является IL-10, приведен на рисунке 4.

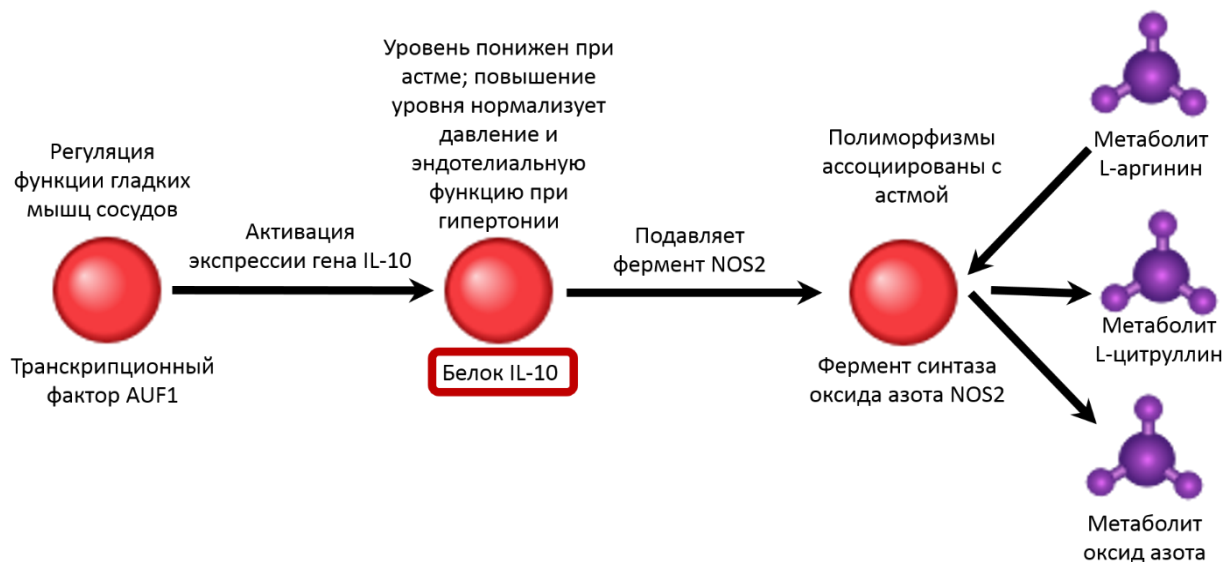


Рисунок 4. Фреймовая модель, связанная с астмой и гипертонией, включающая транскрипционный фактор AUF1, белок IL-10 (кандидат для генотипирования), фермент NOS2, метаболит-субстрат L-аргинин и метаболиты-продукты L-цитруллин и оксид азота.

Известно, что уровень экспрессии гена *IL-10* положительно регулируется транскрипционным фактором AUF1 (Sarkar et al., 2008). Показано, что AUF1 способен регулировать функцию гладких мышц сосудов через β_2 -адренергический рецептор при гипертонии (Babu et al., 2015). В свою очередь, IL-10 подавляет фермент синтазу оксида азота NOS2 (Fieger, 2007), который превращает L-аргинин в L-цитруллин и летучий метаболит оксид азота. Полиморфизмы в гене *NOS2* были ассоциированы с астмой (Holla et al., 2006; Islam et al., 2009), а в воздухоносных путях при астме наблюдался повышенный уровень оксида азота (Turner, 2007; Van Muylem et al., 2013).

Компьютерная реконструкция и анализ генных сетей позволяет выдвигать гипотезы о молекулярных механизмах заболеваний, а также оказывается эффективным инструментом для изучения таких сложных взаимоотношений между заболеваниями, как коморбидные состояния. На основе стандартных методов приоритизации (Endeavour и ToppGene), а также оригинальных критериев, использующих структуру генной сети астма/гипертония и фреймовые модели, были предложены гены-кандидаты для генотипирования и поиска лекарственных мишеней. Наибольшим приоритетом обладали гены *IL10*, *TLR4* и *CAT*, которые играют важную роль в иммунной системе и апоптозе (Saik et al., 2018).

Глава 7. Заключение

Согласно Международной Классификации Болезней (МКБ) выделяют около 20 000 различных заболеваний (Gersenovíc, 1995). Развитие методов диагностики и лечения заболеваний является весьма актуальной задачей. Одним из ключевых этапов при разработке методов лечения и диагностики заболеваний является выявление и исследование молекулярно-генетических механизмов, лежащих в основе патологических процессов. На настоящий момент разработан ряд моделей связей между заболеваниями и участниками молекулярно-генетических систем. Однако до сих пор не рассматривались ассоциативные фреймовые модели, описывающие взаимодействия между заболеваниями и генетической регуляцией метаболических процессов, включающие транскрипционные факторы, ферменты, микроРНК, метаболиты, их связи между собой и с заболеваниями. Следует также отметить, что, несмотря на обилие работ по выявлению связей белков, генов, полиморфизмов с заболеваниями и их представлению в виде молекулярно-генетических сетей, модели, включающие транскрипционные факторы, микроРНК и метаболиты, встречаются значительно реже. Вместе с тем метаболиты, транскрипционные факторы и микроРНК играют важную роль в развитии патологических процессов и могут иметь потенциальное значение для лечения и диагностики заболеваний. Поэтому построение ассоциативных фреймовых моделей, описывающих связь заболеваний с генетической регуляцией метаболических процессов является актуальной задачей.

В данной работе с использованием методов автоматического анализа текстов и баз данных, представляющих большие данные (big data), создана база данных Promedia (Сайк и др., 2010, 2011). База данных Promedia содержит ассоциативные фреймовые модели, описывающие связь между заболеваниями и процессами генетической регуляции метаболических процессов, реализуемых при участии микроРНК, транскрипционных факторов, ферментов и метаболитов. База данных может использоваться для планирования экспериментов по генотипированию и предсказания потенциальных биомаркеров заболеваний.

Было описано свойство транзитивности ассоциаций заболеваний с другими объектами в ассоциативных фреймовых моделях, которое может быть использовано для предсказания биомаркеров заболеваний (Сайк и др., 2018). В работе (Fang et al., 2008) рассматривается транзитивность отношений между заболеваниями, генами и препаратами традиционной китайской медицины. Однако, ранее свойство транзитивности ассоциаций заболеваний с другими объектами в фреймовых моделях для предсказания биомаркеров заболеваний не применялось.

На основании информации из разработанной базы данных Promedia были реконструированы ассоциативные фреймовые модели генетической регуляции метаболических процессов при различных патологиях, таких как первичная открыто-угольная глаукома, лимфедема, преэклампсия, гипертония, астма, болезнь Паркинсона, гепатит С и ревматоидный артрит (Glotov et al., 2015; Saik et al., 2015, 2016, 2017, 2018; Сайк и др., 2011, 2016; Усманов и др., 2017; Янкина и др., 2018),

хорошо согласующиеся с существующими представлениями о механизмах развития рассматриваемых патологий и дополняющие их. На основе анализа фреймовых моделей с использованием свойства транзитивности ассоциаций были предложены новые гены/белки, которые могут быть потенциально ассоциированы с рассматриваемыми заболеваниями.

Построенные фреймовые модели, ассоциированные с ревматоидным артритом, включающие ацетон, позволили предположить, что этот метаболит может быть потенциальным летучим маркером данного заболевания, что согласуется с экспериментальными данными о содержании ацетона в выдыхаемом воздухе у больных ревматоидным артритом (Сайк и др., 2011, Zabek et al., 2016, Cheng et al., 2017). Был проведен поиск генов-кандидатов, важных для коморбидных состояний астмы и гипертонии, на основе анализа генных сетей и фреймовых моделей. Согласно проведенному анализу, максимальным приоритетом обладали гены *IL10*, *TLR4* и *CAT* (Saik et al., 2018).

ВЫВОДЫ

- 1) Впервые с использованием методов автоматического анализа текстов научных публикаций и фактографических баз данных создана база данных фреймовых моделей Promedia, описывающих генетическую регуляцию метаболических процессов, ассоциированных с заболеваниями, в виде регуляторных паттернов в ассоциативных генных сетях, включающих ферменты, транскрипционные факторы, белки-транспортеры, белки-регуляторы активности/стабильности, микроРНК, метаболиты и заболевания. База данных включает более 700 млн фреймовых моделей для более 16 тысяч заболеваний.
- 2) Показана статистическая достоверность выполнения транзитивности ассоциаций заболеваний с объектами во фреймовых моделях (транскрипционными факторами, микроРНК, ферментами и метаболитами) при условии ассоциации соседей данных объектов во фреймовых моделях с рассматриваемыми заболеваниями. Выполнение этого отношения позволяет выдвигать гипотезы, которые могут быть использованы при планировании экспериментов по генотипированию, а также метаболическому профилированию заболеваний на основании информации о связи отдельно взятых участников фреймовых моделей с заболеваниями.
- 3) На основе анализа фреймовых моделей с использованием свойства транзитивности ассоциаций были предложены новые гены/белки и метаболиты, которые могут быть потенциально ассоциированы с рядом социально-значимых заболеваний: первичной открыто-угольной глаукомой (метаболит линолевая кислота), лимфедемой (микроРНК-451), преэклампсией (транскрипционный фактор IRF3), гипертонией (транскрипционные факторы FOS, HIF1A и CREB3L1), астмой (микроРНК-125a), болезнью Паркинсона (микроРНК-320) и гепатитом С (фермент PLA2G4C).
- 4) Впервые на основе анализа транзитивных ассоциаций во фреймовых моделях, ассоциированных с ревматоидным артритом, предложено 17 метаболитов, включая ацетон, являющихся потенциальными летучими маркерами этого заболевания.

Предсказание роли ацетона, как неинвазивного маркера ревматоидного артрита, согласуется с экспериментальными данными по изменению уровня ацетона в выдыхаемом воздухе пациентов.

5) Впервые на основе комплексного анализа ассоциативных генных сетей были предложены гены-кандидаты для генотипирования при коморбидных состояниях астмы/гипертонии, включая *IL10*, *TLR4* и *CAT*.

Список основных работ, опубликованных по теме диссертации:

1. **Сайк О.В.**, Мошкин М.П., Балдин М.Н., Грузнов В.М., Козлов В.А., Самороков С.Н., Деменков П.С., Иванисенко В.А., Колчанов Н.А. PROMEDIA–база данных химических соединений, потенциальных биомаркеров заболеваний, имеющих значение для неинвазивной диагностики // Математическая биология и биоинформатика. – 2011. – Т. 6. – №. 2. – С. 250-263.
2. Ivanisenko V.A., **Saik O.V.**, Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology // BMC systems biology. – 2015. – Т. 9. – №. 2. – С. S2.
3. Glotov A.S., Tiys E.S., Vashukova E.S., Pakin V.S., Demenkov P.S., **Saik O.V.**, Ivanisenko T.V., Arzhanova O.N., Mozgovaya E.V., Zainulina M.S., Kolchanov N.A., Baranov V.S., Ivanisenko V.A. Molecular association of pathogenetic contributors to pre-eclampsia (pre-eclampsia associome) // BMC systems biology. – 2015. – Т. 9. – №. S-2. – С. S4.
4. **Saik O.V.**, Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Interactome of the hepatitis C virus: Literature mining with ANDSystem //Virus research. – 2015. – Т.218. – С.40-48.
5. **Сайк О.В.**, Коновалова Н.А., Деменков П.С., Иванисенко Н.В., Иванисенко Т.В., Иванощук Д.Е., Пономарева М.Н., Коновалова О.С., Подколотная О.А., Лаврик И.Н., Колчанов Н.А., Иванисенко В.А. Молекулярно-генетические механизмы взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открытоугольной глаукоме // Вавиловский журнал генетики и селекции. – 2016. – Т. 20. – №. 6. – С. 840-847.
6. **Saik O.V.**, Konovalova N.A., Demenkov P.S., Ivanisenko T.V., Petrovskiy E.D., Ivanisenko N.V., Ivanoshchuk D.E., Ponomareva M.N., Konovalova O.S., Lavrik I.N., Kolchanov N.A., Ivanisenko V.A. Molecular associations of Primary Open-Angle Glaucoma with potential comorbid diseases (POAG-associome) // Biotecnología Aplicada. – 2016. – Т. 33. – №. 3. – С. 3201-3206.
7. Янкина М.А., **Сайк О.В.**, Деменков П.С., Хуснутдинова Э.К., Рогаев Е.И., Лаврик И.Н., Иванисенко В.А. Анализ взаимодействия генов нейронального апоптоза в ассоциативной генной сети болезни Паркинсона // Вавиловский журнал генетики и селекции. – 2018. – Т. 22. – №. 1. – С. 153-160.
8. **Saik O.V.**, Demenkov P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Goncharova I.A., Dosenko V.E., Zolotareva O.I., Hofestaedt R., Lavrik I.N., Rogaev E.I., Ivanisenko V.A. Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks // BMC medical genomics.– 2018. – Т.11. – №.1. – С.15.