

РЕКОНСТРУКЦИЯ ГЕННЫХ И АССОЦИАТИВНЫХ СЕТЕЙ ЗАБОЛЕВАНИЙ ДЛЯ ПОИСКА ГЕНОВ-МИШЕНЕЙ

Туркина В.А.¹, Ярема П.О.¹, Майорова А.А.¹, Орлова Н.Г.², Савина Е.А.^{1,3}, Орлов Ю.Л.^{1,4}

¹Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет)
ул. Трубецкая, 8, стр.2, г. Москва, 119991, РФ; e-mail: orlov@d-health.institute

²Финансовый Университет при Правительстве РФ
Ленинградский проспект, 49/2, г. Москва, 125167, РФ

³Институт молекулярной биологии им. В.А. Энгельгардта РАН
ул. Вавилова, 32, г. Москва, 119991, РФ

⁴Институт цитологии и генетики СО РАН

просп. ак. Лаврентьева, 10, г. Новосибирск, 630090, РФ

Поступила в редакцию 10.08.2023. DOI: 10.29039/rusjbpс.2023.0647

Аннотация. Компьютерная реконструкция генных сетей – наборов взаимодействующих макромолекул, объединенных общей функцией – представляет собой комплексную задачу биоинформатики. Генные сети как графические структуры представляют удобное средство визуализации, дают качественно новую информацию о наборе исследуемых генов, описание их биологических функций. В более широком плане мы рассматриваем ассоциативные генные сети заболеваний, куда входят не только гены и их продукты – белки, метаболиты, некодирующие РНК, лекарственные соединения, но и общие признаки заболеваний – симптомы, фенотипические проявления. Программные инструменты для компьютерной реконструкции графа генной сети активно развиваются в мире, используются в биомедицине. Мы рассматриваем онлайн-доступные инструменты биоинформатики для реконструкции генных сетей по списку генов и соответствующий компьютерный конвейер обработки данных. Анализ структуры генной сети заболевания позволяет выделить узловые гены, как мишени для терапии. Рассмотрены примеры применения онлайн-программ для анализа генной сети глиомы. Предложенный подход может быть масштабирован на другие функциональные наборы генов, связанные с заболеваниями человека. Используя подобные подходы биоинформатики, рассматривались генные сети комплексных заболеваний, таких как болезнь Паркинсона, деменция, шизофрения, рак молочной железы и другие виды рака.

Ключевые слова: биоинформатика, генные сети, глиома, комплексные заболевания, белок-белковые взаимодействия, медицинская информатика.

ВВЕДЕНИЕ

Компьютерная реконструкция генных сетей по списку генов, ассоциированных с тем или иным заболеванием, позволяет исследовать структуру сети, определять узловые гены и их продукты как возможные мишени терапевтических воздействий [1-3]. Генные сети как графические структуры представляют удобное средство визуализации, дают качественно новую информацию о наборе исследуемых генов, описание их биологических функций [4,5]. В мире активно разрабатываются базы данных и компьютерные инструменты, для анализа генных сетей, в том числе онлайн-доступные программы STRING (<https://string-db.org/>), GeneCards (<https://www.genecards.org/>), DAVID (<https://david.ncifcrf.gov/>), PANTHER, ANDVisio, ANDSystem [6-8].

Работа с онлайн программами биоинформатики, объединение их в общий конвейер для последовательного выполнения при исследовании заболеваний важно в медицинском образовании [3,9]. Использование онлайн-доступных программ незаменимо при дистанционном обучении, может эффективно использоваться для самостоятельных работ по биоинформатике [10]. Работа по анализу генов, реконструкции генной сети, и оформлению результатов с использованием инструментов биоинформатики была выполнена студентами на цифровой кафедре Сеченовского Университета (Первый МГМУ им. И.М. Сеченова Минздрава России) (<https://dk.sechenov.ru/>). Применение компьютерной реконструкции генных сетей студентами для задач биотехнологии выполнялось в Новосибирском государственном университете (<https://www-bionet.ssc.ru/icgenomics/>), РУДН, Школе биомедицины ДВФУ. Необходимо развитие собственных программных инструментов, интеграция баз данных по генным сетям, с применением в образовании, дистанционном обучении [11].

Используя онлайн инструменты биоинформатики как методический подход, рассматривались генные сети комплексных заболеваний, таких как болезнь Паркинсона [12], деменция, шизофрения [1], различные виды опухолей [2], метаболический синдром [13]. В данной работе рассмотрен пример анализа генной сети глиомы. Диффузные глиомы относятся к числу наиболее трудно поддающихся лечению видов рака [14]. Согласно клиническим рекомендациям, стандарт лечения включает в себя хирургическую резекцию опухоли с последующей лучевой и/или химиотерапией, а также использование немногочисленных таргетных методов лечения. Проблемой для эффективного лечения является гематоэнцефалический барьер, который препятствует

проникновению в головной мозг лекарственных средств. Необходимо развитие компьютерных моделей для анализа заболевания и поиска возможных мишеней для терапии.

Известно, что в генезе глиом участвует множество генетических мутаций, которые могут варьировать в зависимости от типа и степени злокачественности опухоли [15]. Одним из наиболее изученных генетических изменений, связанных с глиомами, является мутация гена IDH1. Она наблюдается у приблизительно 70% инфильтрирующих глиом и может прогнозировать более благоприятный исход для пациентов. Также было показано, что мутации в генах IDH1 и IDH2 могут влиять на ответ опухоли на химиотерапию [16]. IDH-мутировавшие глиомы располагаются преимущественно в лобной, височной доле или мозжечке, редко в промежуточном мозге или стволе головного мозга [17]. На МРТ такие глиомы, как правило, ограничиваются одной долей с большими участками поражений и более четкими границами, чем IDH-глиомы дикого типа [18].

Другим важным генетическим изменением является делеция на хромосоме 1p/19q, которая обнаруживается у приблизительно половины опухолей типа "олигодендроглиома" и является взаимоисключающей для астроцитом (TP53). Это изменение связано с более благоприятным прогнозом и может использоваться для выбора подхода к лечению [19].

Кроме того, в генезе глиом могут участвовать мутации в генах, связанных с сигнальными путями роста и пролиферации клеток, таких как гены EGFR, PTEN, TP53, PIK3CA и другие [15,20]. В очень небольшом количестве случаев в клетках с мутациями в генах IDH1 или IDH2 выявлены также изменения в генах PTEN, EGFR, CDKN2A и CDKN2B. Вместе с этим в образцах с немутантными IDH1 и IDH2 встречаемость мутаций гена TP53 была существенно ниже (18%), в то время как мутации генов PTEN, EGFR, CDKN2A и CDKN2B присутствовали намного чаще (74%) [21].

Помимо давно известных и долгое время изучаемых генов, формируются новые группы молекулярных маркеров (KDM1A, HIF1A, EGFR, SMO, MSI2), связанных с канцерогенезом и возможные для использования в таргетной терапии глиальных опухолей [22]. Хотя большинство глиом возникает спорадически без сопутствующего семейного заболевания, существует тесная связь между глиомами и несколькими редкими наследственными синдромами; на эти заболевания приходится менее 5% всех глиом. К таким редким синдромам относятся, прежде всего, синдромы факоматоза (нейрофиброматоз 1, туберозный склероз), синдром Ли-Фраумени, энхондроматоз, синдром семейной атипичной множественной родинки (FAMMM), синдром Рубинштейна-Тайби [23].

МАТЕРИАЛЫ И МЕТОДЫ

Материалами исследования являлись аннотации функций генов, последовательности и их функциональная разметка. Использовались данные из открытых баз данных NCBI, OMIM (<https://omim.org/>), GeneCards (<https://www.genecards.org/>).

Для получения списка генов, связанных с наследственной предрасположенностью к заболеванию, использовался интернет-ресурс OMIM (Online Mendelian Inheritance in Man) (<https://omim.org/>). Полнотекстовые обзоры со ссылками в OMIM содержат информацию обо всех известных менделевских расстройствах и более чем 16 000 генах. С помощью интегрированного ресурса GeneCards выполнялся поиск и анализ функции генов, играющих ключевую роль в развитии заболевания [24].

Для реконструкции генной сети взаимодействий генов использовался ресурс STRING-DB (<https://string-db.org/>). Взаимодействия включают прямые (физические) и косвенные (функциональные) ассоциации. Для поиска публикаций, описывающих взаимодействия генов, их продуктов, их функциональные ассоциации, использовался инструмент GeneMANIA (<http://genemania.org/>).

Для поиска ассоциированных заболеваний использовался открытый онлайн-ресурс MalaCards (<https://www.malacards.org/>) – интегративная база данных о болезнях и расстройствах организованная в виде "карточек болезней". Для анализа категорий генных онтологий использовались онлайн ресурсы международного консорциума Gene Ontology [25] и DAVID (<https://david.ncifcrf.gov/summary.jsp>). Анализ генных онтологий для списка генов выполняется с помощью ресурса PANTHER (<https://pantherdb.org/>). Инструмент выполняет анализ белков на основании эволюционных взаимосвязей. Для визуализации генных онтологий был использован онлайн-инструмент gGOST (<https://biit.cs.ut.ee/gprofiler/gost>), позволяющий построить график точечных значений категорий генных онтологий.

Дополнительно для расчета возможных трехмерных структур белков и определения их достоверности использовалась программа AlphaFold (<https://alphafold.ebi.ac.uk/>). Для целей дополнительной верификации структуры белков использовалась программу PAE viewer (<http://www.subtiwiki.uni-goettingen.de/v4/paeViewerDemo>).

Отметим, что настоящее время развиваются новые инструменты для анализа ассоциативных сетей – Cytoscape (<https://cytoscape.org/>), ANDVisio, ANDSystem [7,8], которые могут быть использованы в расширении исследования – поиске ассоциированных генов по литературе.

РЕЗУЛЬТАТЫ

Проведен анализ литературы с целью выявления генов предрасположенности к глиоме. Одним из ключевых генов, участвующих в трансформации нормальной клетки в опухолевую, является ген-супрессор TP53 [20]. Он

является важным клиническим маркером глиомы и, как известно, играет сложную роль в клетках глиомы. Функциональная его роль заключается в поддержании стабильности генома и генетической однородности клеток во всем организме. По мере роста и деления клеток постоянно возникают и накапливаются нарушения структуры ДНК. Это происходит как в результате естественного мутагенеза, так и ошибок в процессе репликации ДНК.

Эволюционно выработалась система устранения данных ошибок, включающая в себя цепь репарационных белков, которые работают лишь в определенных фазах клеточного цикла [26]. Один из них – белок p53, продукт гена TP53. Данный белок синтезируется во всех клетках организма и активируется при повреждении ДНК. Экспрессия TP53 в свою очередь вызывает задержку клеточного цикла либо для устранения повреждений и репарации ДНК, либо для индукции апоптоза. В норме данный механизм поддерживает целостность генома и препятствует приобретению опухолевого фенотипа [27]. Однако, при отсутствии гена TP53 либо наличии его дефекта, клетка не пересекает G1/S-контрольную точку фазы митоза, в результате чего не вступает в S фазу. Накапливаются популяции клетки с множественными нарушениями структуры ДНК, растет нестабильность генома, появляются все новые клоны клеток. В дальнейшем опухолевые клетки участвуют в своеобразном естественном отборе – они становятся все менее чувствительными к действию цитостатиков, с каждым поколением все сильнее нарушается контактное взаимодействие клеток, возникает метастазирование [28].

Мутантный TP53 обнаруживается уже на ранних стадиях заболевания. Установлено, что его мутации связаны с агрессивным течением онкологического заболевания и устойчивостью опухолевых клеток к химио- и лучевой терапии. Многие авторы отмечают, что высокий уровень экспрессии белка p53 связан с низкой общей выживаемостью пациентов [26,27]. Кроме того, мутировавший TP53 усиливает регуляцию MYC, EGFR, PNCA и снижает регуляцию p21, CD95Fas, PTEN [29].

Важную доказанную роль в патогенезе глиомы также играет рецептор эпидермального фактора роста (EGFR) [30]. EGFR (Epidermal Growth Factor Receptor) представляет собой трансмембранный гликопротеин, который принадлежит к семейству тирозинкиназных рецепторов (RTKs) и играет важную роль в процессах роста и развития клеток. Он был открыт в 1960 году как рецептор, способный связываться с эпидермальным фактором роста (EGF), регулирующий рост и деление эпителиальных клеток [30,31].

В последние годы также была идентифицирована группа микроРНК, участвующих в инициации глиомы путем нацеливания на ген RAS [32]. Ген RAS, который состоит из KRAS, NRAS и HRAS, является одним из наиболее часто активируемых онкогенов при раке человека. KRAS является преобладающим мутантным геном RAS при раке (84% всех миссенс-мутаций RAS), за ним следует NRAS (12%), HRAS мутирует редко (4%). Ras обладает активностью фермента GTP, который может гидролизовать GTP до GTP-протеина [32]. Ген KRAS (гомолог вирусного онкогена саркомы крысы Кирстен 2) представляет собой протоонкоген, который кодирует небольшой белок-трансдуктор GTPase под названием KRAS. Уже доказано, что полиморфизм гена KRAS связан с риском развития и прогрессированием глиомы. KRAS влияет на воспалительный компонент рака, активируя сигнальные пути MAPK и PI3K, что приводит к высвобождению дополнительных цитокинов IL-6/IL-8 и пролиферации раковых клеток [33, 34].

С помощью баз данных биоинформатики составлен список генов, ассоциированных с глиомой. Запрос “glioma” в базе данных OMIM дал 176 результатов - имен генов. Далее список был уточнен и аннотирован с помощью ресурсов GeneCards, GenBank NCBI.

Список 176 генов был загружен через интерфейс онлайн-ресурса PANTHER (<https://pantherdb.org/>) для поиска значимых категорий генных онтологий для этой группы генов. Использовались 3 основные группы генных онтологий (Gene Ontology): GO biological process - биологические процессы; GO molecular function - молекулярные функции; GO cellular component - клеточные компартменты. В качестве референсного генома выбран *Homo sapiens*. Гены микроРНК не были распознаны программой. Программа производила расчет по 144 идентификаторам. Всего в референсном геноме человека использовалось 20589 генов. Применялась коррекция оценки статистической значимости на множественность гипотез по Бонферрони (<https://pantherdb.org/>).

Было выполнено три аналогичных анализа для групп генных онтологий с ограничением значения уровня значимости P-Value до E-10 для биологических процессов, E-02 для молекулярных функций и E-03 для клеточных компартментов для представления наиболее информативных результатов. Анализ категорий генных онтологий с помощью ресурсов DAVID (<https://david.ncifcrf.gov/tools.jsp>), PANTHER (<https://pantherdb.org/>) показал, что самыми значимыми категориями являются: негативная регуляция биологического процесса, регуляция программируемой клеточной гибели, негативная регуляция клеточных процессов, регуляция клеточной гибели. Эти категории относятся к общему клеточному развитию, что связано с нормальным функционированием клеток и тканей. Значимыми категориями являются: связывание идентичных белков, активность димеризации белков, активность гомодимеризации белков, связывание сигнальных рецепторов. Эти категории можно отнести к работе путей передачи межклеточных сигналов, что связано с работой нервной системы.

Такой же список генов был загружен через интерфейс DAVID Bioinformatics Resources (<https://david.ncifcrf.gov/tools.jsp>) для поиска значимых категорий генных онтологий. Было идентифицировано 157 генов. Результаты представлены в обобщенной таблице 1.

Таблица 1. Список категорий генных онтологий по DAVID для глиомы

Группа онтологий	Категория онтологий	Число генов	P-value *
GOTERM_BP_DIRECT	negative regulation of neuron apoptotic process	13	2,60E-06
GOTERM_BP_DIRECT	negative regulation of apoptotic process	19	6,00E-05
GOTERM_BP_DIRECT	cellular response to hypoxia	10	3,60E-04
GOTERM_BP_DIRECT	regulation of cell cycle	12	2,40E-03
GOTERM_BP_DIRECT	positive regulation of transcription, DNA-templated	19	2,70E-03
GOTERM_BP_DIRECT	ovarian follicle development	6	3,80E-03
GOTERM_BP_DIRECT	neuron apoptotic process	7	5,10E-03
GOTERM_MF_DIRECT	identical protein binding	30	6,10E-03
GOTERM_BP_DIRECT	cell aging	5	7,20E-03
GOTERM_MF_DIRECT	protein homodimerization activity	18	6,10E-03
GOTERM_BP_DIRECT	negative regulation of cell proliferation	14	7,40E-03
GOTERM_BP_DIRECT	positive regulation of pri-miRNA transcription from RNA polymerase II promoter	6	7,40E-03
GOTERM_BP_DIRECT	positive regulation of peptidyl-serine phosphorylation	7	9,80E-03

Примечание к таблице: * **P-value** было нормировано с поправкой на множественность гипотез по Бенджамини-Хохбергу.

В построенных таблицах категорий генных онтологий показано присутствие категорий, связанных с апоптозом, что соответствует представлениям о заболевании.

Генная сеть для генов глиомы была реконструирована с помощью ресурса STRING-DB и собственной разработанной программы по визуализации сетей (Дипломные работы студентов по цифровой кафедре Сеченовского университета 2023 года, <https://dk.sechenov.ru/>). Граф включал 150 генов (белков). Гены микроРНК не были распознаны программой. Сеть была перестроена с удалением не связанных генов и заменой нескольких линий связей между узлами сети одной линией с учётом только доказанных взаимодействий со средним уровнем достоверности (параметр STRING-DB, см. <https://string-db.org/>).

Статистика общей построенной сети белок-белковых взаимодействий глиомы: количество узлов – 150; количество ребер - 590; средняя степень связности узла – 7.87; средний локальный коэффициент кластеризации – 0,478; ожидаемое количество ребер - 237; значимость плотности белок-белковых взаимодействий - $< 1,0e-16$. В центре сети находятся белки EGFR и TP53, которые являются известными онкогенами, а также STAT3, KRAS, PIK3CA, IDH1, KDR. Для более детального анализа сеть была перестроена в программе STRING (<https://string-db.org/>) итеративно с учетом только экспериментально доказанных взаимодействий со средним уровнем достоверности (рис. 1).

Несмотря на то, что часть объектов сети не контактировали с другими, сеть достаточно связанная, в ней выделяется центральный сильно связанный кластер генов (здесь и далее, для сети, построенной с помощью STRING, мы говорим о белковых продуктах генов). Это означает, что гены, по крайней мере, частично биологически связаны как группа взаимодействующих макромолекул.

В сети (рис. 1) можно видеть несколько кластеров. Наибольшее число связей - у генов EGFR, TP53, KRAS, из которых EGFR и TP53 являются ранее подробно изученными онкогенами. Сеть была перестроена с удалением не связанных генов и заменой нескольких линий связей между узлами сети в визуализации STRING только одной линией с учётом только доказанных взаимодействий со средним уровнем достоверности. В центре сети - белки EGFR и TP53, которые являются известными онкогенами, а также STAT3, KRAS, PIK3CA, IDH1, KDR.

Для более детального анализа сеть была перестроена итеративно с учетом только экспериментально доказанных взаимодействий. В сети выделяются несколько кластеров. Самый большой кластер включает 7 генов: BCL2, BRCA2, BRE, CHEK2, LZMBTL1, PTEN, TP53. Следующие по количеству вовлеченных узлов кластеры включают гены DNAJA3, DOCK4, EGFR, NF2, PIK3CA, PTPN11 и BRAF, KRAS, LZTR1, NF1 соответственно (рис. 2).

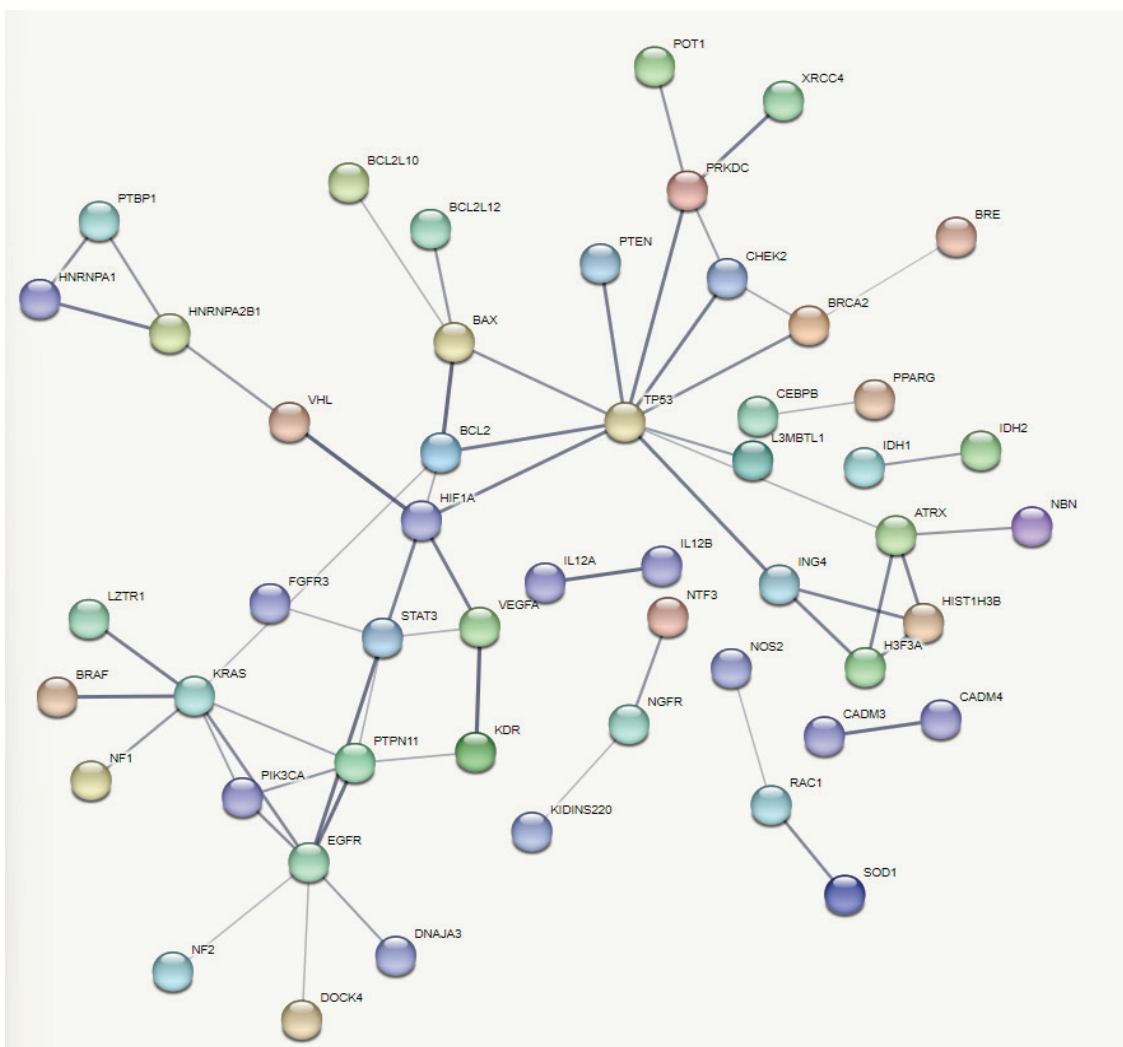


Рисунок 1. Итеративно перестроенная генная сеть генов глиомы (иллюстрация построена с помощью ресурса STRING)

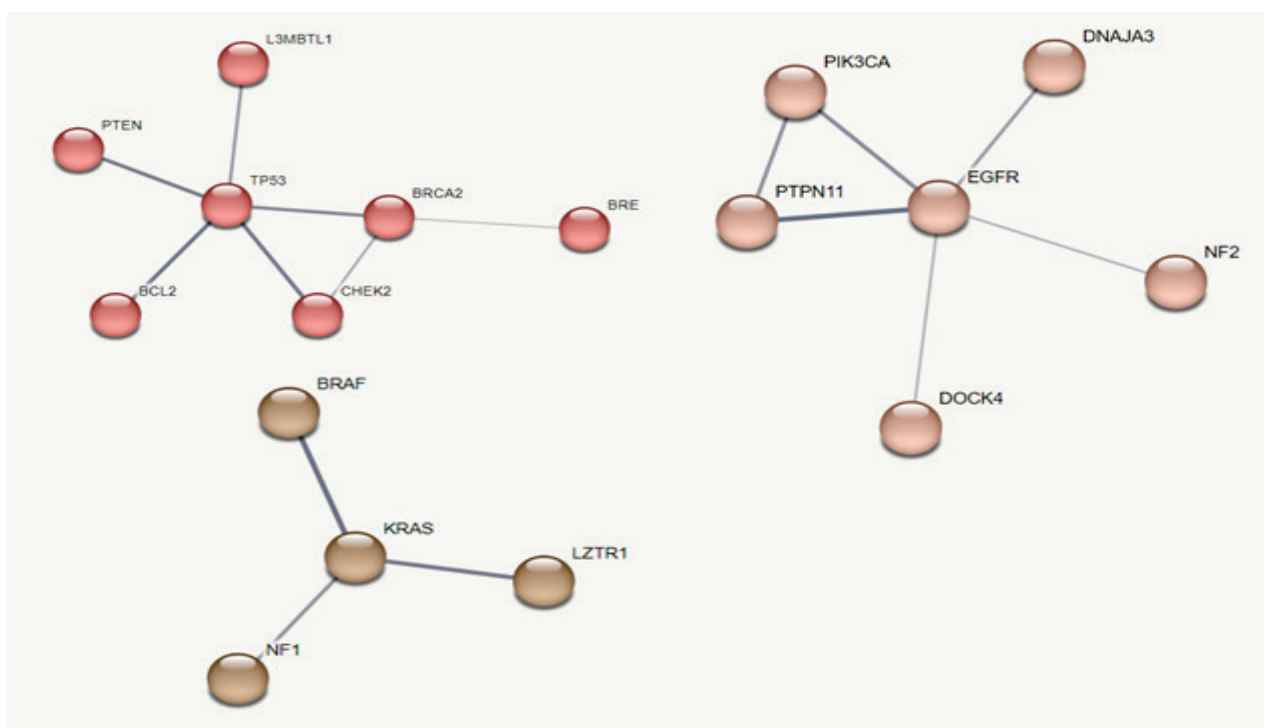


Рисунок 2. Кластеры генной сети глиомы, с узловыми генами TP53, EGFR, KRAS (иллюстрация построена с использованием ресурса STRING)

Наибольшее число связей у генов EGFR, TP53, KRAS, при этом EGFR и TP53 [26] ранее подробно изученные онкогены.

Далее, с помощью ресурса MalaCards (<https://www.malacards.org/>) были определены заболевания, ассоциированные с глиомой. Это глиобластома, и другие опухоли мозга – glioblastoma, malignant astrocytoma, neuroblastoma.

Таким образом, с помощью набора онлайн инструментов биоинформатики была собрана актуальная научная литература по заболеванию глиомой [3], описана роль отдельных генов и их продуктов. Несмотря на то, что глиома занимает небольшую долю от общего числа злокачественных новообразований ЦНС, на сегодняшний день она остается одним из самых агрессивных и тяжелых онкологических заболеваний, которое не поддается эффективному лечению и длительному уходу в ремиссию. Полученная структура геной сети может быть использована для поиска лекарственных средств, как основа для машинного обучения, дающего дополнительную информацию и функциональную аннотацию [35]. Исследование геной сетей, их компьютерная реконструкция – активно развивающееся научное направление [2,9], представленное в нашей работе на примере глиомы.

ЗАКЛЮЧЕНИЕ

В целом, в результате работы получен список генов, ассоциированных с глиомой, с использованием баз данных и онлайн-инструментов биоинформатики, определены значимые категории геной онтологий генов глиомы, показана их связь с онкологическими заболеваниями и реконструирована геной сеть заболевания [2,3]. Выявлены гены PIK3CA, KRAS, которые потенциально могут быть использованы как мишени для таргетной терапии, а также как молекулярно-генетические маркеры заболевания. Рассмотрены трехмерные структуры белков Pik3CA, KRas, и связывающиеся с ними лиганды [3]. Определение генов глиомы может базироваться на экспериментальных данных транскриптомного секвенирования [36,37].

Разработанный конвейер для компьютерной реконструкции геной сетей представляет набор интернет-доступных и частично заменяемых программных инструментов – для составления списка генов – OMIM (<https://omim.org/>), GenBank NCBI (<https://www.ncbi.nlm.nih.gov/gene>) и GeneCards (<https://www.genecards.org/>), для анализа геной онтологий – DAVID (<https://david.ncifcrf.gov/tools.jsp>), PANTHER (<https://pantherdb.org/>), для реконструкции геной сети по списку генов – STRING-DB (<https://string-db.org/>), Metascape (<https://metascape.org>), GeneMANIA (<http://genemania.org/>), MalaCards (<https://www.malacards.org/>), следуя стандартам, предложенным в работах [2,9]. Варианты анализа генов комплексных заболеваний используя новую версию [38] отечественной программы ANDDigest (<https://anddigest.sysbio.ru/>) [7] использовались для анализа генов карциномы печени [39].

Исследование генетической составляющей заболевания начинается с определения релевантных генов по базам данным. Поиск биомедицинской литературы в PubMed, Scopus, ScienceDirect для анализа заболеваний также является активной областью разработки программных средств, используя различные платформы [40,41].

Отметим появление новых онлайн инструментов составления и анализа списка генов, дополняющих стандартный поиск – Geneshot и Enrichr (<https://maayanlab.cloud/geneshot/>). Реконструкция геной сети по списку генов также может быть эффективно выполнена новыми средствами, такими как NDEx (<https://www.ndexbio.org/>). Необходима дальнейшая интеграция баз данных биоинформатики с помощью графов для сопоставления данных, получения интегрированной информации по аннотации функций генов, РНК и белков – отметим новый онлайн инструмент BioGraph (<http://andromeda.matf.bg.ac.rs:54321/>) [42].

Представленные подходы анализа геной сетей и их применения широко обсуждались на серии конференций по биоинформатике и регуляции структуры генома, и системной биологии – Bioinformatics of Gene Regulation and Structure/Systems Biology (BGRS/SB), проходящей в Новосибирске с 1998 года (<https://bgrssb.icgbio.ru/2022/>), и представленной в специальных выпусках международных научных журналов *International Journal of Molecular Sciences* ("25 Anniversary of Bioinformatics of Genome Regulation and Structure Conference Series", https://www.mdpi.com/journal/ijms/special_issues/0LGA6103S5), *Frontiers in genetics* (<https://www.frontiersin.org/research-topics/40408/bioinformatics-of-genome-regulation-and-systems-biology-volume-iii>) [43], *Journal of Bioinformatics and Computational Biology* (<https://dblp.org/db/journals/jbcb/jbcb11.html>) [44], *Journal of Integrative Bioinformatics* (<https://www.degruyter.com/journal/key/jib/17/4/html>) [45,46].

Институтом цифровой медицины Сеченовского университета проводится разработка программных инструментов для медицинской информатики [47,48], работает новый формат дополнительного образования – Цифровая кафедра (<https://dk.sechenov.ru/>). В 2023 году на Кафедре защищено несколько дипломов по разработке программных средств реконструкции и визуализации геной сетей, и анализу их структуры. Применение программного конвейера инструментов биоинформатики для глиомы представлено в данной работе. Использование геной сетей как параметра при создании предсказательных моделей машинного обучения показало свою эффективность для поиска лекарственных средств [49,50] и должно развиваться далее, опираясь на отечественные программные ресурсы [6,38].

Благодарности. Работа поддержана грантом бюджетным проектом ИЦиГ СО РАН FWNR-2022-0020 "Системная биология и биоинформатика: реконструкция, анализ и моделирование структурно-функциональной организации и эволюции геной сетей человека, животных, растений и микроорганизмов". Авторы признательны П.А. Иванову-Ростовцеву и А.Ю. Потаповой за техническую помощь.

Список литературы / References:

1. Дохоян А.Ю., Глущенко М.В., Орлов Ю.Л. Реконструкция геновой сети шизофрении для поиска генов-мишеней. *Ульяновский медико-биологический журнал*, 2022, т. 3, с. 6-22, doi: 10.34014/2227-1848-2022-3-6-22 [Dokhoyan A.Y., Glushchenko M.V., Orlov Y.L. Reconstruction of schizophrenia gene network to search for target genes. *Ulyanovsk Medical Biological Journal*, 2022, vol. 3, pp. 6-22 (In Russ.)].
2. Gubanov N.V., Orlova N.G., Dergilev A.I., Oparina N.Y., Orlov Y.L. Glioblastoma gene network reconstruction and ontology analysis by online bioinformatics tools. *Journal of Integrative Bioinformatics*, 2021, vol. 18, p. 20210031, doi: 10.1515/jib-2021-0031.
3. Туркина В.А., Майорова А.А., Дергилев А.И., Красильникова А.А., Ланских Д.В., Кумейко В.В., Орлов Ю.Л. Компьютерный анализ структуры геновой сети глиобластомы и опухолей мозга. *Сборник научных трудов VII Съезда Биофизиков России*: в 2 томах, том 1 - Краснодар: Типография ФГБОУ ВО «КубГТУ», 2023, с. 294-295, doi: 10.26297/SbR6.2023.001 [Turkina V.A., Maiorova A.A., Dergilev A.I., Krasilnikova A.A., Lanskih D.V., Kumeiko V.V., Orlov Yu.L. Computer analysis of the structure of the gene network of glioblastoma and brain tumors. Collection of scientific papers of the VII Congress of Biophysicists of Russia: in 2 volumes, vol. 1 - Krasnodar: Printing house of FGBOU VO "KubGTU", 2023, pp. 294-295 (In Russ.)].
4. Anashkina A.A., Leberfarb E.Y., Orlov Y.L. Recent Trends in Cancer Genomics and Bioinformatics Tools Development. *International Journal of Molecular Sciences*, 2021, vol. 22, p. 12146, doi: 10.3390/ijms22212146.
5. Klimontov V.V., Koshechkin K.A., Orlova N.G., Sekacheva M.I., Orlov Y.L. Medical Genetics, Genomics and Bioinformatics-2022. *International Journal of Molecular Sciences*, 2023, vol. 24, no. 10, p. 8968, doi: 10.3390/ijms24108968.
6. Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol.*, 2011, vol. 11, no. 3-4, pp. 149-161, doi: 10.3233/ISB-2012-0449.
7. Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*, 2020, vol. 21, suppl. 11, p. 228, doi: 10.1186/s12859-020-03557-8.
8. Ivanisenko V.A., Saik O.V., Ivanisenko N.V. et al. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol.*, 2015, vol. 9, suppl. 2, p. S2, doi: 10.1186/1752-0509-9-S2-S2.
9. Орлов Ю.Л., Шадеркин И.А., Орлова Н.Г., Рябков И.В., Лебедев Г.С. Цифровизация медицинского образования и телемедицина. *Физико-химические методы в междисциплинарных экологических исследованиях*. Всероссийский симпозиум и школа-конференция молодых ученых, 27 октября - 3 ноября, 2021, Севастополь, Россия. Сборник трудов симпозиума. М.: Издательский дом "Граница", 2021, 355 с. 340-341 [Orlov Y.L., Shaderkin I.A., Orlova N.G., Ryabkov I.V., Lebedev G.S. Digitalization of Medical Education and Telemedicine. *Physico-Chemical Methods in Interdisciplinary Environmental Research*. All-Russian Symposium and School-Conference of Young Scientists, October 27 - November 3, 2021, Sevastopol, Russia. Symposium Proceedings. Moscow: "Granitsa" Publishing House, 2021, pp. 340-341 (In Russ.)].
10. Orlova N.G., Orlov Y.L. Problems of developing online training courses for students in digital disciplines using bioinformatics as an example. In: *Proceedings of the International Conference "Scientific research of the SCO countries: synergy and integration"*. Scientific publishing house Infinity, 2022, pp. 58-65.
11. Орлов Ю.Л., Анашкина А.А., Рябков И.В., Лисненко А.А., Лебедев Г.С. Цифровизация образования в медицине: дистанционное преподавание биоинформатики и научные исследования с использованием интернет-технологий. *Наука в инновационном процессе*: Материалы Международной научно-практической конференции (Москва, 1–2 декабря 2021 г.). М.: ИПРАН РАН, 2021, с. 222-226, doi: 10.37437/9785912941719-21-ss2 [Orlov Y.L., Anashkina A.A., Ryabkov I.V., Lisenko A.A., Lebedev G.S. Digitalization of Medical Education: Remote Teaching of Bioinformatics and Scientific Research Using Internet Technologies. *Science in the Innovative Process*: Proceedings of the International Scientific and Practical Conference (Moscow, December 1–2, 2021). Moscow: IPHRAS RAS, 2021, pp. 222-226 (In Russ.)].
12. Орлов Ю.Л., Галиева А.Г., Орлова Н.Г., Иванова Е.Н., Мозылева Ю.А., Анашкина А.А. Реконструкция геновой сети болезни Паркинсона для поиска генов-мишеней. *Биомедицинская химия*, 2021, т. 67, вып. 3, с. 222-230, doi: 10.18097/PBMC20216703222 [Orlov Y.L., Galieva A.G., Orlova N.G., Ivanova E.N., Mozylova Y.A., Anashkina A.A. Reconstruction of the Parkinson's disease gene network for target gene identification. *Biomedical Chemistry*, 2021, vol. 67, iss. 3, pp. 222-230 (In Russ.)].
13. Тийс Р.П., Осипова Л.П., Галиева Э.Р., Личман Д.В., Воронина Е.Н., Мелихова А.В., Орлов Ю.Л., Филипенко М.Л. Полиморфизм вариантов гена N-ацетилтрансферазы 2 (NAT2) и анализ геновой сети. *Биомедицинская химия*, 2021, т. 67, вып. 3, с. 213-221, doi: 10.18097/PBMC20216703213 [Tiis R.P., Osipova L.P., Galieva E.R., Lichman D.V., Voronina E.N., Melikhova A.V., Orlov Y.L., Filipenko M.L. Polymorphism of N-acetyltransferase 2 (NAT2) gene variants and gene network analysis. *Biomedicinskaya himiya*, 2021, vol. 67, no. 3, pp. 213-221 (In Russ.)].
14. Osborn A.G., Louis D.N., Poussaint T.Y., Linscott L.L., Salzman K.L. The 2021 World Health Organization Classification of Tumors of the Central Nervous System: What Neuroradiologists Need to Know. *AJNR Am. J. Neuroradiol.*, 2022, vol. 43, no. 7, pp. 928-937, doi: 10.3174/ajnr.A7462.

15. Byun Y.H., Park C.-K. Classification and Diagnosis of Adult Glioma: A Scoping Review. *Brain & Neurorehabilitation*. Korean Society for Neurorehabilitation, 2022, vol. 15, no. 3, pp. 3-5.
16. Chang S., Yim S., Park H. The cancer driver genes IDH1/2, JARID1C/ KDM5C, and UTX/ KDM6A: crosstalk between histone demethylation and hypoxic reprogramming in cancer metabolism. *Exp Mol Med.*, 2019, vol. 51, no. 6, pp. 1-17, doi: 10.1038/s12276-019-0230-6.
17. Qi S., Yu L., Li H. et al. Isocitrate dehydrogenase mutation is associated with tumor location and magnetic resonance imaging characteristics in astrocytic neoplasms. *Oncol Lett.*, 2014, vol. 7, no. 6, pp. 1895-1902, doi: 10.3892/ol.2014.2013.
18. Carrillo J.A., Lai A., Nghiemphu P.L. et al. Relationship between tumor enhancement, edema, IDH1 mutational status, MGMT promoter methylation, and survival in glioblastoma. *AJNR Am J Neuroradiol.*, 2012, vol. 33, no. 7, pp. 1349-1355, doi: 10.3174/ajnr.A2950.
19. Gritsch S., Batchelor T.T., Gonzalez Castro L.N. Diagnostic, therapeutic, and prognostic implications of the 2021 World Health Organization classification of tumors of the central nervous system. *Cancer*, 2022, vol. 128, no. 1, pp. 47-58, doi: 10.1002/cncr.33918.
20. Voropaeva E.N., Pospelova T.I., Orlov Y.L., Churkina M.I., Berezina O.V., Gurazheva A.A., Ageeva T.A., Seregina O.B., Maksimov V.N. The Methylation of the p53 Targets the Genes MIR-203, MIR-129-2, MIR-34A and MIR-34B/C in the Tumor Tissue of Diffuse Large B-Cell Lymphoma. *Genes*, 2022, vol. 13, no. 8, p. 1401, doi: 10.3390/genes13081401.
21. Коновалов Н.А., Асютин Д.С., Шайхаев Е.Г., Капровой С.В., Тимонин С.Ю. Молекулярные биомаркеры астроцитом головного и спинного мозга. *Acta Naturae* (русскоязычная версия), 2019, т. 11, к. 2, № 41, с. 17-27 [Kononov N.A., Vasyutin D.S., Shaikhaev E.G., Kaprov S.V., Timonin S.Yu. Molecular biomarkers of astrocytomas of the brain and spinal cord. *Acta Naturae*, 2019, vol. 11, no. 2, no. 41, pp. 17-27 (In Russ.)].
22. Пушкин А.А., Тимошкина Н.Н., Росторгуев Э.Е. Экспрессионный статус 15-ти генов в глиальных опухолях головного мозга. *Исследования и практика в медицине*, т. 6, № Спецвыпуск, 2019, с. 230 [Pushkin A.A., Timoshkina N.N., Rostorguev E.E. Expression status of 15 genes in glial tumors of the brain. *Research Practical Medicine Journal*, vol. 6, no. Special issue, 2019, p. 230 (In Russ.)].
23. Davis M.E. Epidemiology and Overview of Gliomas. *Semin. Oncol. Nurs.*, 2018, vol. 34, no. 5, pp. 420-429, doi: 10.1016/j.soncn.2018.10.001.
24. Stelzer G., Rosen N., Plaschkes I. et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics*, 2016, vol. 54, pp. 1.30.1-1.30.33, doi: 10.1002/cpbi.5.
25. Aleksander S.A., Balhoff J. et al. The Gene Ontology Consortium, The Gene Ontology knowledgebase in 2023. *Genetics*, 2023, vol. 224, no. 1, iyad031, pp. 2-12, doi: 10.1093/genetics/iyad031.
26. Voropaeva E.N., Pospelova T.I., Voevoda M.I., Maksimov V.N., Orlov Y.L., Seregina O.B. Clinical aspects of TP53 gene inactivation in diffuse large B-cell lymphoma. *BMC Med Genomics*, 2019, vol. 12, suppl. 2, pp. 35, doi: 10.1186/s12920-019-0484-9.
27. Xu Y., Tong X., Yan J., Wu X., Shao Y.W., Fan Y. Short-Term Responders of Non-Small Cell Lung Cancer Patients to EGFR Tyrosine Kinase Inhibitors Display High Prevalence of TP53 Mutations and Primary Resistance Mechanisms. *Transl. Oncol.*, 2018, vol. 11, no. 6, pp. 1364-1369, doi: 10.1016/j.tranon.2018.08.010.
28. Побырзин В.В. Изменение экспрессии гена-супрессора TP53 в тканях крыс при экспериментальном аскаридозе на различных сроках наблюдения во время воспроизведения опухолевой модели глиомы крыс C6 in situ. *Международный научно-исследовательский журнал*, 2022, т. 9, № 123, с. 1, doi: 10.23670/IRJ.2022.123.20 [Pobyarzhin V.V. Changes in TP53 suppressor gene expression in rat tissues in experimental ascariasis at different observation periods during reproduction of the rat glioma C6 tumor model in situ. *International research journal*, 2022, vol. 9, no. 123, p. 1 (In Russ.)].
29. Tran P.M.H., Tran L.K.H., Satter K.B., Purohit S., Nechtman J., Hopkins D.I., Dos Santos B., Bollag R., Kolhe R., Sharma S., She J.X. Retrospective Validation of a 168-Gene Expression Signature for Glioma Classification on a Single Molecule Counting Platform. *Cancers (Basel)*, 2021, vol. 13, no. 3, p. 439, doi: 10.3390/cancers13030439.
30. Cohen S. Isolation of a Mouse Submaxillary Gland Protein Accelerating Incisor Eruption and Eyelid Opening in the New-born Animal. *J. Biol. Chem.*, 1962, vol. 237, no. 6, pp. 1555-1562.
31. Levi-Montalcini R., Booker B. Excessive growth of the sympathetic ganglia evoked by a protein isolated from mouse salivary glands. *Proc. Natl. Acad. Sci. U.S.A.*, 1960, vol. 46, no. 3, p. 373.
32. Tan S., Spear R., Zhao J., Sun X., Wang P. Comprehensive Characterization of a Novel E3-Related Gene Signature with Implications in Prognosis and Immunotherapy of Low-Grade Gliomas. *Front Genet.*, 2022, vol. 13, p. 905047, doi: 10.3389/fgene.2022.905047.
33. Wen J., Zhao W., Shu X. A novel cuproptosis-related LncRNA signature: Prognostic and therapeutic value for low grade glioma. *Front. Oncol.*, 2023, vol. 12, p. 1087762, doi: 10.3389/fonc.2022.1087762.
34. Xie J., Sun J., Feng J., Yang F., Wang J., Wen T., Nie Q. Kernel Differential Subgraph Analysis to Reveal the Key Period Affecting Glioblastoma. *Biomolecules*, 2020, vol. 10, p. 318, doi: 10.3390/biom10020318.
35. Zhao B.-W., Su X.-R., Hu P.-W., Huang Y.-A., You Z.-H., Hu L. iGRLDTI: an improved graph representation learning method for predicting drug-target interactions over heterogeneous biological information network. *Bioinformatics*, 2023, vol. 39, no. 8, p. 2, btad45, doi: 10.1093/bioinformatics/btad451.
36. Babenko V.N., Gubanov N.V., Bragin A.O., Chadaeva I.V., Vasiliev G.V., Medvedeva I.V., Gaytan A.S., Krivoschapkin A.L., Orlov Y.L. Computer Analysis of Glioma Transcriptome Profiling: Alternative Splicing Events.

Journal of Integrative Bioinformatics, 2017, vol. 14, no. 3, p. 20170022, doi: 10.1515/jib-2017-0022.

37. Babenko V.N., Bragin A.O., Spitsina A.M., Chadaeva I.V., Galieva E.R., Orlova G.V., Medvedeva I.V., Orlov Y.L. Analysis of differential gene expression by RNA-seq data in brain areas of laboratory animals. *Journal of Integrative Bioinformatics*, 2016, vol. 13, no. 4, pp. 292, doi: 10.2390/biecoll-jib-2016-292.

38. Ivanisenko T.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. The New Version of the ANDDigest Tool with Improved AI-Based Short Names Recognition. *Int J Mol Sci.*, 2022, vol. 23, no. 23, p. 14934, doi: 10.3390/ijms232314934.

39. Khlebodarova T.M., Demenkov P.S., Ivanisenko T.V., Antropova E.A., Lavrik I.N., Ivanisenko V.A. Primary and Secondary micro-RNA Modulation the Extrinsic Pathway of Apoptosis in Hepatocellular Carcinoma. *Mol. Biol. (Mosk)*, 2023, vol. 57, no. 2, pp. 165-175, doi: 10.1134/S0026893323020103.

40. Brix T.J., Berentzen M., Becker L., Storck M., Varghese J. Development of a Command Line Interface for the Analysis of Result Sets from Automated Queries to Literature Databases. *Stud Health Technol Inform.*, 2023, vol. 302, pp. 162-166, doi: 10.3233/SHTI230095.

41. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*, 2011, vol. 2011, baq036, pp. 1-12, doi: 10.1093/database/baq036.

42. Veljkovic A.N., Orlov Y.L., Mitic N.S. BioGraph: Data Model for Linking and Querying Diverse Biological Metadata. *International Journal of Molecular Sciences*, 2023, vol. 24, no. 8, p. 6954, doi: 10.3390/ijms24086954.

43. Orlov Y.L., Baranova A.V. Editorial: Bioinformatics of Genome Regulation and Systems Biology. *Front Genet.*, 2020, vol. 11, pp. 625, doi: 10.3389/fgene.2020.00625.

44. Orlov Y.L., Hofstaedt R.M., Kolchanov N.A. Introductory note for BGRS/SB-2014 special issue. *J. Bioinform. Comput. Biol.*, 2015, vol. 13, no. 1, p. 1502001, doi: 10.1142/S0219720015020011.

45. Dergilev A.I., Orlova N.G., Dobrovolskaya O.B., Orlov Y.L. Statistical estimates of multiple transcription factors binding in the model plant genomes based on ChIP-seq data. *Journal of Integrative Bioinformatics*, 2021, vol. 19, no. 1, p. 20200036, doi: 10.1515/jib-2020-0036.

46. Orlov Y., Dobrovolskaya O., Chen M., Hofstadt R. Bioinformatics of genome regulation and structure – 2020 papers collection. *Journal of Integrative Bioinformatics*, 2020 vol. 17, no. 4, p. 20200038, doi: 10.1515/jib-2020-0038.

47. Карпов О.Э., Клименко Г.С., Лебедев Г.С. Применение интеллектуальных систем в здравоохранении. *Современные наукоемкие технологии*, 2016, т. 7, с. 38-43 [Karpov O.E., Klimenko G.S., Lebedev G.S. Application of intelligent systems in healthcare. *Modern high technologies*, 2016, vol. 7, pp. 38-43 (In Russ.)].

48. Lebedev G., Polikarpov A., Golubev N., Tyurina E., Serikov A., Selivanov D., Orlov Y. The Geographic Information System of the Russian Ministry of Health. In: Czarnowski I., Howlett R., Jain L. (eds) *Intelligent Decision Technologies. IDT 2020. Smart Innovation, Systems and Technologies*, 2020, vol. 193, pp. 403-411. Springer, Singapore, doi: 10.1007/978-981-15-5925-9_34.

49. Xing X., Yang F., Li H. et al. Multi-level attention graph neural network based on co-expression gene modules for disease diagnosis and prognosis. *Bioinformatics*, 2022, vol. 38, no. 8, pp. 1-6, 2178-2186, doi: 10.1093/bioinformatics/btac088.

50. Oloulade B.M., Gao J., Chen J., Al-Sabri R., Wu Z. Cancer Drug Response Prediction with Surrogate Modeling-Based Graph Neural Architecture Search. *Bioinformatics*, 2023, btad478, doi: 10.1093/bioinformatics/btad478.

RECONSTRUCTION OF GENE AND ASSOCIATIVE NETWORKS OF DISEASES TO SEARCH FOR TARGET GENES

Turkina V.A.¹, Iarema P.O.¹, Mayorova A.A.¹, Orlova N.G.², Savina E.A.^{1,3}, Orlov Y.L.^{1,4}¹ I.M. Sechenov First Moscow State Medical University of the Ministry of Health of Russia (Sechenov University)*Trubetskaya 8-2, Moscow, 119991, Russia; e-mail: vasundel@mail.ru*² Financial University under the Government of the Russian Federation*Leningradsky Ave, 49.2, Moscow, 125167, Russia*³ Engelhardt Institute of Molecular Biology of RAS*Vavilova str., 32, Moscow, 119991, Russia*⁴ Institute of Cytology and Genetics SB RAS*Lavrentieva str., 10, Novosibirsk, 630090, Russia; e-mail: orlov@d-health.institute*

Received 10.08.2023. DOI: 10.29039/rusjbp.2023.0647

Abstract. Computer reconstruction of gene networks—sets of interacting macromolecules with common functions – is a complex task in bioinformatics. Gene networks, presented as graphical structures, offer a convenient means of visualization and provide qualitatively new insights into the set of studied genes and their biological functions. Expanding further, we explore associative gene networks related to diseases. These networks encompass not only genes and their products (proteins, metabolites, non-coding RNA and drug compounds) but also common disease indicators (symptoms, phenotypic manifestations). Software tools aimed at reconstructing gene network graphs are undergoing active development worldwide and find significant utility in biomedicine. We evaluate available online bioinformatics tools for gene network reconstruction based on gene lists and corresponding computer analysis pipelines. We delve into examples showcasing the utilization of online programs for analyzing the glioma gene network. The proposed approach can be extended to other functional gene sets linked to diseases. Scrutinizing the structure of disease-associated gene networks enables the identification of pivotal genes, which can serve as therapeutic targets. Employing similar bioinformatics strategies, gene networks of intricate disorders such as Parkinson's disease, dementia, schizophrenia, breast cancer, and other cancers have been explored. We delve into the applications of computer programs for gene network analysis and address pedagogical aspects of the field.

Key words: *bioinformatics, gene networks, glioma, complex diseases, protein-protein interactions, medical informatics.*