

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВА-
ТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ ПЕРВЫЙ МОС-
КОВСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ
ИМЕНИ И.М. СЕЧЕНОВА МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОС-
СИЙСКОЙ ФЕДЕРАЦИИ (СЕЧЕНОВСКИЙ УНИВЕРСИТЕТ)

ИНСТИТУТ ЦИФРОВОЙ МЕДИЦИНЫ

КАФЕДРА ИНФОРМАЦИОННЫХ И ИНТЕРНЕТ-ТЕХНОЛОГИЙ

Группа 801-04

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**КОМПЬЮТЕРНАЯ РЕКОНСТРУКЦИЯ ГЕННЫХ СЕТЕЙ,
АССОЦИИРОВАННЫХ С БОЛЕЗНЬЮ ПАРКИНСОНА**

Исполнитель: Белова О.В.

Научный руководитель: Орлов Ю.Л.

Москва 2023

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	2
ГЛАВА 1. ОБЗОР ГЕННЫХ СЕТЕЙ, СВЯЗАННЫХ С БОЛЕЗНЬЮ ПАРКИНСОНА	5
1.1. Структуры генных сетей	5
1.2. Методы исследования генных сетей.....	15
1.3. Определение признаков болезни Паркинсона в генных сетях.....	17
ВЫВОДЫ ПО 1-Й ГЛАВЕ	21
ГЛАВА 2. КОМПЬЮТЕРНАЯ РЕКОНСТРУКЦИЯ ГЕННОЙ СЕТИ	22
2.1. Анализ нейронных сетей для диагностики заболеваний.....	22
в медицине	22
2.2. Определение специфики болезни Паркинсона для генных сетей	33
2.3. Разработка структуры генной сети.....	37
ВЫВОДЫ ПО 2-Й ГЛАВЕ	44
Глава 3. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ГЕННОЙ СЕТИ	45
3.1. Формирование техничек требований к программе	45
3.2. Выбор языка программирования	57
3.3. Разработка алгоритма функционирования генной сети, ассоциированных с болезнью Паркинсона	58
3.4. Разработка программы для реконструкции генной сети	61
ВЫВОДЫ ПО 3-Й ГЛАВЕ	64
ГЛАВА 4. ОЦЕНКА ЭКОНОМИЧЕСКОЙ.....	65
ЭФФЕКТИВНОСТИ ПРОГРАММЫ	65
4.1. Сравнение разработанной сети с существующими сетями	65
4.2. Определение затрат на разработку программы	71
4.3. Оценка сроков окупаемости программы	73
ВЫВОДЫ ПО 4-Й ГЛАВЕ	74
ГЛАВА 5. ТЕСТИРОВАНИЕ ПРОГРАММЫ	76
5.1. Формирование набора тестов	76
5.2. Выполнение тестов в программе.....	80
5.3. Оценка быстродействия программы и точности результатов.....	82
ВЫВОДЫ ПО 5-Й ГЛАВЕ	82
ЗАКЛЮЧЕНИЕ	83

ВВЕДЕНИЕ

Болезнь Паркинсона – одно из наиболее распространенных неврологических заболеваний. Эпидемиологические исследования указывают на то, что болезнь Паркинсона наблюдается во всех этнических группах и встречается с равной частотой у мужчин и женщин. Заболеваемость составляет около 20 на 100 000 населения, распространенность – 120-180 на 100 000 и достигает 1 % в популяции лиц старше 60 лет. Заболевание встречается, как правило, в пожилом и старческом возрасте. Средний возраст начала заболевания составляет 65,3±12,6 года, распространенность в популяции лиц старше 85 лет – 2205,3 на 100 000.

В России каждый год диагностируют болезнь Паркинсона почти у 10 тысяч человек. Во всем мире паркинсонизмом болеет больше 10 миллионов человек, и в перспективе к 2050 г. их число может увеличиться до 50 миллионов. Приведенные данные говорят о необходимости коррекции современных методов терапии болезни Паркинсона, создания современных эффективных лекарственных препаратов.

Оба препарата достаточно эффективны и используют для лечения симптоматического паркинсонизма. Оба препарата будут эффективны на ранней стадии болезни Паркинсона, а также в комбинации с леводопой, усиливая его действия. Мирапекс имеет меньшее количество противопоказаний, но больше побочных эффектов по сравнению с препаратом Юмекс.

Проблемой для экспериментальных исследований патогенеза болезни Паркинсона является невозможность забора материала от пациентов с данной патологией, поэтому практически полностью отсутствует возможность наблюдения за процессами, протекающими в живых клетках человека при БП. Известны модели болезни Паркинсона *in vivo* (на живых организмах) и *in vitro* («в пробирке»). Наиболее приближенными к процессам человеческих клеток являются модели патологии *in vivo* на животных, однако при их использовании возникают сложности с воспроизводимостью результатов,

трудности содержания чистых линий животных и обоснования использования моделей *in vivo* перед этическим комитетом.

В то же время, в последнее время все большее применение находят модели *in vitro*. Так, клеточные культуры являются универсальным методом для исследования «физиологических» и патологических явлений, выяснения механизмов передачи сигнала, регуляции экспрессии генов, клеточной пролиферации, а также механизмов их гибели. Данные модели полностью не исключают модели *in vivo*, а являются хорошим к ним дополнением, позволяя исследовать физиологические явления и механизмы патогенеза заболеваний, выяснить механизмы передачи сигнала, регуляции экспрессии генов, клеточной пролиферации и гибели [22].

Среди моделей болезней Паркинсона *in vivo* различают модели генетические (нокаутные и трансгенные модели), нейротоксические (системное введение нейротоксинов) и стереотаксические (стереотаксическое введение ротенона, параквата, 6-OHDA, MPP⁺, MPTP, метамфетамина, дегуелина и других нейротоксинов). Из моделей *in vitro* главным образом используются модели с приложением экзогенных (ротенон, паракват и MPTP) и эндогенных (6-гидроксидофамин, MPP⁺, L-DOPA) нейротоксинов. Для моделирования болезни Паркинсона *in vitro* используются культуры нейронов, астроцитов и клеток микроглии, составляющие функциональную сеть в цитоархитектонике головного мозга, контактирующие друг с другом посредством нейрон-глиальных взаимодействий и поддерживающие гомеостаз головного мозга.

Ротеноновые модели БП *in vitro*. С экологической точки зрения особый интерес представляет широкое использование органических пестицидов, в частности, пестицида ротенона. Было показано, что ротенон является причиной селективной дегенерации дофаминергических нейронов при системном применении у крыс, ингибируя митохондриальный комплекс I и вызывая дестабилизацию микротрубочек [10, 17].

Предмет исследования: генные сети при болезни Паркинсона у людей преклонного возраста

Объект исследования: пациенты преклонного возраста с болезнью Паркинсона

Материалы и методы исследования: исследовательская работа с источниками литературы, интернет – ресурсами, разработка программного обеспечения.

Задачи исследования:

1. По литературным данным изучить теоретические аспекты болезни Паркинсона
2. По литературным данным изучить причины возникновения болезни Паркинсона у пациентов преклонного возраста
3. Выполнить анализ генных сетей при болезни Паркинсона
4. Разработать алгоритм и программное обеспечение для реконструкции генных сетей при болезни Паркинсона
5. Выполнить тестирование разработанного программного обеспечения и оценить точность прогноза заболевания с помощью программы

Дипломная работа состоит из введения, пяти глав, заключения, списка литературы, содержит рисунки, таблицы.

ГЛАВА 1. ОБЗОР ГЕННЫХ СЕТЕЙ, СВЯЗАННЫХ С БОЛЕЗНЬЮ ПАРКИНСОНА

1.1. Структуры генных сетей

Болезнь Паркинсона в 20% случаев имеет генетические причины (наследственную предрасположенность и мутации), а в 80% случаев - факторы окружающей среды. Клинические проявления дрожательного паралича и синдрома паркинсонизма возникают в результате перенесенных острых и хронических инфекций нервной системы (эпидемический энцефалит Экономо, клещевой, вирусный и другие виды энцефалитов). Причинами болезни могут служить церебральный атеросклероз, сосудистые заболевания головного мозга, опухоли, травмы нервной системы, длительное использование препаратов фенотиазинового ряда (аминазин, трифтазин), производных раувольфии, метилдофа - лекарственный паркинсонизм.

Паркинсонизм может развиваться при острой или хронической интоксикации окисью углерода и марганца. В возникновении акинетико-ригидного синдрома может иметь значение наследственно обусловленное нарушение обмена катехоламинов в мозге или неполноценность ферментных систем, контролирующих этот обмен. Часто выявляется семейный характер заболевания при аутосомно-доминантном типе наследования. Различные экзо- и эндогенные факторы (атеросклероз, инфекции, интоксикации, травмы) способствуют проявлению генных дефектов в механизмах обмена катехоламинов в подкорковых ядрах и возникновению заболевания.

Заболевание впервые было описано в 1817 году английским врачом Джеймсом Паркинсоном, который назвал его треморным параличом. Болезнь Паркинсона относится к группе нейро-дегенеративных заболеваний, что означает, что нейроны в головном мозге человека, вырабатывающие дофамин в качестве нейромедиатора, умирают при этой болезни. Доля пациентов с болезнью Паркинсона в возрастной группе старше 60 лет составляет 1%, а старше 85 лет - от 2,6% до 4%. Чаще всего первые симптомы заболевания появляются примерно в возрасте 55-60 лет.

Однако в некоторых случаях болезнь может развиваться в возрасте до 40 лет (болезнь Паркинсона с ранним началом) или до 20 лет (юношеская форма болезни). В России ежегодно болезнь Паркинсона заболевает почти 10 000 человек. Во всем мире более 10 миллионов человек с болезнью Паркинсона, и в будущем их число может вырасти до 50 миллионов к 2050 году.

Болезнь Паркинсона имеет генетические причины (наследственную предрасположенность и мутации в 20% случаев и факторы окружающей среды в 80% случаев). Клинические проявления паркинсонизма и синдрома баллизма возникают в результате перенесенных острых и хронических инфекций нервной системы (эпидемический энцефалит Экономо, клещевой, вирусный энцефалит и другие виды). [2,с.67] Причинами заболевания могут быть церебральный атеросклероз, сосудистые заболевания головного мозга, опухоли, травмы нервной системы, длительный прием препаратов фенотиазинового ряда (аминазин, трифтазин), производных раувольфии, метилдопа - болезнь Паркинсона. Паркинсонизм может развиваться при остром или хроническом отравлении оксидом углерода и марганцем.

Если возникает акинетический ригидный синдром, важным может быть наследственное нарушение метаболизма катехоламинов в головном мозге или неадекватность ферментных систем, контролирующих этот обмен. Семейная природа заболевания часто проявляется по аутосомно-доминантному типу наследования. Различные экзо- и эндогенные факторы (атеросклероз, инфекции, интоксикация, травмы) способствуют проявлению генетических дефектов механизмов обмена катехоламинов в подкорковых ядрах и возникновению заболевания. [11,с.28]

Болезнь Паркинсона - медленно прогрессирующее хроническое неврологическое заболевание, относится к дегенеративным заболеваниям экстрапирамидной моторной системы. Вызвано прогрессирующим разрушением и гибелью нейронов, вырабатывающих нейромедиатор дофамин. Болезнь Паркинсона является наиболее частым нейродегенеративным заболеванием после болезни Альцгеймера. Этиология болезни Паркинсона окончательно не

выяснена. Около 15 % людей с болезнью Паркинсона имеют семейный анамнез данного заболевания. При микроскопическом исследовании пораженных областей выявляют уменьшение числа нервных клеток. В них определяется наличие телец Леви, которые образуются вследствие скопления в цитоплазме белка α -синуклеина. Основные препараты, устраняющие двигательные нарушения при болезни Паркинсона: леводопа (чаще в комбинации с периферическими ингибиторами ДОФА-декарбоксилазы), агонисты дофаминовых рецепторов и ингибиторы МАО-В.

На сегодняшний день известны мутации в 15 генах, ответственных за моногенные формы болезни Паркинсона (Hernandez et al. 2016). Однако все эти известные моногенные формы в совокупности объясняют лишь 30% моногенных и 3- 5% генетически сложных случаев этой болезни. Составление списка генов, связанных с наследственной предрасположенностью к болезни Паркинсона. Интернет-ресурс OMIM (Online Mendelian Inheritance in Man) (<https://omim.org/>) использовался для анализа генов менделевского наследования у человека (Amberger et al., 2019), поиск проводился по ключевому слову Parkinson disease. В настоящий момент генетические исследования болезни Паркинсона ведутся двумя, не исключаящими друг друга направлениями. Первое направление руководствуется гипотезой общего варианта заболевания и полагает, что генетическая основа болезни является совокупным результатом большого числа общих генетических факторов, каждый из которых оказывает относительно небольшое влияние на риск заболевания, но в совокупности создает значительный риск. Второе направление исследований предполагает, что наибольший риск в активном развитии болезни Паркинсона дают редкие генетические варианты, для которых существуют высоко функциональные вредные аллели (Billingsley et al, 2018). Болезнь Паркинсона не является нейродегенеративным заболеванием в 95% случаев. Это функциональное расстройство ЦНС (<http://griza.nevrologica.ru/parkinson>).

Использовался интернет-ресурс OMIM (Online Mendelian Inheritance in Man) (<https://omim.org/>) для поиска генов предрасположенности к заболева-

нию. Был найден список из 229 генов и выполнен расчет категорий и анализ генных онтологий с помощью ресурса PANTHER (Protein ANalysis THrough Evolutionary Relationships) (<http://pantherdb.org/>) (Mi et al., 2013). Из 229 исходных генов распознано 170 идентификаторов, 59 идентификаторов были не распознаны или не могли быть картированы однозначно. Всего в референсном геноме по PANTHER использовался 20851 ген. С помощью PANTHER мы построили таблицу онтологий для категорий биологических процессов, в целях получения наиболее информативных результатов значения p -value ограничились $E=20$. Наиболее значимыми категориями для генов болезни Паркинсона являются общая регуляция клеточной смерти, регуляция клеточной смерти нейронов, регуляция апоптоза и программируемой клеточной смерти, негативная регуляция клеточной смерти, что подтверждает этиологию заболевания - гибель нейронов черной субстанции.

Наиболее значимыми категориями для генов болезни Паркинсона являются категории общей регуляции клеточной смерти, регуляции клеточной смерти нейронов, регуляции апоптоза и программируемой клеточной смерти, негативная регуляция клеточной смерти. Эти данные подтверждают ключевые этиологические признаки заболевания, среди которых центральным аспектом патофизиологии болезни Паркинсона является прогрессирующая гибель дофаминовых нейронов среднего мозга и их аксональных проекций.

Процессы аутофагии являются одним из основных путей внутриклеточной деградации α -синуклеина и актуальные исследования показывают, что дисфункциональная аутофагия при болезни Паркинсона является одним из основных факторов риска развития заболевания (Hale et al. 2016).

Наиболее значимые категории в проявлении болезни Паркинсона - соматодендритный компартмент, формирование нейронов, аксонов, пресинапсы, синапсы и другие, тесно связанные с нервной тканью, категории. Таким образом, анализ категорий генных онтологий с помощью PANTHER для генов болезни Паркинсона показал иерархию важных категорий, регулирующих процессы клеточной смерти, развития нервных клеток, а также онтоло-

гии, связанные с метаболическими процессами, митохондриями и цитоплазмой.

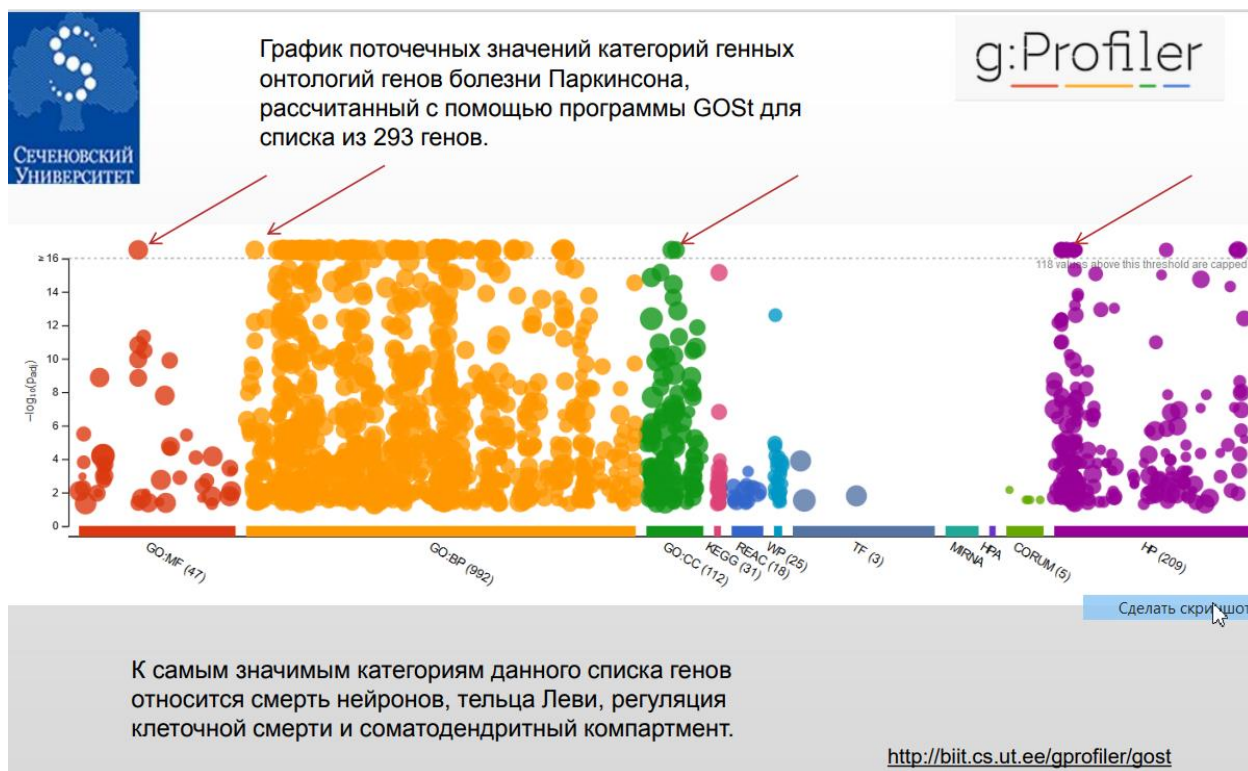


Рисунок 1. График категорий генных онтологий

К самым значимым категориям данного списка генов относится смерть нейронов, тельца Леви, регуляция клеточной смерти и соматодендритный компартмент.

Для реконструкции генной сети взаимодействий генов болезни Паркинсона использовались ресурсы GeneMANIA и STRING-DB. На рисунке 2 представлена генная сеть из 187 генов болезни Паркинсона, реконструированная с помощью GeneMANIA.

К настоящему времени известно, что ген SNCA, кодирующий α -синуклеин, является плеiomорфным, и любые, как редкие мутации так и общие, вариации в этом локусе изменяют риск развития заболевания.

В центре построенной сети находятся гены (белки), имеющие большое число связей с другими элементами - SNCA, CASP3, GFRA1, HTT, PARK7. Эта тенденция подтверждается актуальными исследованиями ассоциаций генов-кандидатов (Billingsley K.J. et al. 2018), в которых наиболее статистически значимыми сигналами, связанными с болезнью Паркинсона являются

распространенные варианты, расположенные близко к SNCA, LRRK2и
MART, а также низкочастотные кодирующие варианты в GBA

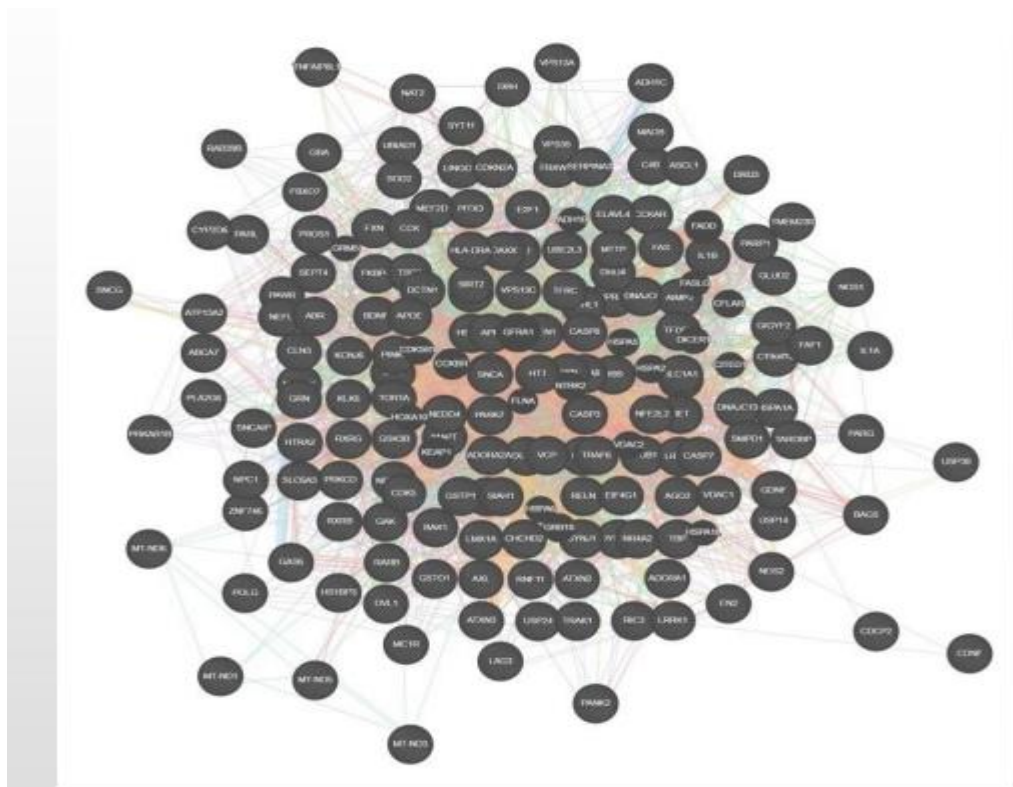


Рисунок 2. Генная сеть при болезни Паркинсона (187 генов)

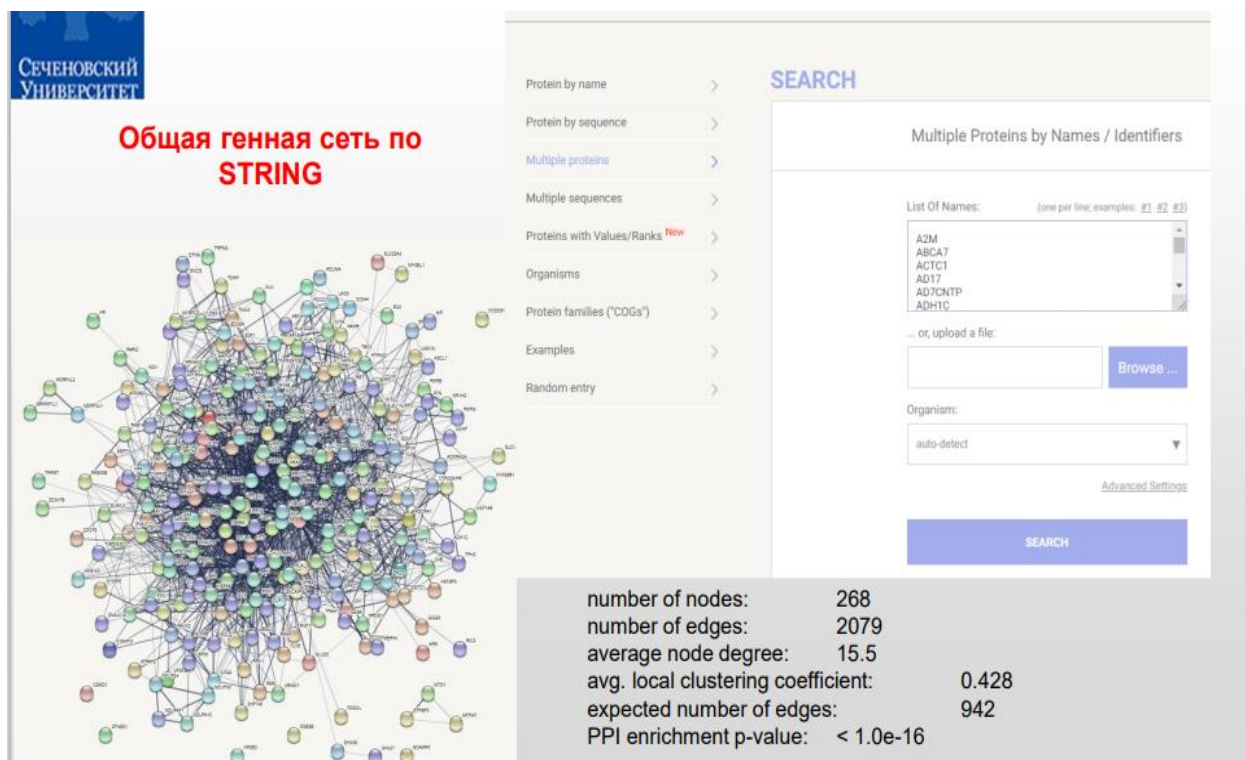


Рисунок 3. Общая генная сеть

Реконструкция генной сети по STRING-DB

На рисунке 4 можно выделить несколько кластеров сети, самый большой из которых включает гены PARK2, SNCA, APOB, UBE2D1. Этот кластер занимает центральную позицию полученной сети и связан с наибольшим числом других объектов и кластеров – с UBE2W, UBE2D1, STUB1, RNF144B и др.

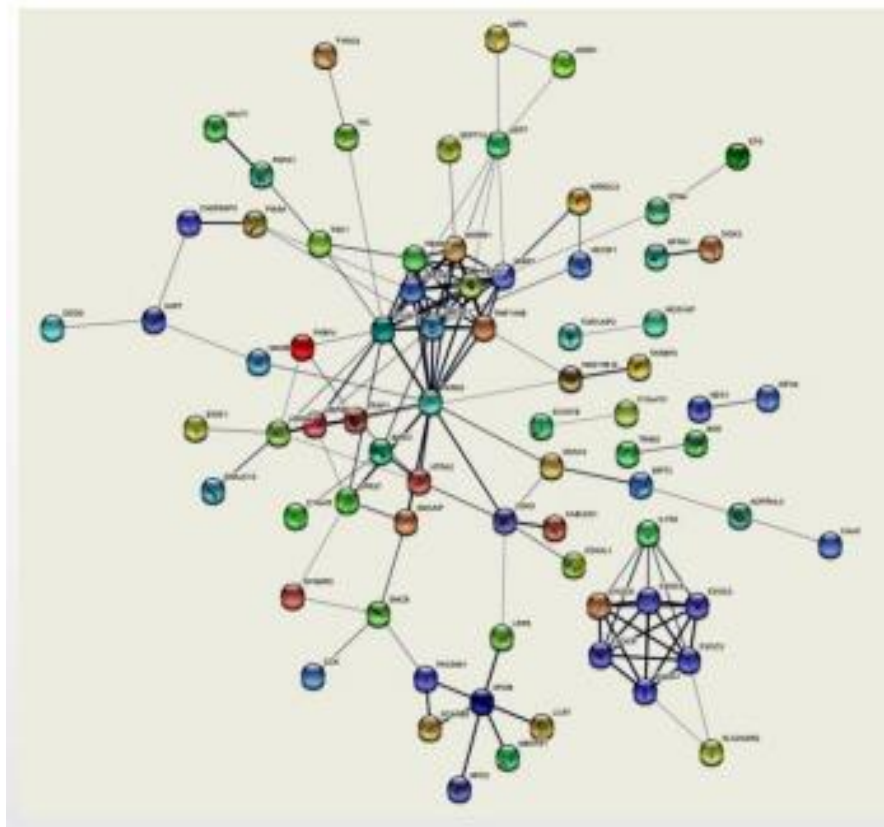


Рисунок 4. Реконструкция генной сети

Можно отметить небольшой, не связанный в общую сеть кластер, включающий гены EXOC5, EXOC6, EXOC8, EXOC3 (гены белкового комплекса, нацеливающего экзоцитарные везикулы на места стыковки на плазматической мембране) и несколько других.

Первоначальный вариант реконструированной с помощью ресурса STRING-DB сети оказался умеренно разреженным и часть объектов сети не контактировала с другими объектами. Тем не менее, в первичной реконструкции выделялся центральный плотно связанный кластер генов и связанный с несколькими небольшими кластерами генов белок PARK2. Статистика

показала, что сеть имеет неслучайно большое число связей (со значимостью $<2.44e-15$), средняя степень связности узла сети (белка) - 2.36, коэффициент кластеризации 0.415.

По данным ресурса GeneCards наиболее значимыми являются следующие 10 генов. Первое место в этом иерархическом списке занимает ген SNCA кодирующий белок альфасинуклеин, мутаций в этом гене приводят к развитию аутосомно-доминантных форм заболевания, тяжесть заболевания коррелирует с числом копий гена SNCA. Мутации в гене LRRK2 были идентифицированы как причины аутосомно-доминантной природы болезни Паркинсона как наиболее распространенной моногенной формы болезни, выявленной к настоящему времени (Paisán-Ruíz et al., 2004; Zimprich et al., 2004). Генетические варианты в LRRK2 связаны с большинством всех известных наследуемых проявлений болезни Паркинсона.

Связь между заболеваемостью болезнью Паркинсона и локусом, содержащим гены MAPP1 наиболее интересна для изучения, поскольку мутации в MAPP1 и сопутствующая им таупатия преимущественно связаны с деменцией. Растущее количество доказательств показывает, что MAPP1 H1 и его субгаплотип H1c связаны с повышенным риском развития болезни Паркинсона. Предполагается, что специфические для гаплотипа различия в экспрессии и потенциально альтернативном сплайсинге транскриптов MAPP1 влияют на клеточные функции на разных уровнях, что в конечном итоге повышает их восприимчивость (Skipper et al. 2004). Широко распространено предположение, что белки, кодируемые рецессивными генами, ассоциированными с ранним началом болезни Паркинсона, PINK1, PARK2, DJ-1 и FBXO7 участвуют в системе контроля качества митохондрий и ее регуляции (Mullin and Schapira 2015), считается, что митохондриальная дисфункция играет ключевую роль в ранних событиях начала болезни Паркинсона.

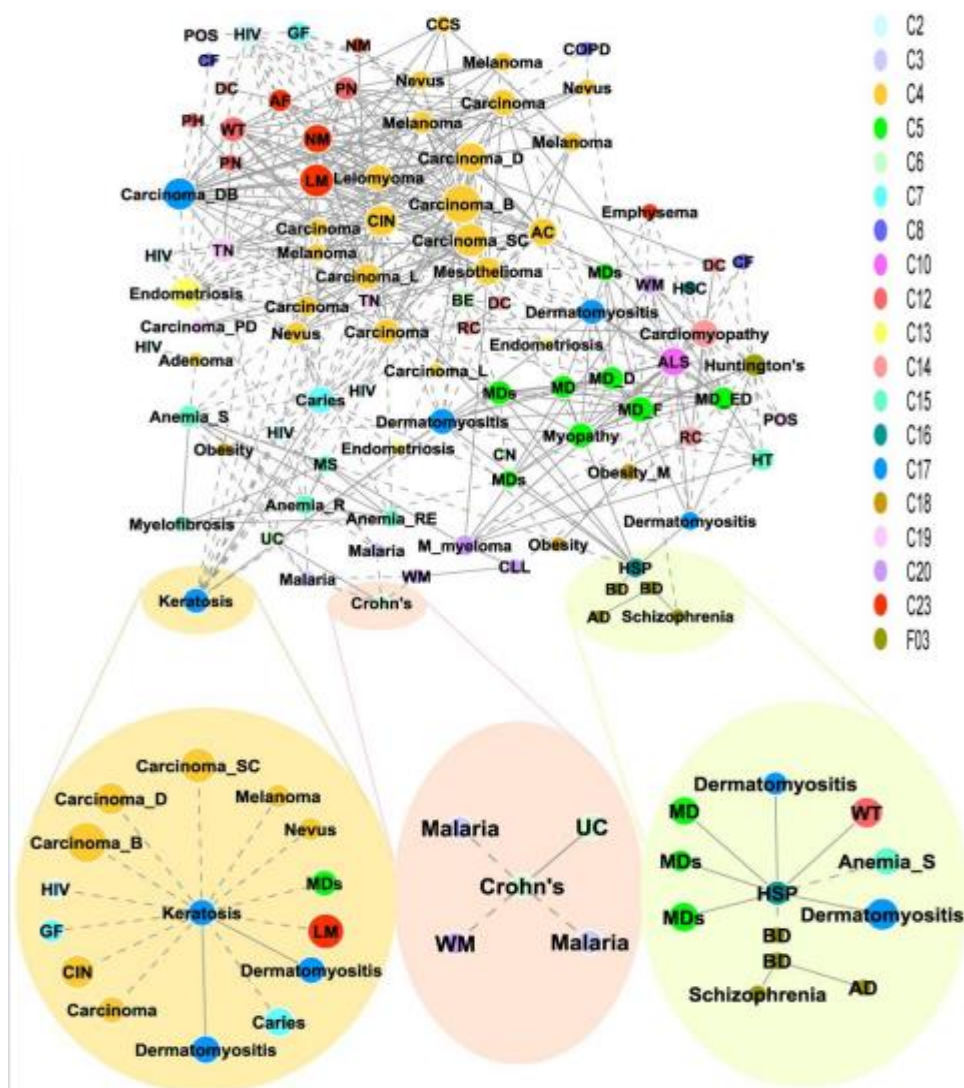


Рисунок 5. Сеть связи заболеваний и лекарственных средств

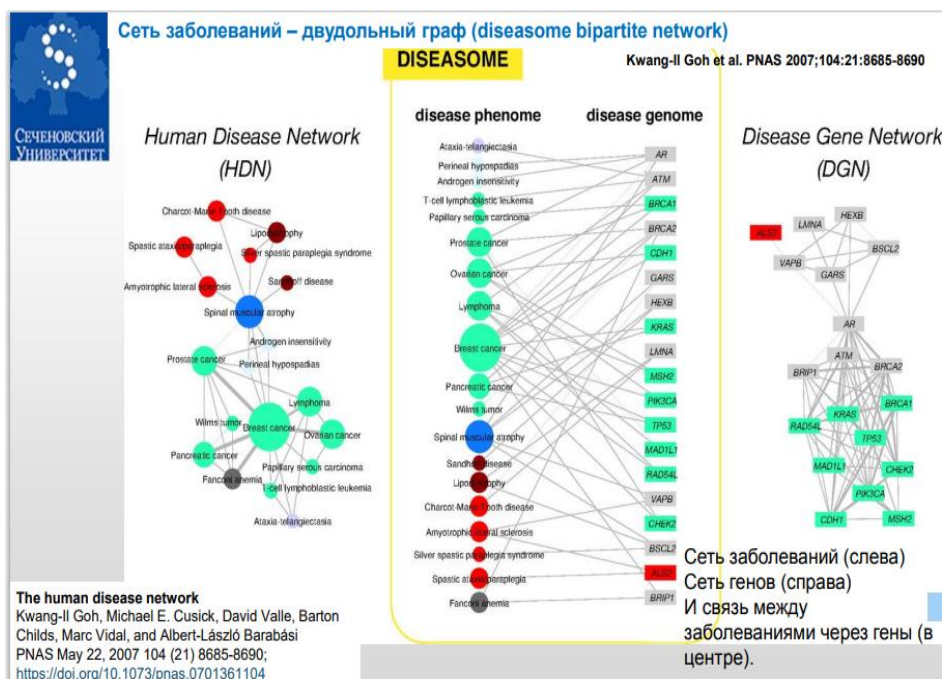


Рисунок 6. Сеть заболеваний – двудольный граф

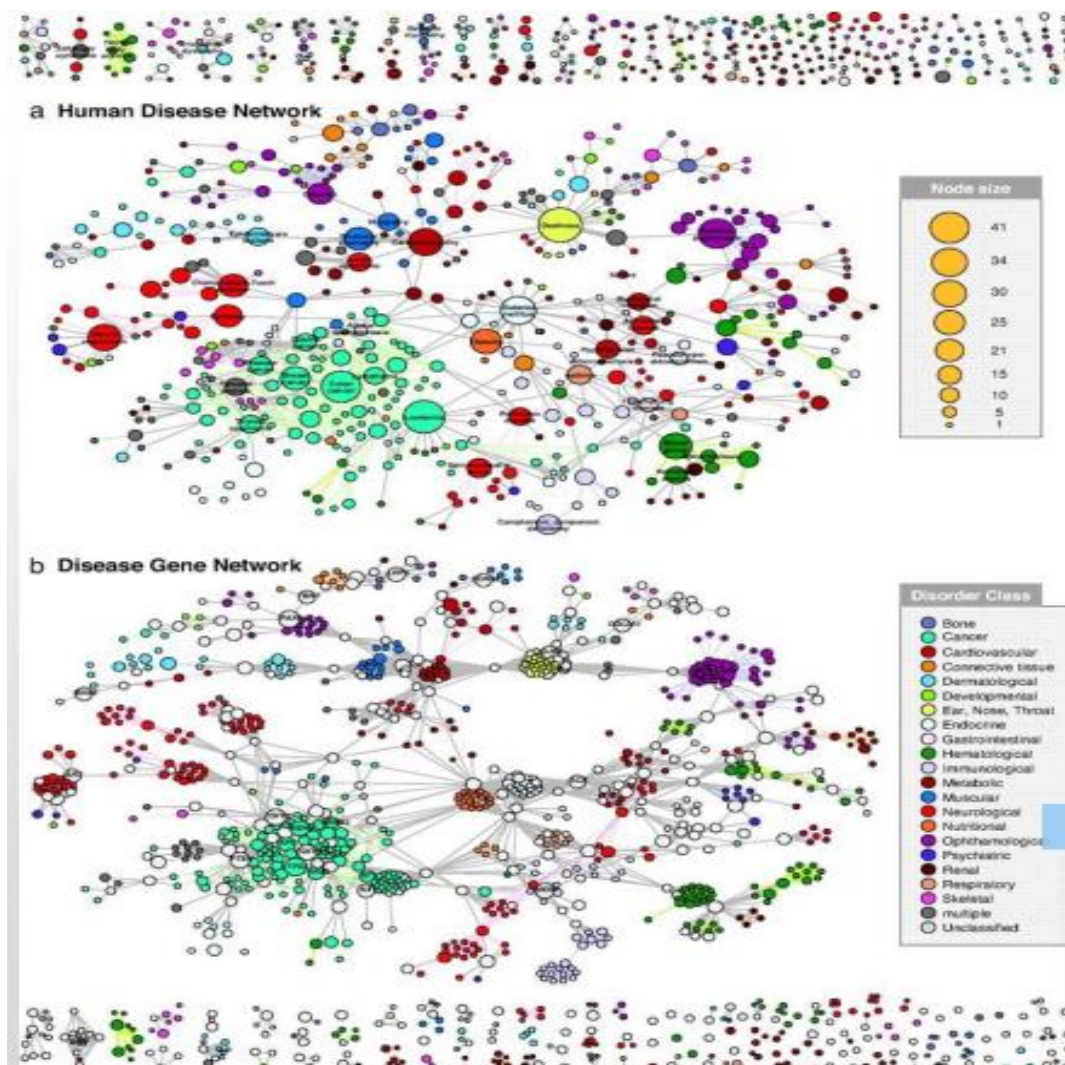


Рисунок 7. Сеть заболеваний человека (HDN) и сеть генов, относящихся к заболеваниям (DGN)

Составление списка ключевых генов, ассоциированных с развитием болезни Паркинсона, проведенный анализ категории генных онтологий для такого списка и реконструкция генной сети способны ускорить путь к лекарствам, модифицирующим степень выраженности болезни. Существующие методы терапии позволяют продлить время жизни пациента и улучшить качество его жизни, но пока они не направлены непосредственно на причину заболевания, а носят скорее заместительный характер, что говорит о необходимости дальнейшего исследования генетических основ болезни. Появились работы показывающие, что генная терапия становится перспективным подходом для лечения неврологических расстройств, включая болезнь Паркинсона (Niethammer M. et al. 2018).

1.2. Методы исследования генных сетей

В клинических испытаниях представлено много новых методов лечения БП, и некоторые из них сосредоточены на генно-терапевтических подходах, которые компенсируют потерю дофамина или защищают дофаминовые нейроны нервной системы от дегенерации.

Есть несколько важных причин для того, чтобы сосредоточиться на генной терапии как на лечении БП. Прежде всего, в настоящее время нет лекарства от этого заболевания. Во-вторых, были идентифицированы некоторые гены, которые могут модулировать фенотип нейрона или действовать как нейропротекторы. Кроме того, окружающая среда мозга не может позволить себе повторные инъекции в область, где черная субстанция встречается с полосатым телом, нигростриатумом.

Таким образом, генная терапия может быть единственным привлекательным методом лечения, вирусные векторы, используемые в терапии, диффундируют и способны трансдуцировать полосатое тело.

В настоящий момент нет методов лабораторного или инструментального исследования, которые были бы обязательны у каждого пациента с подозрением на БП. В последние годы больным с БП часто проводят КТ или МРТ головного мозга, однако чаще всего в этом нет необходимости, и в большинстве случаев диагноз может быть установлен на основе клинических данных.

Тем не менее, если клиническая картина у больного с синдромом паркинсонизма отклоняется от классического варианта, свойственного БП, в частности, отсутствует типичная реакция на дофаминергические средства, — необходимо проведение нейровизуализации.

При начале заболевания до 50 лет важно исключить гепатолентикулярную дегенерацию, о которой могут свидетельствовать роговичное кольцо Кайзера–Флейшера, низкий уровень церулоплазмина, повышение интенсивности сигнала от базальных ганглиев и мозжечка на T2-взвешенных МРТ изображениях, повышенная экскреция меди с мочой.



Рисунок 8. Алгоритм терапии

Диагностическую значимость может иметь и транскраниальная сонография глубинных структур мозга, выявляющее при БП гиперэхогенные изменения в проекции черной субстанции, связанные с накоплением железа и установленные в 92% случаев клинически вероятной БП, однако его результаты можно интерпретировать лишь в клиническом контексте.

Из практически важных, но пока отсутствующих в нашей стране, методов диагностики следует упомянуть позитронно-эмиссионную томографию (ПЭТ) и однофотонную эмиссионную компьютерную томографию (ОФЭКТ), позволяющие изучить синаптическую передачу на всех уровнях, а также проводить мониторинг патологического процесса. При выявлении снижения накопления F18-флюородопы при ПЭТ и β -CIT при ОФЭКТ в полосатом теле можно говорить о вовлечении в патологический процесс пресинаптических nigrostriарных терминалей (первичный паркинсонизм). Определение снижения накопления 11C-раклоприда (лиганд D2рецепторов) при ПЭТ будет указывать на уменьшение количества дофаминовых рецепторов в полосатом теле (паркинсонизм «плюс»).

Поскольку на данный момент способность замедлять процесс дегенерации за счет нейропротективного эффекта (возможность защищать интактные клетки от повреждения) или нейрорепаративного эффекта (возможность восстанавливать активность частично поврежденных клеток) ни у одного из применяющихся средств убедительно не доказана, лечение основывается пока на симптоматическом действии. Тем не менее потенциальная возможность нейропротективного эффекта, подтверждаемая экспериментальными или клиническими данными, должна учитываться при назначении лечения. В настоящее время широкое распространение получила концепция, подчеркивающая важность раннего назначения дофаминергической терапии – немедленно после установления диагноза – с тем, чтобы быстрее скорректировать нейрохимический дисбаланс в мозге и поддержать процессы компенсации.

Если ранее подчеркивалась необходимость возможно более длительного сохранения монотерапии, то в настоящее время преимущества такого подхода не представляются очевидными – по сравнению с ранним переходом к комбинации средств с разным механизмом действия. Вопрос о необходимости монотерапии или комбинированной терапии должен решаться индивидуально.

В любом случае при выборе препаратов и их дозы следует стремиться не к полному устранению симптомов, а к существенному улучшению функций, позволяющему поддерживать бытовую и профессиональную активность. В то же время следует избегать внесения в схему лечения сразу нескольких изменений (например, повышения дозы сразу нескольких средств или добавления сразу нескольких препаратов), это позволяет оценить отдельно эффективность и безопасность каждого из назначаемых средств.

1.3. Определение признаков болезни Паркинсона в генных сетях

В основе развития БП лежит повреждение дофаминергических нейронов черной субстанции и альфа-синуклина (α Syn), содержащихся в тельцах Леви. Существуют две формы болезни Паркинсона: наследственные и спора-

дические. Первая форма обусловлена генетическими абберациям в гене α Syn или его геометрическим дублированием или утроением. Причина спорадической БП не известна, но определены некоторые генетические и экологические факторы, такие как пестициды ротенон и паракват, токсин МРТР (1-метил-4-фенил-1,2,3,6-тетрагидропиридином), синтетический героин.

Недавний обзор Pan-Montojo и Reichmann предполагает важную роль токсичных веществ окружающей среды в этиологии спорадических БП. Хотя точное влияние генетических и экологических факторов в спорадической БП не известно, но все же некоторые элементы развития заболевания были выявлены. Самое главное из них нейровоспаление, окислительный стресс и неправильный фолдинг и агрегация α Syn. [1,2,10,11,15,16] Нейроны способны передавать различные формы белка α Syn друг к другу через аксоны, что позволяет распространить потенциально токсичный олигомерный белок, который может быть основным механизмом, лежащим в основе возникновения БП. Затем стало известно, что токсин может вызвать местное воспаление и окислительный стресс в кишечнике, тем самым инициируя осаждение α Syn, впоследствии который распространяется в ЦНС. Гипотетически, токсический α Syn может привести к гибели нейронов.

Это повлечет за собой порочный круг нейровоспаления. Известно, что нейроны в районах, пострадавших от телец Леви, имеют особые характеристики высокую метаболическую нагрузку, которая, делает эти нейроны особенно чувствительны к окислительному стрессу. Эти нейроны имеют высокие уровни эндогенных α Syn, они имеют длинные и сильно разветвленные аксоны с отсутствием или плохой миелинизацией. [6,14] Способность телец Леви распространяться через нервную систему, поднимает вопрос, что точный механизм переноса ЛП между нейронами, и почему распространение телец Леви следует определенному образцу, в обе нейронные клеточные линии. Первичные нейроны способны выделять α Syn мономеры, олигомеры и фибриллы через нетрадиционный кальций-зависимы экзоцитоз из крупных плотных ядер везикул или через экзосомы.

После того, как α Syn проявится в их среде, он имеет возможность занять свободные экзосомы и олигомеры эндоцитоза, после чего они разрушаются в лизосомах (SH-SY5Y клетки), в то время как мономеры, могут диффундировать через клеточные мембраны и не разрушатся. Антероградно или ретроградно через аксоны и переходят к другим нейронам, обеспечивая потенциальную магистраль для распространения между связанных областей нервной системы у пациентов с БП. Недавнее исследование показывает, что нейрон к нейрону передачи α Syn может быть инициировано путем связывания трансмембранного белка ген-лимфоцитами активации 3 (LAG3). Кроме того, мыши, лишённые LAG3 показали задержка α Syn PFF-индуцированной патологии и снижение токсичности. [18] Распространение α Syn от ENS до центральной нервной системы также была изучена в животных моделях.

Когда белок α Syn впрыскивали в стенке желудка и двенадцатиперстной кишки у крыс, он был способен распространяться через блуждающий нерв. Кроме того, внутрижелудочный ротенон обработка мышей приводило к α Syn включений в ENS, DMV, и SN, и потери клеток в SN. Эти результаты показывают, что блуждающий нерв участвует в и важное значение для распространения α Syn патологии от ENS к ЦНС у крыс и мышей. [11] Было высказано предположение, что α Syn действует прионами.

В этой теории, патологическая, неправильно уложенный α Syn является инфекционным белком распространения токсичности путем формирования токсичного шаблона, способствует неправильному сворачиванию белка соседнего α Syn, превращая его в токсичный белок. Токсичность α Syn в иной форме, до сих пор не определились и остается темой множества экспериментов. [17] Существуют также клинические доказательства, что тельца Леви, находящиеся в носовых отделах и желудочно-кишечном тракте, потенциально предшествуют развитию данного заболевания. Поэтому у больных возникают типичные жалобы нарушения пищеварения, а также проблемы с обонянием на ранних стадиях, еще до начала двигательных нарушений. Другие исследования показали, что связь между тельцами Леви и клинических симп-

томов БП должна быть поставлена под сомнение. Только 45 % людей с широко распространенной тельцами Леви в мозге диагностированы с деменцией или двигательными симптомами, и только около 10 % людей с тельцами Леви в черной субстанции и / или базальных отделах переднего мозга с диагнозом БП.



Рисунок 9. Симптомы болезни Паркинсона

Кроме того, нейродегенеративные расстройства в черной субстанции могут предшествовать образованию телец Леви. [5,9,13,19] За последние 10 лет исследования роли α Syn в патогенезе болезни Паркинсона обнаружили, что существуют дополнительные мутации миссенс (p.E46K, p.H50Q, p.G51D и p.A53E) и с дублированием α Syn, а также статистически значимую связь между некодирующими вариантами внутри и вокруг гена α Syn в спорадических заболеваниях. Совсем недавно, захватывающая работа с использованием генной инженерии показала, что полиморфный тип в интроне альфа-синуклеина является транскрипционным геном, в котором аллель вызывает положительную регуляцию экспрессии альфа-синуклеина. Heiko Braak and colleagues использовали иммуногистохимию α Syn в спорадических случаях БП, чтобы показать, что тельца Леви, содержащие α Syn, существуют и в дру-

гих отделах центральной нервной системе, например, в дорсальном моторном ядре блуждающего нерва. Для классических моторных знаков БП характерны признаки истощения дофамина в черной субстанции и полосатом теле. [4,7,8,12].

ВЫВОДЫ ПО 1-Й ГЛАВЕ

Накопление генетических данных в области изучения болезни Паркинсона в настоящее время сильно прогрессирует от определения факторов риска до уверенного прогнозирования возникновения заболевания. Для поиска новых мишеней терапии необходима реконструкция генной сети заболевания, кластеризация генов в сети, выявление ключевых генов, обладающих наибольшим числом контактов в сети. С помощью онлайн-инструментов биоинформатики OMIM, PANTHER, g:Profiler, GeneMANIA и STRING-DB мы проанализировали актуальный на данный момент массив данных, связанных с болезнью Паркинсона, рассчитали категории генных онтологий для большого списка генов, визуализировали их и построили генные сети, содержащие выявленные ключевые объекты и их взаимосвязи. Биологическая интерпретация полученных результатов всё еще остается сложной задачей. Анализ генов, связанных с болезнью Паркинсона, определение их положения в генной сети (связанности) позволяет оценить их перспективность в качестве генов-мишеней для лекарственных воздействий.

ГЛАВА 2. КОМПЬЮТЕРНАЯ РЕКОНСТРУКЦИЯ ГЕННОЙ СЕТИ

2.1. Анализ нейронных сетей для диагностики заболеваний в медицине

Наиболее эффективно ИНС используются в области кардиологии. Так, в работе Т.В. Чаши с соавторами [3] рассматривается возможность применения ИНС для прогнозирования течения постгипоксических нарушений сердечно-сосудистой системы у новорожденных детей. В данном исследовании созданы две трехслойные сети, на вход которых подавалась запись вариабельности сердечного ритма в виде последовательности значений интервалов КЯ. Обучение сетей проводилось методом случайного поиска, генетическим алгоритмом и методом обратного распространения ошибки. Последний метод оказался наиболее эффективным: авторам удалось добиться точности распознавания

СТ-Т-нарушений и дисфункции миокарда в 91 и 82 % случаях соответственно.

Особое внимание уделяется использованию ИНС для диагностики ишемической болезни сердца. Так, А.Г. Сбоев с соавторами [4] показал, что многослойный персептрон с двумя скрытыми слоями является оптимальной топологией для диагностики ишемической болезни сердца. Обучение проводилось с использованием генетической оптимизации для количества нейронов в скрытых слоях. Полученная точность диагностики коронарного атеросклероза и ишемической болезни сердца составила 96 и 94 % соответственно, что превышает точность нейросетевой модели М. Со-lak et al. [5]. О.Yu. Atkov, S.G. Gorokhova et al.

[6] в своем исследовании разработали несколько нейросетевых моделей, диагностическая точность которых варьировала от 64 до 94 %. Лучшей моделью оказался многослойный персептрон с двумя скрытыми слоями, обученный с помощью генетического алгоритма. Аналогичные показатели точ-

ности (95,5 %) диагностирования ишемической болезни сердца получены Н. Niranjana Murthy, М. Meenakshi

[7]. Структура модели, используемая в этом исследовании, представляет собой многослойную архитектуру с 13 входными нейронами, 13 скрытыми нейронами и 1 выходным нейроном. В качестве функции активации выбрана сигмоидная функция. Н. Moghaddasi et al. [8] в качестве нейросетевой модели для диагностики ишемии сердца использовали многослойный персептрон, обучение которого осуществлялось с помощью алгоритма Broyden-Fletcher-Goldfarb-Shanno (BFGS). Точность работы модели составила 73,39 %, при этом чувствительность - 93,44 % и специфичность - 28,34 %. Для повышения эффективности диагностики ишемической болезни сердца Z. Arabasadi et al. [9] предлагают гибридный метод, сочетающий генетический алгоритм и ИНС. Используя такую методологию, авторы данного исследования достигли точности 93,85 %, при этом чувствительность модели составила 97 %, а специфичность - 92 %.

С развитием нейросетевых технологий разрабатывались новые архитектуры ИНС, а также новые алгоритмы их обучения. В 2006 году появились технологии глубокого обучения сетей. Данный подход стал широко использоваться только после 2012 года. Так, в 2017 году ученые А. Caliskan и М. Yuksel [10] опубликовали научную работу, в которой описали возможность применения глубоких нейронных сетей в диагностике ишемической болезни сердца. Диагностика в данном исследовании рассматривалась как задача классификации пациентов по двум группам - «пациент здоров» и «пациент болен». Сеть обучалась в два этапа. В качестве обучающих параметров использовались возраст и пол пациента, а также показатели лабораторного анализа крови и ЭКГ. Нейронная сеть обучена на двух наборах данных, в первом случае сеть классифицировала пациентов с точностью 87,6 %, во втором - с точностью 89,7 %.

Л.Н. Ясницким с соавторами [11] предложена нейросетевая модель, которая по 51 входному параметру, характеризующему пациента и его симпто-

мы, позволяет выявить 9 заболеваний сердечно-сосудистой системы. На основе данной ИНС авторы разработали демонстрационный прототип диагностической системы, который размещен на сайте Пермской научной школы искусственного интеллекта (www.permai.ru). Эти же авторы в своем более позднем исследовании [12] модифицировали ранее разработанную нейросетевую систему, дополнив ее возможностью прогнозировать течение заболеваний в различные периоды их развития. Кроме того, увеличено число входных параметров до 62, а количество возможных диагнозов уменьшено до 6. В качестве практической значимости разработанной нейросетевой системы авторы указывают возможность моделировать различные варианты прогноза развития заболевания для каждого обследуемого пациента.

В литературе также имеются примеры применения ИНС для анализа патологических изменений в кровеносных сосудах [13], прогнозирования риска развития артериальной гипертензии [14], диагностики заболевания сахарным диабетом [15, 16].

Большое количество исследований посвящено разработке и применению ИНС в онкологии. Нейросетевые технологии применяются для диагностики рака печени [17] и поджелудочной железы [18]; дифференциальной диагностики опухолей щитовидной железы [19]; диагностики опухолей головного мозга [20]; прогнозирования рака мочевого пузыря [21]; определения групп повышенного риска заболевания раком молочной железы [22]; выявления рака предстательной железы [23, 24]. F. Ahmed [25] описал опыт использования ИНС для диагностики и прогнозирования выживаемости при раке толстой кишки.

В области пульмонологии ИНС применяются для дифференциальной диагностики интерстициальных заболеваний легких [26, 27], острой тромбоэмболии легких [28, 29]. В работе Л.С. Макаровой и Е.Г. Семеряковой [30] представлены и проанализированы результаты дифференциальной диагностики бронхиальной астмы, полученные с помощью двух методов: нейронных сетей и дискриминантного анализа. В данном исследовании модель, по-

строенная с помощью дискриминантного анализа, показала наилучшие результаты. Диагностирование пациентов на основе многослойного персептрона дало неудовлетворительные результаты. Авторы связывают это с недостаточным количеством примеров в обучающей выборке. О.В. Алексеевой и соавторами [31, 32] предложен способ дифференциальной диагностики рецидивирующей бронхолегочной патологии у детей с помощью нейросетевого анализа. Тестирование разработанных авторами нейросетевых систем показало достаточно высокую их прогностическую способность - 95 и 92 % соответственно.

Нейросетевые технологии применяются для диагностики заболеваний желудочно-кишечного тракта. Р. Maclin, J. Dempsey [33] и S. Kazmierczak et al. [34] использовали в своих исследованиях технологию нейронных сетей для дифференциальной диагностики заболеваний печени. П.И. Миرونным с соавторами

[35] исследованы возможности ИНС в прогнозировании развития абдоминального сепсиса у больных тяжелым острым панкреатитом. В работе использовалась трехслойная сеть, обученная с помощью алгоритма обратного распространения ошибки по данным 100 больных. При чувствительности 63,3 % ИНС продемонстрировала высокую специфичность - 88,2 %. Установлено, что используемая нейросетевая модель позволяет с высокой точностью прогнозировать развитие абдоминального сепсиса у данной категории больных.

Встречаются отдельные работы, посвященные анализу возможности применения ИНС для диагностики заболеваний позвоночника и костной системы. П.Н. Афонин с соавторами

[36] использовали нейросетевые технологии для прогнозирования ближайших и отдаленных результатов лечения больных гематогенным остеомиелитом позвоночника. В обоих случаях сети по своей структуре представляли многослойный персептрон. В процессе построения ИНС применен алгоритм генетического отбора входных признаков. В качестве исходных параметров

сети для прогнозирования результатов лечения остеомиелита позвоночника (на момент выписки) отобрано 5 показателей, для прогнозирования степени выраженности нарушения жизнедеятельности в отдаленном периоде - 18 показателей. Точность прогнозирования составила 92,3 % для ближайшего периода (момент выписки больного из стационара) и 90,6 % для отдаленного периода (через 1 год после завершения стационарного лечения). А.А. Ефимовым и соавторами [37] предложен нейросетевой алгоритм классификации пациентов в зависимости от стадий остеопороза с точностью 95,2 %.

В неврологии ИНС применяются для диагностики и классификации нейродегенеративных расстройств, таких как болезни Альцгеймера, Паркинсона, Гентингтона, Пика. D. Mantzaris et al. [38] предложен метод оценки когнитивных нарушений при болезни Альцгеймера на основе анализа ЭЭГ с использованием ИНС в сочетании с генетическим алгоритмом. F. Bertè et al. [39] применяли различные архитектуры ИНС для диагностики типа деменции у пациентов. В данном исследовании выделено 6 типов деменции: умеренные когнитивные нарушения, болезнь Альцгеймера, лобно-височная деменция, сосудистые когнитивные нарушения, болезнь Альцгеймера с сосудистыми когнитивными нарушениями и лобно-височная деменция с сосудистыми когнитивными расстройствами. Моделью с лучшей предсказательной способностью оказалась вероятностная нейронная сеть, состоящая из входного (30 нейронов), скрытого (65 нейронов) и выходного (6 нейронов) слоев. Данная ИНС точно диагностировала тип деменции у пациентов в 97,25 % случаев. Стоит отметить, что нейронная сеть в виде многослойного персептрона также показала хороший результат: точность прогнозирования типа деменции составила 95,60 %.

M. Quintana et al. [40] применяли ИНС для классификации пациентов по группам: здоровые, пациенты с умеренными когнитивными нарушениями, пациенты с болезнью Альцгеймера. В качестве классификатора использовался многослойный персептрон, который включал в себя 3 слоя: входной (12 нейронов), скрытый (4 нейрона) и выходной (1 нейрон). В качестве функции

активации выбран сигмоид. Обучение ИНС проводилось с помощью алгоритма обратного распространения ошибки.

На первом этапе сеть обучена на данных пациентов из всех трех групп. В результате установлено, что персептрон корректно классифицирует пациентов в 66,67 % случаев. На втором этапе для обучения сети использовались данные пациентов из здоровой группы и группы с умеренными когнитивными нарушениями. В результате сеть точно классифицировала пациентов в 98,33 % случаев. На последнем этапе сеть обучали на данных здоровых пациентов и группы пациентов с болезнью Альцгеймера. В этом случае точность классификации составила 100 %. А. Lins et al. [41] применяли ИНС для выбора параметров,

позволяющих прогнозировать развитие умеренных когнитивных нарушений и деменции у лиц пожилого возраста. Авторами предложена нейросетевая модель, использующая в качестве диагностических факторов следующие параметры: пол, возраст, уровень образования, время обучения и баллы, полученные при тестировании.

Н.С. Резниченко [42] описан опыт применения ИНС для диагностики синдрома дефицита внимания с гиперактивностью. Сеть обучена на данных психологических карт 25 детей. Точность диагностики составила 70 %. Автором сделан вывод о перспективности использования разработанной нейросетевой модели в диагностике синдрома дефицита внимания с гиперактивностью. После увеличения объема обучающей выборки и повторного обучения данной сети Н.С. Резниченко и С.Н. Шиловым [43] получены более точные показатели работы модели: ИНС выдавала правильный прогноз в 89 % случаев.

Нейросетевые технологии постепенно стали приходить в сферу психологии. Например, М.А. Беребиным и С.В. Пашковым [44] исследована возможность применения ИНС для дифференциальной диагностики и прогноза нарушений психической адаптации у сотрудников силовых структур. В данной работе сеть организована в виде трехслойного персептрона с сигмои-

дальними передаточными функциями нейронов. Первый слой сети состоял из 13 нейронов, второй (скрытый) - также из 13 нейронов, а выходной слой - из 2 нейронов, что соответствует числу диагностируемых классов уровня психической адаптации. После обучения ИНС протестирована, точность прогнозирования составила 100 %.

Е.В. Славутской и Л.А. Славутским [45] предложен нейросетевой алгоритм для селективной оценки гендерных различий в эмоционально-волевой и интеллектуальной сферах школьников 10-11 лет. В данном исследовании авторы используют двухслойную сеть с прямой передачей сигнала и обратным распространением ошибки. Полученные результаты свидетельствуют об эффективности применения ИНС, т. к. данный подход позволяет достаточно точно выделять наиболее значимые психологические признаки, определяющие гендерные различия обследуемых.

Таким образом, подходящим типом структуры ИНС, предназначенной для решения задач медицинской диагностики и прогнозирования, является персептрон с сигмоидными активационными функциями, на вход которого подается информация о пациенте, а на выходе выводится диагноз заболевания. Алгоритм обратного распространения ошибки и генетический алгоритм наиболее часто используются для обучения многослойных персептронов при диагностике различных заболеваний.

Стоит отметить, что ИНС можно использовать как математические модели рассматриваемой предметной области. Изменяя входные параметры нейросетевой модели и наблюдая за поведением выходных сигналов, можно изучать предметную область, выявлять и исследовать медицинские закономерности, которые извлекла ИНС при обучении.

В работе Аверина О.И. рассмотрен способ применения искусственной нейронной сети для определения патологии в сосудах. Данная работа является отличным примером использования модели искусственной нейронной сети [8]. Результаты обучения и проверки работоспособности спроектированной сети показывает её успешное применение для решения поставленных за-

дач и способность находить сложные закономерности и взаимосвязи между различными объектами, относящимися к одному классу данных. После того как сеть обучена, она становится надежным и недорогим инструментом анализа патологий в кровеносных сосудах.

Мнение врачей пока более авторитетно по сравнению с «мнением» нейросетей, но как второе мнение и как инструмент, упрощающий работу врача, они эффективны. Это хорошо видно на примере нейронной сети, которая распознает почечно-клеточный рак и почечные кисты. Веса и связи настроены учеными; нейросеть обучалась на цифровых данных ультразвукового исследования порядка 100 реальных случаев рака, которые уже достоверно определили. После этого нейронной сети дали в обработку 52 случая (17 злокачественных, 30 кист и 5 других) из данных УЗИ в больнице Мемфиса. Обученная нейросеть без ошибок определила 47 случаев, которых не было в данных, используемых для обучения. В этом исследовании довольно мало информации для обучения, но тем не менее эффективность работы сети велика. Самые большие трудности, почему нейронные сети пока нельзя повсеместно использовать — это то, что они еще не до конца проверены. Для их проверки требуется очень много данных, а не так-то просто создавать базы данных с медицинскими изображениями, чтобы научить нейронные сети опознавать болезни, — это не фотографии котиков и собачек. Пока в таких серьезных вопросах, как рак, нейронные сети могут быть только вторым мнением [3]. Но сейчас прикладывается много усилий, чтобы создавать базы данных, на которых нейросети могут лучше обучаться и на которых можно их использовать. Не так давно ученые создали нейросеть, которую обучили на 4000 гистологических снимках, и теперь она может определять тип рака легких с точностью 97% — даже опережая в этом живых специалистов. На рисунке 4 видно, что вывод специалистов очень близко совпадает с выводом нейросети: точки — это области, отмеченные нейросетью, а сплошные области выделили специалисты.

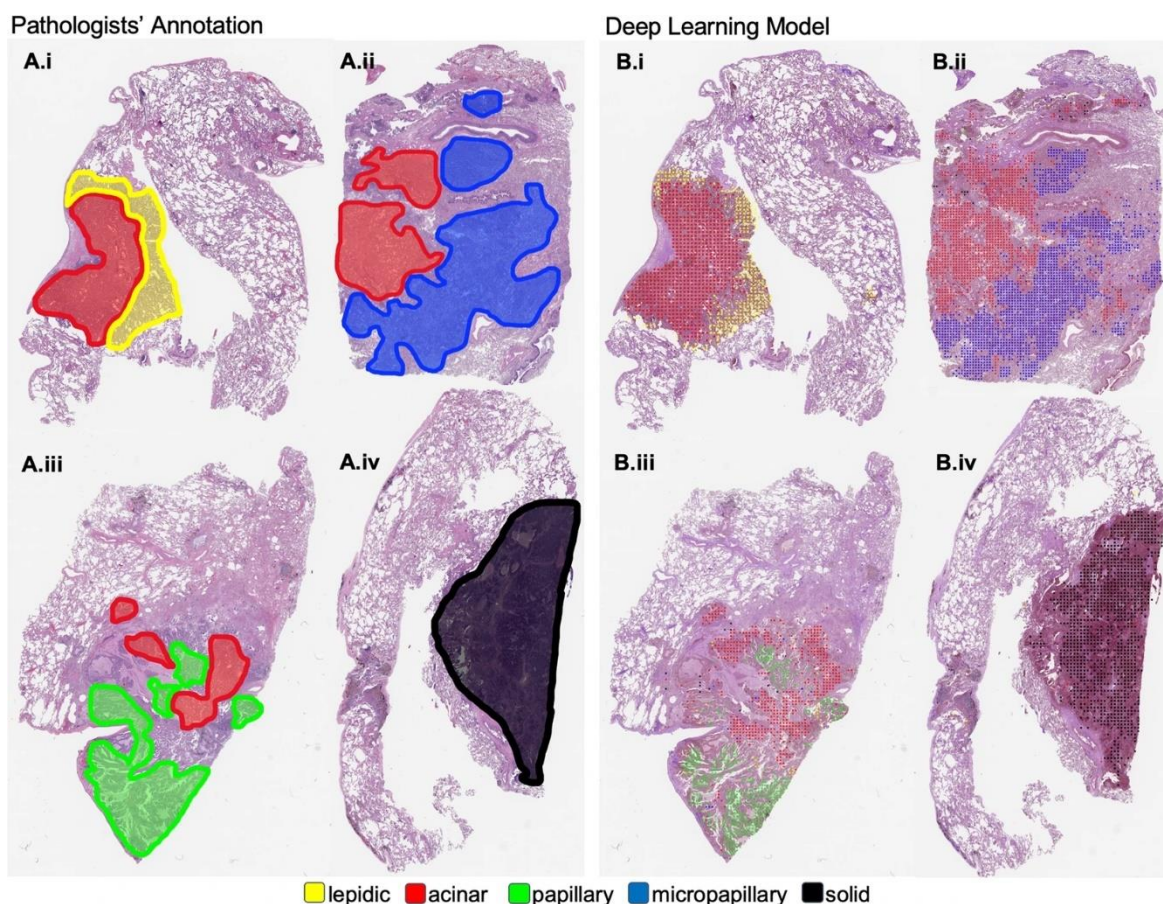


Рисунок 2.1. Оценка одних и тех же гистологических срезов профессиональными медиками (слева) и нейросетью (справа). [2]

Вывод относительно рака: нейронные сети могут выделять те образцы, с которыми должен ознакомиться специалист, и выступать в качестве второго мнения. Аналогичным образом применяют нейронные сети для мониторинга серьезных хронических заболеваний, например, диабета [2].

Нейросети можно применять для самых разных диагностик, не только рака, хотя принцип их работы сохраняется.

Взять хотя бы болезни сердца: для определения диагноза классическим методом является ЭКГ с последующей расшифровкой медиком.

К сожалению, не всегда по полученной записи ЭКГ можно определить заболевание. В особенности это относится к аритмии — довольно опасному заболеванию, при котором нарушается частота сердечных сокращений и их регулярность. Чтобы зафиксировать такое состояние, люди иногда сутками ходят с электродами для записи ЭКГ — и все равно в некоторых случаях аритмия остается не выявленной, что опасно и печально. Сейчас учеными

Стэнфорда во главе с Эндрю Энджи создана нейронная сеть, которая может даже точнее и быстрее врачей диагностировать аритмию. Когда такое нововведение внедрят в медицинскую практику, это сильно ускорит диагностику аритмии и упростит работу врачей. Эта нейронная сеть обучалась на 30 000 тридцатисекундных записях ЭКГ, а потом еще на 300 более длинных. Кроме этого, создаются нейронные сети для диагностики заболеваний позвоночника. Для обучения одной из таких нейронных сетей использовалось 250 наборов записей о состоянии пациентов. В результате точность тестирования оказалась равна 83%, что очень неплохо для такого небольшого набора исходных данных [4].

Нейросети против резистентных бактерий

Как можно использовать нейросети, которые определяют объект в медицине, мы рассмотрели. Но в медицине также применяют нейросети, которые обобщают информацию и находят закономерности среди огромных объемов данных. Люди могут не видеть закономерность в горах данных — а они есть, и нейросеть их найдет. Такие нейросети применяют для решения очень важных и актуальных проблем. Известно, что многие антибиотики начинают работать все хуже. Это происходит из-за того, что бактерии становятся устойчивы к антибиотикам. Когда бактерий становится очень много, они быстро размножаются и мутируют. После приема антибиотиков, особенно не по правилам, без соблюдения рекомендаций и назначений врача, есть ненулевая вероятность, что какая-нибудь бактерия, обладающая некоторой мутацией, выживет и даст начало огромному количеству бактерий, которые тоже будут выживать после приема антибиотика. Выжившая сильная бактерия даст начало целой колонии резистентных бактерий, которым не страшен антибиотик. Находить новые антибиотики — единственный способ бороться с проблемой. На данный момент ничего лучше еще не придумали. С каждым разом новые виды антибиотиков становится искать все сложнее и сложнее; низко висящих плодов становится все меньше, приходится рассматривать все больше веществ. Как это часто бывает, в случае, где нужно проанализировать

гигантские объемы информации, биологам помогают методы программирования.

Нейросети смогли найти новый антибиотик галицин, а также группы других веществ — потенциальных антибиотиков. Как всегда, нейросеть обучали по базе данных с уже известными антибиотиками и другими веществами. После этого через нее прогнали огромное количество разнообразных веществ, о способности к убийству бактерий которых ничего не известно.

И нейросеть выдала кандидатов, одним из которых оказался галицин — лекарство, которое изначально пробовали использовать для лечения диабета, но испытания показали его неэффективность. Зато испытания на бактериях доказали, что это новый антибиотик широкого спектра действия [5].

На рисунке 5 кратко описана суть процесса. Вначале сеть получила миллион молекул, после чего они прошли обработку в черном ящике; затем сеть выдала модель возможного антибиотика, после чего его проверили на бактериях. (Подробнее о поиске новых антибиотиков с помощью машинного обучения читайте в статье «Поиск новых антибиотиков с помощью машинного обучения» [6]).

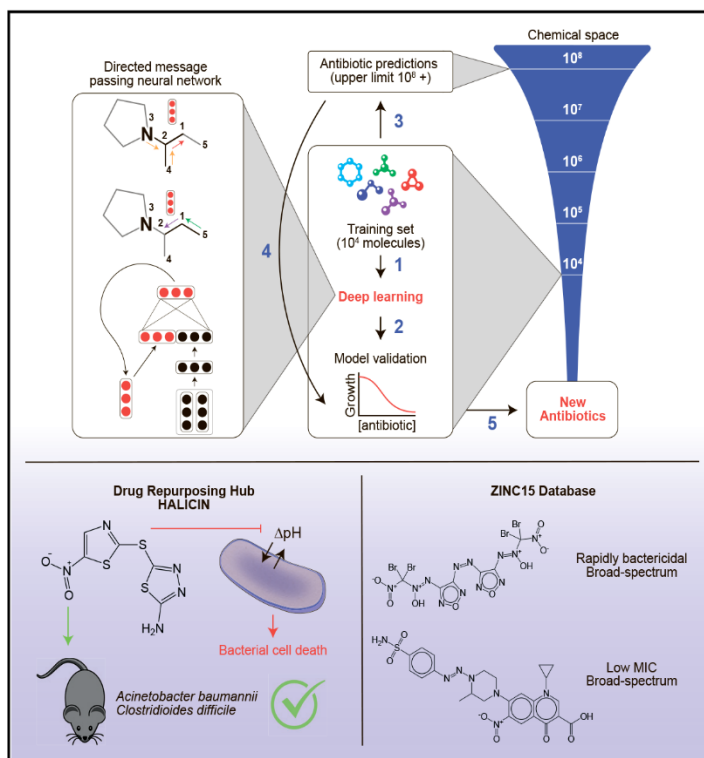


Рисунок 2.2. Схема работы нейронной сети.

Пока внедрение происходит медленно; требуется больше данных, чтобы приобрести уверенность в эффективности нейросетей и в том, что нейронные сети не будут делать роковых ошибок. Кроме того, внедрение нейросетей потребует изменения алгоритма работы медучреждений. Дополнительной проблемой является то, что по сути мы не можем узнать точно, как именно нейронная сеть пришла к выводу, поскольку скрытый слой, хоть его и настраивают люди, представляет из себя черный ящик, тогда как врач всегда может рассказать, почему он поставил такой диагноз. Нейронным сетям будет сложно доверять — мало ли что они там решили?

Возможности нейронных сетей очень велики, и если все сложится удачно, они станут превосходными помощниками врачей, смогут выполнять за них многие задачи, выступать в роли второго мнения в сложных случаях. В перспективе нейросети сделают диагностику гораздо более быстрой и эффективной, а мониторинг заболеваний будет возможен в домашних условиях. Они смогут сделать жизнь проще и врачам, и пациентам.

2.2. Определение специфики болезни Паркинсона для генных сетей

Активный интерес к изучению болезни Паркинсона обусловлен тем, что это заболевание представляет собой медицинскую и экономическую проблему для общества и на данный момент не существует методов лечения, способных остановить или обратить вспять сопровождающий это заболевание нейродегенеративный процесс.

Рост объемов генетических данных дает основу для поиска ассоциаций с заболеваниями, что отражается в пополнении таких баз данных как OMIM (<https://omim.org/>), GeneCards (<https://www.genecards.org/>). Развитие экспериментальных технологий секвенирования ведет к росту транскриптомных данных, что позволяет реконструировать генные сети / пути передачи сигнала на основе ко-экспрессии. Существующие онлайн-инструменты биоинформатики позволяют решать многие практические задачи по реконструкции генных сетей без использования дополнительных программных

средств (используется в учебном курсе для студентов Сеченовского Университета – Первого Московского Государственного Медицинского Университета Минздрава РФ). Активный интерес к изучению болезни Паркинсона обусловлен тем, что это заболевание представляет собой медицинскую и экономическую проблему для общества и на данный момент не существует методов лечения, способных остановить или обратить вспять сопровождающий это заболевание нейродегенеративный процесс.

Болезнь Паркинсона (БП) характеризовалась поздним началом, прогрессирующей потерей дофаминовых нейронов и двигательными расстройствами. Прогрессирование БП повлияло на нервную функцию и целостность. На сегодняшний день большинство исследований в основном касались заместительной терапии допамином, но появление дискинезии, вызванной L-допой, затрудняло использование препарата. И механизм БП настолько сложен, что трудно решить проблему простым добавлением лекарств. Исследователи начали сосредотачиваться на генетических основах болезни Паркинсона, ища новый метод, который может повлиять на процессы нейродегенерации в ней.

Выше рассмотрены современные методы доставки, используемые в генной терапии БП, мы также обобщили основные мишени генной терапии при лечении БП, такие как нейротрофический фактор (для регенерации), синтез нейротрансммиттера (для продления действия L-дофа) и потенциальные белки, которые могут быть мишенью для модуляции с помощью генной терапии. Наконец, мы обсудили методы РНК-интерференции, используемые при болезни Паркинсона, они могут действовать как новый класс лекарств. Мы в основном фокусируемся на эффективности и инструментальных особенностях различных генной терапии при лечении БП.

Преимущество генной терапии заключалось в том, что мы можем доставить ген в качестве агента в определенную область мозга для изменения функции и лечения БП (Стейт и Виссел, 2014), избегая при этом побочных эффектов (Аллен и Фейгин, 2014). Хотя генная терапия в настоящее время в

основном экспериментальная, многообещающее будущее заставляет многих исследователей рассматривать ее как новый класс лекарств от БП. Прежде чем мы рассмотрим генную терапию при БП, мы должны сначала уделить некоторое внимание способу доставки генной терапии (Эртель и Шульц, 2016).

Хотя меры, которые мы приняли для устранения БП, требуют большого внимания, сами меры ограничены множеством факторов, например, выбором вектора доставки. В принципе, мы можем разделить способ доставки на два типа: вирусный и невирусный опосредованные пути (Latchman and Coffin, 2001; Muramatsu et al., 2002; Lewis et al., 2014).

Генная терапия GDNF

GDNF привлек много внимания благодаря своим нейрогенеративным и нейропротекторным эффектам. *In vitro* это способствовало выживанию культивируемого вентрального дофаминергического нейрона среднего мозга (Лин и др., 1993; Кларксон и др., 1997; Эггерт и др., 1999). Затем последовали некоторые исследования, подтверждающие положительное влияние экспрессии GDNF на дегенерацию нигростриата и связанные с ней двигательные симптомы у животных, моделирующих БП (Eslamboli et al., 2003, 2005). Учитывая, что большинство менее продвинутых пациентов с БП все еще полностью реагируют на терапию L-допой, они могут не согласиться на необратимую меру. Но очень раннее вмешательство генной терапии после постановки диагноза может принести им пользу. Некоторые исследователи используют векторы мифепристора и AAV-5, которые экспрессируют GDNF, для установления прерывистого и обратимого режима контроля экспрессии GDNF, в этой системе мифепристон использовался в качестве генного переключателя, чтобы вызвать временное воздействие на экспрессию. Животные, которым вводили векторы, конститутивно экспрессирующие GDNF, показали долгосрочное и стабильное улучшение уровня GDNF, в то время как животные, которым вводили мифепристон для индукции кратковременной экспрессии, также показали надежное, но краткосрочное улучшение (Kirik et al.,

2016). Эти результаты показывают, что новые прерывистые обратимые методы также оказывают значительное и стабильное влияние на экспрессию генов по сравнению с традиционным необратимым способом.

Для менее продвинутых пациентов с БП это означает, что у них будет более безопасный и полезный выбор для предотвращения нейродегенерации. Некоторые ученые используют обработанных MPTP макаков-резусов и пожилых обезьян с паркинсоническими симптомами в качестве модельных животных, обе группы используют вектор AAV2-GDNF, введенный в путамен, в результате оказывается, что AAV2-GDNF усиливает двигательную активность и увеличивает дофаминергические терминалы в путаме (Eberling and Kells, 2009; Johnston et al., 2009).). Хотя в генной терапии GDNF есть много преимуществ, значительная потеря веса, вызванная экспрессией nigral GDNF, вызывает беспокойство. Терапия GDNF требует большего понимания и дальнейшего развития на всех уровнях. Таблица 2а представляет собой сумму генной терапии GDNF.

Таблица 2а.Описание генной терапии GDNF

Model animal	Study duration	Target area	Interference of gene therapy
6-OHDA lesioned rats	17 weeks	Striatum	Constitutive GDNF expression AAV vector Mifeprestone induced once Mifeprestone induced twice
MPTP-rhesus monkeys monkeys	12 months	Putamen	AAV2-GDNF vector
aged naïve monkeys	6 months	Putamen substantia nigra	AAV2-GDNF vector
Interference of gene therapy	Restoration of the motor control		Impact on dopaminergic pathway
	Cylinder test	APO-induced rotation	
Constitutive GDNF expression AAV vector	From 20 to 38%right forepaw use	Stable reduction of the Apo induced rotation by about 80%	DA level almost completely restored than the contralateral hemisphere
Mifeprestone induced once	From 20 to 38% right forepaw use, then declined to 25% at 15 weeks after lesion	A reduction by about 50 ~ 60% which lasts 5 weeks, and then increased at 7 weeks	Not significantly increased at the end of the experiment
Mifeprestone induced twice	From 20 to 35%right forepaw use	Rotation behavior was stable reduced over the time course of study	Significant lower than those achieved by constitutive GDNF expression
AAV2-GDNF vector	Enhanced locomotor activity		Increased density of Dopaminergic terminals in the putamen
AAV2-GDNF vector	An increase of basal locomotor activity		Small increase of AADC activity

Реконструкция регуляторных генных сетей на основе данных об экспрессии генов, белок-белковых взаимодействиях, функциональной аннотации с помощью методов биоинформатики позволяет строить модели заболеваний, что важно для определения эффективности диагностики и терапии и возможности направленного лекарственного воздействия на гены-мишени.

Активный исследовательский интерес к изучению болезни Паркинсона обусловлен тем, что это заболевание представляет собой медицинскую и экономическую проблему для общества и на данный момент не существует методов лечения, способных остановить или обратить вспять сопровождающий это заболевание нейродегенеративный процесс. Изучение причин возникновения болезни Паркинсона продолжается уже несколько десятилетий и на первых порах фокусировалось только на экологических причинах или последствиях вирусных инфекций, что позволяло считать болезнь Паркинсона архетипическим примером негенетического заболевания. В дальнейшем пристальное изучение возможных генетических причин заболевания привело к выявлению на сегодняшний день мутаций в 15 генах, ответственных за моногенные формы болезни Паркинсона [1]. Однако все эти известные моногенные формы в совокупности объясняют лишь 30% моногенных и 3-5% генетически сложных случаев этой болезни. Это означает, что современный подход к поискам причин болезни Паркинсона должен включать изучение широкого спектра сложных полигенных нарушений, которые индуцируются как генетическими, так и экологическими факторами.

Определение ключевых генов заболевания, полученных с помощью анализа структуры геной сети, необходимо для дальнейшего подбора и изучения потенциальных вариантов фармакологических веществ, способных взаимодействовать с белками, выявленными поиском, что в перспективе приведёт к созданию лекарственных средств на их основе. Решение этих задач необходимо начинать со сбора и построения актуального списка генов, ассоциированных с развитием болезни Паркинсона, с анализа категорий генов онтологий для такого списка и реконструкции геной сети.

2.3. Разработка структуры геной сети

Интернет-ресурс OMIM (Online Mendelian Inheritance in Man) (<https://omim.org/>) использован для анализа генов менделевского наследования у человека [2], поиск проводился по запросу “Parkinson disease”. Расчёт категорий генов онтологий с помощью ресурса PANTHER (Protein

ANalysis THrough Evolutionary Relationships) (<http://pantherdb.org/>) [3] выполнен анализ генных онтологий для заданного списка генов. Визуализация генных онтологий Для визуализации генных онтологий использовался ресурс GOST g:Profiler (<http://biit.cs.ut.ee/gprofiler/gost>), выполняющий статистический анализ для интерпретации предоставленных пользователем списков генов.

Данные о генах, связанных с наследственной предрасположенностью к болезни Паркинсона, пополняются и агрегируются в базах данных, таких как OMIM, GeneCards (<https://www.genecards.org/>), MalaCards. Используя ресурс OMIM, по запросу “Parkinson disease” получен актуальный список 293 наименований генов. Далее выполнен расчёт категорий и анализ генных онтологий для этого списка генов с помощью ресурса PANTHER (Protein ANalysis THrough Evolutionary Relationships) (<http://pantherdb.org/>) [3]. Из набора исходных генов часть идентификаторов не распознаны системой или не могли быть определены в геноме человека однозначно. Всего в референсном геноме человека по PANTHER использовался 20851 ген. С помощью PANTHER мы построили таблицу онтологий (таблица 1) для категорий биологических процессов. Для получения наиболее информативных результатов значения p ограничились $1,0E-25$ (коррекция Бонферрони), а превышение наблюдаемого к ожидаемому по случайным причинам (FC, fold change) — не менее 3. Приведённые в таблице 1 данные показывают, что наиболее значимыми для генов болезни Паркинсона являются категории общей регуляции клеточной смерти, регуляции клеточной смерти нейронов, регуляции апоптоза и программируемой клеточной смерти, негативная регуляция клеточной смерти. Эти данные подтверждают ключевые этиологические признаки заболевания, среди которых центральным аспектом патофизиологии болезни Паркинсона является прогрессирующая гибель дофаминовых нейронов среднего мозга и их аксональных проекций. Процессы аутофагии являются одним из основных путей внутриклеточной деградации α -синуклеина и актуальные исследования показывают, что дисфункциональная аутофагия при болезни

Паркинсона является одним из основных факторов риска развития заболевания [5].

В таблице 1 представлены также категории более общих процессов, таких как регуляция метаболических и катаболических процессов, регуляция организации клеточных компартментов, а также поведения, что, безусловно, свидетельствует о системном характере заболевания. В настоящее время все генетические исследования болезни Паркинсона ведутся по двум взаимно дополнительным направлениям. Первое направление руководствуется гипотезой общего варианта заболевания и полагает, что генетическая основа болезни является совокупным результатом большого числа общих генетических факторов, каждый из которых оказывает относительно небольшое влияние на риск заболевания, но в совокупности создаёт значительный риск. Второе направление исследований предполагает, что наибольший риск в активном развитии болезни Паркинсона дают редкие генетические варианты, для которых существуют высоко функциональные вредные аллели [6].

При расчёте категорий генных онтологий для молекулярных функций порог по значимости категорий принят на уровне $1,0E-05$ (таблица 2), также с поправкой Бонферрони и с учётом параметра FC (таблица 1). Приведённые в таблице 2 данные показывают, что наиболее значимыми являются категории связывания ферментов, белкового связывания, связывание лигазы убиквитина и убиквитин-подобных белков, а также связывание сигнального белка и белков теплового шока. Менее значимыми оказались категории регуляторов молекулярных функций и связывания протеаз.

При построении таблицы онтологий для клеточных компартментов для генов болезни Паркинсона задан порог значимости $1,0E-14$, аналогично параметрам, приведенным выше для биологических процессов и молекулярных функций (таблица 3).

Таблица 1. Категории генных онтологий для генов болезни Паркинсона по PANTHER (биологические процессы), $p=1,0E-20$

Категории генных онтологий для биологических процессов	Класс онтологий GO	Число генов	FC*	p (Бонферрони)
регуляция гибели клетки (regulation of cell death)	GO:0010941	107	4,85	1,59E-40
позитивная регуляция клеточного метаболизма (positive regulation of cellular metabolic process)	GO:0031325	140	3,24	5,10E-37
позитивная регуляция метаболических процессов (positive regulation of metabolic process)	GO:0009893	149	3,00	1,36E-36
регуляция гибели нейронов (regulation of neuron death)	GO:1901214	52	12,52	1,20E-34
регуляция метаболизма белков (regulation of protein metabolic process)	GO:0051246	121	3,49	3,12E-33
регуляция клеточного белкового обмена (regulation of cellular protein metabolic process)	GO:0032268	116	3,56	3,37E-32
регуляция программируемой гибели клеток (regulation of programmed cell death)	GO:0043067	93	4,59	3,63E-32
позитивная регуляция метаболизма азотистых соединений (positive regulation of nitrogen compound metabolic process)	GO:0051173	129	3,16	7,67E-32
регуляция транспортировки (regulation of transport)	GO:0051049	99	4,22	8,09E-32
регуляция процессов апоптоза (regulation of apoptotic process)	GO:0042981	91	4,61	2,03E-31
регуляция организации клеточных компонентов (regulation of cellular component organization)	GO:0051128	110	3,49	3,45E-29
регуляция клеточных катаболических процессов (regulation of cellular catabolic process)	GO:0031329	69	6,04	5,84E-29
регуляция локализации (regulation of localization)	GO:0032879	117	3,20	2,53E-28
позитивная регуляция процессов белкового обмена (positive regulation of protein metabolic process)	GO:0051247	88	4,29	6,94E-28
негативная регуляция гибели клеток (negative regulation of cell death)	GO:0060548	72	5,40	1,48E-27
позитивная регуляция процессов клеточного белкового обмена (positive regulation of cellular protein metabolic process)	GO:0032270	85	4,37	2,73E-27
регуляция катаболических процессов (regulation of catabolic process)	GO:0009894	72	5,28	5,73E-27
регуляция процессов модификации белков (regulation of protein modification process)	GO:0031399	89	4,08	8,85E-27
поведение (behavior)	GO:0007610	55	7,52	1.80E-26

Данные таблицы 3 позволяют выявить наиболее значимые категории в проявлении болезни Паркинсона — соматодендритный компартмент, формирование нейронов, аксонов, пресинапсы, синапсы и другие, связанные с нервной тканью, категории.

Таблица 2. Категории генных онтологий для генов болезни Паркинсона по PANTHER (молекулярные функции), порог значимости 1,0E-03

Категории генных онтологий для молекулярных функций	Класс онтологий GO	Число генов	FC	<i>p</i> (Бонферрони)
связывание ферментов (enzyme binding)	GO:0019899	92	3,49	1,72E-23
специфическое связывание белкового домена (protein domain specific binding)	GO:0019904	41	4,39	1,87E-11
связывание киназ (kinase binding)	GO:0019900	41	4,06	2,22E-10
связывание белков теплового шока (heat shock protein binding)	GO:0031072	19	10,95	3,05E-10
связывание убиквитин протеин-лигазы (ubiquitin protein ligase binding)	GO:0031625	26	6,56	5,45E-10
связывание убиквитин-подобной протеин-лигазы (ubiquitin-like protein ligase binding)	GO:0044389	26	6,16	2,05E-09
связывание протеинкиназы (protein kinase binding)	GO:0019901	37	4,10	3,09E-09
связывание протеазы (protease binding)	GO:0002020	14	7,71	3,56E-05
связывание тау-белка (tau protein binding)	GO:0048156	9	14,98	9,56E-05

Таблица 3. Категории генных онтологий для генов болезни Паркинсона по PANTHER (клеточные компартменты), порог значимости 1,0E-10

Категории генных онтологий для клеточных компартментов	Класс онтологий GO	Число генов	FC*	<i>p</i> (Бонферрони)
митохондрии (mitochondrion)	GO:0005739	75	3,41	7,74E-18
соматодендритный компартмент (somatodendritic compartment)	GO:0036477	54	4,65	2,02E-17
проекция нейронов (neuron projection)	GO:0043005	67	3,65	5,65E-17
перикарион (cell body)	GO:0044297	43	5,51	7,63E-16
синапс (synapse)	GO:0045202	62	3,45	2,44E-14
аксон (axon)	GO:0030424	43	4,95	3,11E-14
нейрональное клеточное тело (neuronal cell body)	GO:0043025	38	5,56	8,18E-14

Рассмотрим функциональную аннотацию генов для болезни Паркинсона, интегрируя источники информации с помощью ресурса GeneCards [13]. Из более чем 9000 генов (по запросу “Parkinson disease”), наиболее значимыми являются 15 генов (таблица 4).

Важное место в этом списке занимает ген SNCA, кодирующий белок α -синуклеин; мутации в этом гене приводят к развитию аутосомно-доминантных форм заболевания, а тяжесть заболевания коррелирует с числом копий гена SNCA. Мутации в гене LRRK2 идентифицированы как причины аутосомно-доминантной природы болезни Паркинсона как наиболее распространенной моногенной формы болезни, выявленной к настоящему времени [14]. Генетические варианты в LRRK2 связаны с большинством всех известных наследуемых проявлений болезни Паркинсона. Совместные исследования показали, что риск развития болезни Паркинсона у лиц, наследующих вариант LRRK2 p.G2019S, варьирует от 28% в 59 лет до 51% в 60 лет и

достигает 75% в 80 лет [15]. Эти же гены представлены в генной сети (рисунок 3) с наибольшим числом связей.

Связь между заболеваемостью болезнью Паркинсона и локусом, содержащим гены MАРТ, наиболее интересна для изучения, поскольку мутации в MАРТ и сопутствующая им таупатия преимущественно связаны с деменцией. Растущее количество доказательств показывает, что MАРТ Н1 и его субгаплотип Н1с связаны с повышенным риском развития болезни Паркинсона. Предполагается, что специфические для гаплотипа различия в экспрессии и потенциально альтернативном сплайсинге транскриптов MАРТ влияют на клеточные функции на разных уровнях, что, в конечном итоге, повышает их восприимчивость [16]. Широко распространено предположение, что белки, кодируемые рецессивными генами, ассоциированными с ранним началом болезни Паркинсона (PINK1, PARK2, DJ-1 и FBXO7) участвуют в системе контроля качества митохондрий и её регуляции [17]; считается, что митохондриальная дисфункция играет ключевую роль в ранних событиях начала болезни Паркинсона.

Реконструкция генных сетей, основанная на анализе актуальных данных генов, ассоциированных с болезнью Паркинсона, приводит к выявлению сетевых структур, функционально связанных с генами риска заболевания. Обнаружение функциональных связей открывает путь к созданию новых лекарственных препаратов с точным направленным действием и минимумом побочных эффектов, выявляет новые стратегические пути к терапевтической стимуляции гаснущих нейронных процессов.

. Системное исследование этиологии болезни Паркинсона требует объединения клинических и биоинформатических исследований с их многофакторной интеграцией данных для построения более точных моделей манифестации и прогрессирования этого заболевания.

Этиология генетически сложной болезни Паркинсона представляет огромный набор возможных взаимодействий генетических рисков и окружающих факторов.

РЕКОНСТРУКЦИЯ ГЕННОЙ СЕТИ БОЛЕЗНИ ПАРКИНСОНА

Таблица 4. Наиболее релевантные гены болезни Паркинсона по данным GeneCards

№	Ген	Описание гена	Счёт GIFtS*	Счёт релевантности**
1	LRRK2	Обогащённая лейцином повторяющаяся киназа 2 / Leucine Rich Repeat Kinase 2	50	144,68
2	SNCA	Альфа-синуклеин / Synuclein Alpha	50	142,38
3	PRKN	Паркин (лигаза) / Parkin RBR E3 Ubiquitin Protein Ligase	40	132,53
4	PARK7	Белок 7 болезни Паркинсона / Parkinsonism Associated Deglycase	47	115,67
5	GBA	Бета-глюкозилцерамидаза / Glucosylceramidase Beta	48	108,63
6	PINK1	PTEN-индуцированная киназа 1 / PTEN Induced Kinase 1	48	102,39
7	MAPT	Тау-белок, ассоциированный с микротрубочками / Microtubule Associated Protein Tau	51	99,66
8	SYNJ1	Синаптозин 1 / Synptojanin 1	46	84,67
9	UCHL1	Убиквитин карбоксил-терминальная гидролаза L1 / Ubiquitin C-Terminal Hydrolase L1	52	78,49
10	LOC106627981	Рекомбинантный район GBA / GBA Recombination Region	2	77,69
11	APOE	Аполипротеин E / Apolipoprotein E	51	74,04
12	SNCAIP	Белок, взаимодействующий с альфа-синуклеином / Synuclein Alpha Interacting Protein	45	73,42
13	PSEN1	Пресенилин 1 / Presenilin 1	52	73,12
14	APP	Предшественник бета-амилоида / Amyloid Beta Precursor Protein	52	65,24
15	VPS13C	Гомолог В белка вакуолярной сортировки белков 13 / Vacuolar Protein Sorting 13 Homolog C	35	64,49

Примечание: * — счёт (оценка) функциональной аннотации гена GIFtS (GeneCards Inferred Functionality Scores) по представленности в базах данных, следуя [13]; ** — счёт релевантности оценивает соответствие поисковому запросу в базе данных для болезни Паркинсона “Parkinson disease”.

Составление списка ключевых генов, ассоциированных с развитием болезни Паркинсона, проведенный анализ категории генных онтологий для этого списка и реконструкция генной сети позволили выбрать гены-мишени для лекарственных воздействий, модифицирующих степень выраженности болезни. Существующие методы терапии позволяют продлить время жизни пациента и улучшить её качество, но пока не направлены непосредственно на причину заболевания, а носят скорее заместительный характер, что говорит о необходимости дальнейшего исследования генетических основ болезни. Появились работы, показывающие, что генная терапия становится перспективным подходом для лечения неврологических расстройств, включая болезнь Паркинсона [18].

ВЫВОДЫ ПО 2-Й ГЛАВЕ

Накопление генетических данных в области изучения болезни

Паркинсона в настоящее время сильно прогрессирует, продолжается рост числа публикаций по данной теме от определения факторов риска до уверенного прогнозирования возникновения самого заболевания [19]. Компьютерные методы применяются не только для функционального анализа генов, их классификации и установления взаимосвязей, но и широко используются при экспериментальном изучении структуры и взаимодействия генов и регуляции их экспрессии в клетках мозга [20-22].

Сопоставление списков генов и генных сетей для различных заболеваний — болезни Паркинсона, болезни Альцгеймера, интеграция существующих баз данных в таких ресурсах как GeneCards [13] и MalaCards [23] (<http://www.malacards.org/>) даёт возможность выявления ключевых (узловых) генов, общих для этих заболеваний в качестве возможных мишеней для терапии. Применение всего спектра экспериментальных и компьютерных методов способно уже в ближайшем будущем приблизить создание адекватных лекарственных средств, облегчающих состояние больных болезнью Паркинсона.

Глава 3. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ГЕННОЙ СЕТИ

3.1. Формирование технических требований к программе

Реконструкция генных сетей на основе микрочиповых данных — это глобальная и сложная задача, на данный момент не существует однозначных методов ее решения. Возможным подходом является комбинирование разных моделей, но решать какая модель адекватнее можно только исходя из биологических или экспериментальных соображений. В работе рассматриваются методы, включающие статистический анализ микрочиповых таблиц и модели регуляции с помощью дифференциальных уравнений. Статистический метод разработан специально для анализа данных микрочиповых экспериментов и уже зарекомендовал себя, как достаточно надежный.

В качестве моделей регуляции рассмотрена самая простая линейная модель с задержкой и взятая из литературы стохастическая. Модели регуляции рассмотрены исключительно с математической точки зрения, конечно, для полноценного анализа необходимо привлечение биологии.

Генная сеть — это совокупность совместно экспрессирующихся, влияющих друг на друга генов, выполняющих в организме некоторую специфическую функцию. Примером может служить процесс апоптоза, т.е. программируемой клеточной смерти. Апоптоз нужен для уничтожения старых или неправильно функционирующих (с повреждениями ДНК) клеток. Значительное усиление этого процесса ведет к массовой гибели клеток, в частности, вирус СПИДа провоцирует апоптоз клеток иммунной системы, в то время как ослабление, как правило к раку. Генную сеть апоптоза составляют, как гены-инициаторы апоптоза, так и гены, блокирующие его. Решение о гибели клеток происходит по совокупности их работы.

Важной задачей является изучение процессов межгенных взаимодействий и моделирование работы генной сети. Будучи построенным, такие модели могут дать информацию о реакции организма на различные условия, на точные причины многих генетически обусловленных болезней, таких как раковые опухоли, что крайне важно при создании новых лекарств, которые бу-

дуют направлены на регулирование работы конкретных генов с минимумом возможных побочных действий.

Исходными данными для решения этой задачи служит информация, получаемая об уровне генной экспрессии в результате микрочипового эксперимента.

Технология ДНК-микрочипов появилась сравнительно недавно, около 20 лет назад. Однако, широкие перспективы ее применения в биологии и медицине привели к ее бурному развитию в последнее десятилетие. Сейчас уже можно говорить, что именно эта технология служит основным поставщиком информации для анализа изменения экспрессии генов.

ДНК-микрочип — это миниатюризированные матрицы, в которых на подложке в определенном порядке распределены и прикреплены фрагменты ДНК, соответствующие отдельным генам или их частям. Такие организованные матрицы позволяют проводить эксперименты по одновременному анализу структуры и экспрессии тысяч генов.

С помощью микрочипов можно изучать:

- экспрессию генов в различных тканях.
- изменение экспрессии генов в ходе клеточного цикла и в ходе развития.
- экспрессию генов в норме и при патологии (в нормальных и раковых клетках).
- изменение экспрессии генов с течением времени как результат внешнего воздействия (взаимодействие клетки с патогеном, лекарством).
- наличие чужеродной ДНК/РНК вирусов для диагностики различных заболеваний (например, ЗППП)
- анализ мутаций и полиморфизмов.

Основываясь на характере числовых данных в чипе, можно построить две независимые классификации микрочипов:

- статические (данные замерены единовременно).
- динамические (данные распределены во времени).

А также

- с конкурентной гибридизацией (тестовый образец против контроля).
- без конкурентной гибридизации (только тестовый образец).

Здесь возможны любые комбинации. В статических экспериментах проводятся единовременные замеры экспрессии генов в различных образцах (пробах), в то время как в динамических, экспрессия каждой пробы замеряется многократно в некоторые последовательно идущие моменты времени. Данные, полученные в результате статического эксперимента (например, замеры экспрессии генов в раковых клетках, взятых у различных пациентов), предполагаются независимыми и одинаково распределенными, в то время как данные, полученные из динамического эксперимента, зачастую проявляют сильную автокорреляцию между последовательно расположенными точками.

В случае конкурентной гибридизации сравнивается экспрессия в исследуемых образцах (например, опухолевых тканях) и в норме. Данные представляют собой числовое отношение. Гены в двух образцах конкурируют за место на подложке. Также существуют микрочипы, в которых исследуется экспрессия генов только в исследуемых тканях без контроля.

Хотя, за прошедшее десятилетие данная технология сильно эволюционировала, стала достаточно надежной, все же очень часто разнообразные технические сбои и неточности в планировании эксперимента приводят к различным артефактам в результатах, что может сильно осложнять их последующий анализ.

Выявление генов, изменивших уровень экспрессии

Задача выявления генов, изменивших свою экспрессию вследствие некоего воздействия либо в течение клеточного цикла очень важна, т.к. позволяет выявить, например последствия приема того или иного лекарства или отобрать гены с отличающимся поведением для последующего анализа. Перед тем как начинать строить зависимости между генами необходимо отобрать те из них, которые нас интересуют, а именно гены, изменившие

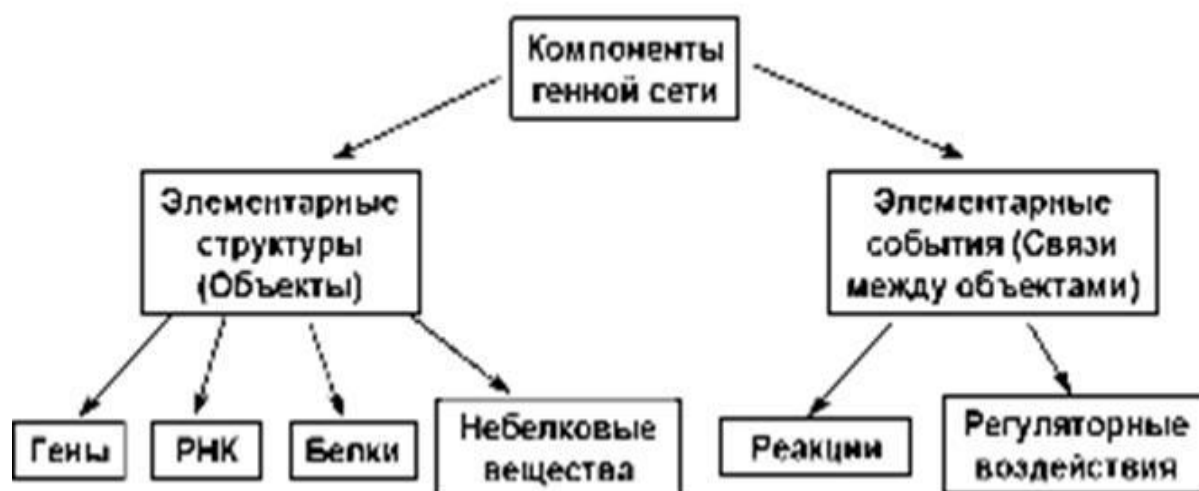
свою экспрессию. Т.к. работать с целым списком (обычно порядка 40000 клонов) слишком трудоемко и к тому же бессмысленно. Опишем те методы, которые были реализованы в плагине и использовались для анализа.

Задача ставится следующим образом: необходимо выделить из общей группы те пробы, которые значимо изменили свою экспрессию либо с течением клеточного цикла, либо в результате воздействия на пациентов. Часто данные представляют собой отношение экспрессии объекта у пациента к экспрессии данного объекта в норме.

Для построения модели любой ГС используют информацию об ее элементах и связях между ними. Это означает, что необходимо знать какие гены участвуют в формировании сети, их регуляторные элементы (промоторы и др.), синтезируемые ими белки, пути передачи сигналов в клетке. Привлечение этой информации идет путем обобщения литературных данных и получения новых данных в ходе натурных экспериментов с конкретными ГС. Далее, составляют уравнения, характеризующие динамику ГС, и подбирают значения параметров – констант, входящих в эти уравнения.

Основной проблемой стало отсутствие эффективных компьютерных технологий для ведения информации об их структурно-функциональной организации, прежде всего для эукариот. С целью накопления и обобщения информации создан целый ряд баз данных. В них содержится качественная, описательная информация (о структурно-функциональной организации генной сети, механизмы различных процессов, организация регуляторных участков генов и т.д.), а также количественная информация (параметры ферментативных реакций, массы белков, и т.д.).

В качестве примера приведена база данных GeneNet. База данных является частью системы GeneExpress. Классификация компонентов ГС в БД GeneNet представлена на рисунке 3.1.



**Рисунок 3.1. Классификация компонентов генной сети
в системе GeneNet**

Для моделирования ГС в GeneNet разработан специальный язык и удобный графический интерфейс. Можно, в частности, указать элемент модели ГС, изменить его характеристики и проследить на графике ответные изменения в поведении различных компонентов ГС.

Технология GeneNet позволяет эксперту реконструировать ГС на основе экспериментальных данных, опубликованных в научных статьях. В базе данных описано несколько сотен генов, белков и небелковых веществ 43 организмов по данным из сотен публикаций. В результате удалось построить модели некоторых ГС: липидного метаболизма, кроветворения, противовирусного ответа, функционирования эндокринной системы, регуляции развития семени, азотфиксации, теплового шока, регуляции холестерина и др.

Графические схемы функционирования генных сетей

Генные сети в упрощенных вариантах моделей описывают на языке генетических элементов (ГЭ). Каждый из них содержит ген, соответствующие ему мРНК и белки. Функционирование генетического элемента в простейшем случае – процесс синтеза белка. В генетический элемент неявно включаются все процессы, обеспечивающие синтез мРНК, белка, формирование мультимера белка (мономерная форма белка есть частный случай

мультимерной). Белки обеспечивают регулирующие взаимодействия между ГЭ.

Регуляторный механизм ГС – элементарный процесс, посредством которого некоторая совокупность генетических элементов действует на конкретный генетический элемент, уменьшая или увеличивая скорость синтеза его продукта. В реальной ГС кроме белков генетические элементы, возможно, содержат низкомолекулярные соединения, которые также регулируют состояния других ГЭ.

Схему функционирования ГС, включающую ГЭ и их воздействия, можно задать ориентированным графом. Он состоит из конечного множества вершин (g_i) и набора упорядоченных их пар, изображаемых в виде стрелок. Любая стрелка называется дугой, или ориентированным ребром. Она отражает направление и характер воздействия (отрицательное или положительное регуляторное воздействие) одного ГЭ на другой. Пример схемы функционирования простой гипотетической ГС – на рис. 6. Отметим, что ген 1 не регулирует экспрессию гена 2 непосредственно, но может повлиять на нее через ген 3. В некоторых случаях отрицательные регуляторные связи на схемах функционирования ГС обозначают пунктирными, а положительные – сплошными стрелками.

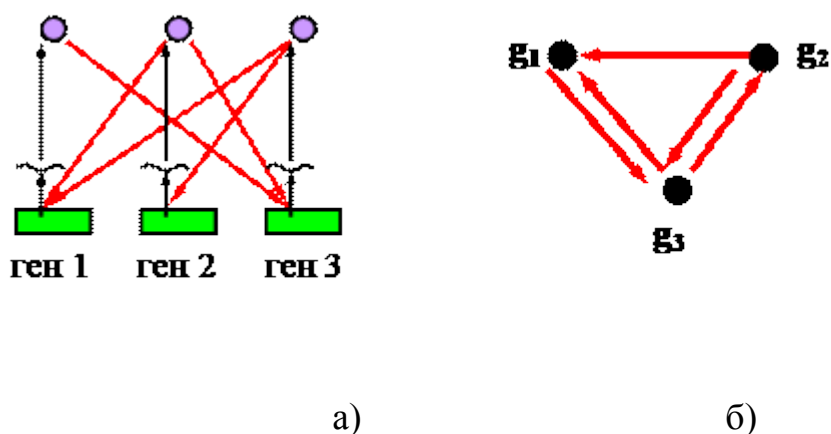


Рисунок 3.2 а) Исходная ГС из трех генов (внизу), их белков (наверху) и регуляторных воздействий (стрелки) на другие гены; б) соответствующий ей ориентированный граф – структурная схема функционирования ГС.

На рисунке 3.3 в этих терминах, но в других обозначениях, показан фрагмент схемы реальной ГС прорастания семени растения. Элементы гиббереллины (gibberelins) и абсцизовая кислота (ABA) соответствуют низкомолекулярным соединениям, остальные ГЭ – гены и их белки. Стрелки, помеченные кружками с плюсом – положительные (усиливающие) регуляторные воздействия, с минусом – отрицательные (ингибирующие).

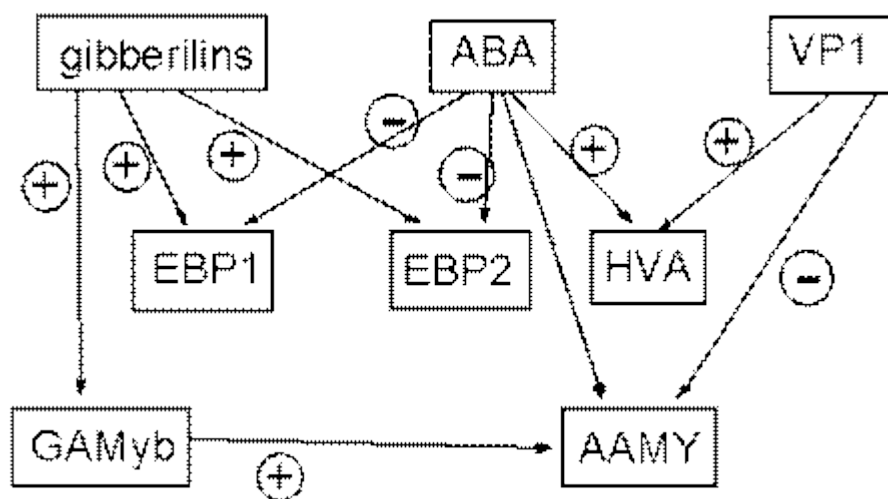


Рисунок 3.3. Граф фрагмента внутриклеточной ГС

Графическое моделирование – это простейший способ описания и первичного анализа структурно-функциональной организации ГС. В некоторых случаях непосредственно по структурному графу можно выявить важные регуляторные контуры, мотивы, а также предсказывать свойства ГС, связанные с предельными режимами ее функционирования. Например, в некоторых случаях удастся оценить число точек покоя или предельных циклов.

В других случаях непосредственно по графу ГС удастся выявить сочетание из двух или большего числа продуктов синтеза, которые даже при длительном функционировании будут присутствовать в ГС. Такие сочетания называются предельными циклами, т.к. их концентрации продолжают циклически меняться во времени. Однако по графу ГС невозможно оценить закономерности изменения концентраций во времени, для этого их необходимо дополнить непрерывными моделями.

Таким образом, графическое моделирование используется совместно с более сложными математическими моделями.

Дискретный способ моделирования генных сетей

Существует два основных количественных детерминистических подхода к описанию ГС – дискретный и непрерывный. В некоторых случаях эти математические модели объединяют, а также пытаются учесть случайную составляющую процессов в ГС с помощью стохастических моделей. Последние применяют, в частности, для таких генных сетей, в которых значительные события обеспечиваются очень малым количеством регуляторного агента (например, транскрипционного фактора).

В простейших вариантах дискретных моделей ГС предполагается, что каждый ген в ГЭ может находиться лишь в одном из двух состояний: 1 - активен (экспрессируется) и 0 - неактивен. Для простоты предполагается, что процесс функционирования ГС состоит из элементарных этапов (шагов). На очередном шаге состояние каждого гена сохраняется или меняется на противоположное. Это зависит от активизирующих и ингибирующих воздействий на него со стороны других генов, активных перед очередным шагом. Исчерпывающей характеристикой ГС в рамках подобной модели является вектор состояний. Это набор нулей и единиц, упорядоченный по номерам ГЭ, представленных на структурном графе ГС.

Алгоритм определения состояния каждого ГЭ по шагам формализован в компьютерной программе. Для этого используют специальные функции, оперирующие только с 0 и 1 (алгебра логики – булева алгебра), или пороговые функции.

Главным преимуществом дискретных моделей ГС является их простота. Такие модели позволяют проследивать динамику ГС (изменение вектора состояний) в течение длительного времени, прогнозировать предельные режимы функционирования регуляторных контуров генных сетей. Но, безусловно, подобные модели являются довольно грубым упрощением.

Непрерывные модели

Наиболее сложная задача – изучение динамики количественных переменных ГС. Сюда относятся концентрации продуктов: «входных» - действующих на ГС извне, «внутренних» - регуляторов процессов и «выходных» - выделяемых сетью во вне. Для количественного описания динамики этих непрерывных количественных переменных во времени необходимы сложные модели. Наиболее полное описание обеспечивают дифференциальные уравнения и их системы, хотя этот подход имеет ряд недостатков.

С помощью непрерывных моделей пытаются ответить на некоторые вопросы теории ГС. Среди них:

- выявление условий существования в ГС множества состояний устойчивых во времени, т.е. состояний, равновесных по значениям основных переменных, а также выявление условий переключения между такими состояниями.

- изучение «степени устойчивости» положений равновесия и влияния на нее различных факторов (мутаций, гетерозиготности, стрессов и пр.). Понятно, например, что наличие отрицательной ОС увеличивает, а положительной - снижает устойчивость равновесия.

- нахождение условий возникновения колебательных, циклических режимов функционирования ГС. В частности, оценка минимального уровня сложности сети и необходимости присутствия в ней положительных и отрицательных ОС.

Кроме теоретических исследований ведутся работы по математическому моделированию реальных ГС и сравнению результатов с данными экспериментов.

Для построения непрерывных моделей требуется формализовать, например, с помощью дифференциальных уравнений «элементарные события» ГС: функционирование регуляторных механизмов, связанных с управлением процессами транскрипции, трансляции, биохимические реакции, включая формирование белков, передачу внешних и внутренних сигналов, работу контуров управления и пр.

В рамках непрерывных и дискретных моделей возможен гибридный подход к моделированию динамики ГС. В нем разделяют сеть на 2 подсистемы. Первая – регуляторная, включает совокупность регуляторных генов и их макромолекулы. Ее состояния моделируются на языке дискретных моделей (например, с использованием пороговых функций). Вторая – исполняющая, содержит гены, обеспечивающие выполнение различных биохимических процессов. Она моделируется системами дифференциальных уравнений. В итоге общая модель формулируется на языке кусочно-непрерывных систем уравнений. Частные примеры подобных элементарных событий и уравнений представлены на рисунках 3.4-3.6.

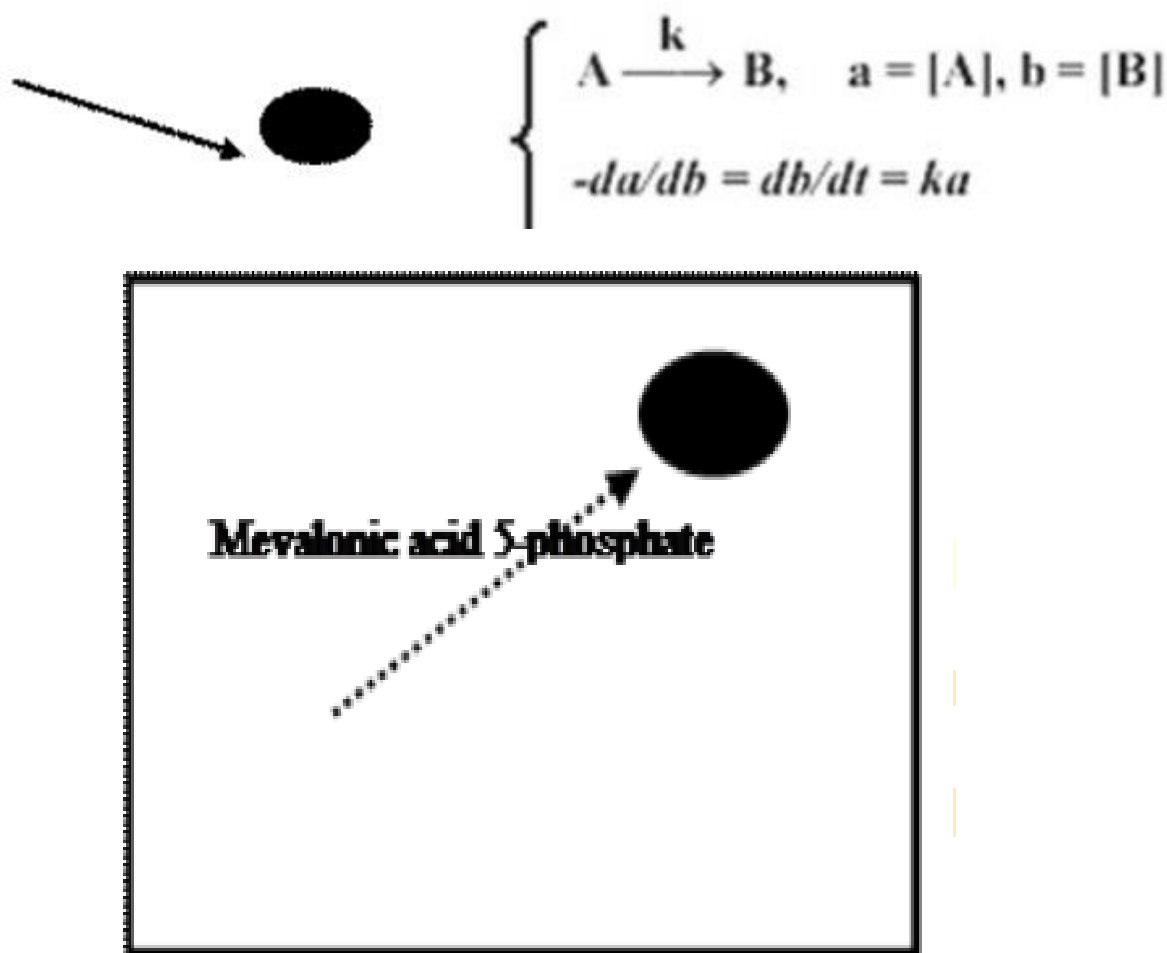


Рисунок 3.4. Схема и дифференциальные уравнения – описание химико-кинетической закономерности простейшей мономолекулярной биохимической реакции фосфорилирования. Здесь a и b – концентрации

исходного и результирующего продуктов реакции, k – параметр, характеризующий скорость прохождения реакции.

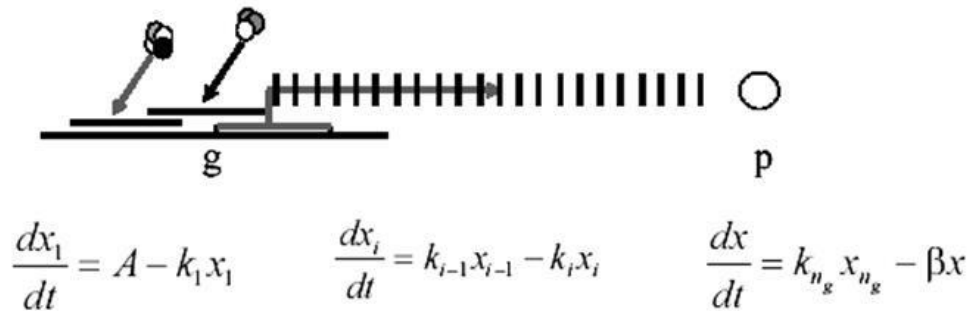
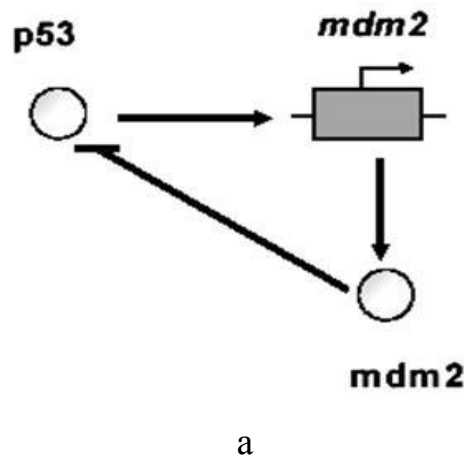
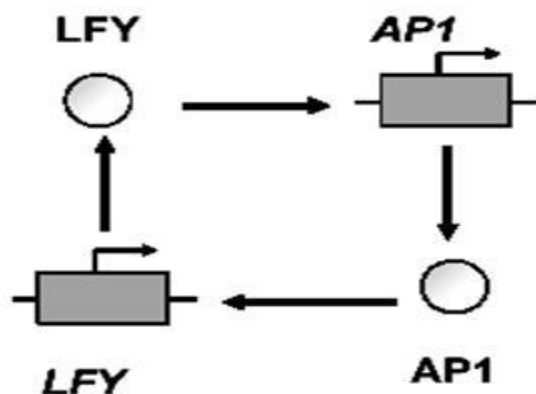


Рисунок 3.5. Схема и дифференциальные уравнения, описывающие функционирование генетического элемента сети при многоэтапном формировании регуляторного белка p .

Здесь n_g – количество промежуточных стадий синтеза белка, x_i , – концентрации промежуточных продуктов синтеза белка, x – концентрация белка, A – параметр – скорость инициации конститутивного синтеза, k_i – параметры – скорость перехода из i -го промежуточного состояния в $(i+1)$ -ое промежуточное состояние, β – параметр – скорость деградации белка p .





б

Рисунок 3.6 Схема некоторых контуров ГС с обратной связью

а) Отрицательная ОС: транскрипционный фактор гена p53 активирует транскрипцию его гена-ингибитора mdm2. Уравнение для переменной x – концентрации транскрипционного фактора гена имеет вид: $dx/dt = \alpha / (1 + Kc \cdot xc)$. α , K , c – параметры реакций. Скорость синтеза белка (dx/dt) убывает с ростом концентрации x .

б) Положительная ОС: взаимное усиление транскрипции двух генов. Уравнение для переменной x – концентрации транскрипционного фактора гена AP1 имеет вид: $dx/dt = \alpha / (1 + Kc1xc) / (1 + Kd2xd)$. Скорость синтеза белка (dx/dt) возрастает с ростом концентрации x .

В рамках непрерывных и дискретных моделей возможен гибридный подход к моделированию динамики ГС. В нем разделяют сеть на 2 подсистемы. Первая – регуляторная, включает совокупность регуляторных генов и их макромолекулы. Ее состояния моделируются на языке дискретных моделей (например, с использованием пороговых функций). Вторая – исполняющая, содержит гены, обеспечивающие выполнение различных биохимических процессов. Она моделируется системами дифференциальных уравнений. В итоге общая модель формулируется на языке кусочно-непрерывных систем уравнений.

Подобные принципы построения моделей функционирования ГС заложены в обобщенном химико-кинетическом методе моделирования. Вначале проводится иерархическая декомпозиция исходной ГС до элементарных

структур и процессов. Затем процессы формализуются в терминах стандартных блоков – систем уравнений. Из них проводят сборку имитационной компьютерной модели всей ГС.

Динамика отдельных молекулярно-генетических систем в ОХКММ формально описывается моделями общего вида:

$$\frac{dx_i}{dt} = F_i(X, K) - x_i G_i(X, K) \quad (3.1)$$

Здесь $X = (x_1, \dots, x_n)$ – вектор динамических переменных модели ГС, т.е. концентраций регуляторных продуктов. $K = (k_1, \dots, k_m)$ – вектор параметров, входящих в уравнения, $F_i(X, K)$ и $G_i(X, K)$ – рациональные функции.

В ОХКММ учитываются временные задержки и локализация процессов в однородных областях, закономерности пассивного и активного переноса веществ и энергии между областями, возможная полиплоидность и гетерозиготность генов, случайные составляющие процессов и пр. В результате изучение динамики функционирования ГС можно рассматривать как математическую проблему анализа решений сложных систем уравнений.

3.2. Выбор языка программирования

Благодаря быстродействию, простому синтаксису и богатому набору инструментов Python стал любимым языком учёных и исследователей. Разработаны специальные библиотеки для Python, которые позволяют строить графики, проводить исследования и вычисления:

- SciPy — набор инструментов для научных вычислений.
- NumPy — расширение, которое позволяет оперировать матрицами и многомерными массивами.
- pandas — библиотека для аналитики данных.
- Matplotlib — библиотека для построения графиков различной сложности.

Для языка программирования Python существует множество свободных ML-инструментов:

- TensorFlow;
- scikit-learn;
- Keras;
- pandas.

3.3. Разработка алгоритма функционирования генной сети, ассоциированных с болезнью Паркинсона

Генные сети могут функционировать как на уровне отдельных клеток (тогда данный тип клеток будет включать все структурные компоненты генной сети), так и на уровне ткани, органа либо целого многоклеточного организма. В этом случае в реализации обратных связей генной сети будут участвовать гены и белки, экспрессирующиеся в различных типах клеток. Рассмотрение структурно-функциональной организации генных сетей (Колчанов и др., 2000) позволяет выделить следующие базовые принципы их организации.

Существование большого разнообразия молекулярных механизмов, обеспечивающих функционирование обратных связей, в том числе за счет изменения интенсивности процессов транскрипции, сплайсинга, трансляции, посттрансляционных модификаций белков, ферментативных реакций, транспорта и т. д. 2. Наличие в каждой сети «центральных» генов и белков (ключевых регуляторов), обеспечивающих координацию функций остальных генов этой сети (рис. 3.7, б). 3. Кассетный механизм активации транскрипции больших групп генов транскрипционными факторами (рис. 3.7, б). 4. Авторегуляция генной сети, осуществляемая благодаря функционированию регуляторных контуров с положительными и отрицательными обратными связями и направленная на поддержание определенного, характерного для каждой генной сети типа динамики (рис. 3.8, а, б). Некоторые типы динамики критических параметров генных сетей приведены на рис. 3.8, в–е. Молекулярная основа работы регуляторных контуров, управляющих функционированием генных сетей и определяющих тип их динамики, – наличие сайтов-мишеней в

ДНК, РНК и белках, с которыми могут взаимодействовать различные молекулярные компоненты генной сети и внешние регуляторные факторы. Особенность процесса формирования любого сложного фенотипического признака – наличие множества иерархических уровней его регуляции. В качестве примера приведем генные сети регуляции стероидогенеза у млекопитающих (рис. 3.9). Сеть наиболее высокого иерархического уровня функционирует в гипоталамусе, регулируется транскрипционным фактором SF1 и продуцирует релизинг-факторы, контролирующие работу гипофиза. Генная сеть промежуточного уровня в гипофизе продуцирует гормоны, контролирующие гонады и надпочечники, и регулируется тем же фактором. Наконец, генная сеть самого нижнего уровня, собственно продуцирующая стероидные гормоны в гонадах и надпочечниках (кортикостерон, эстрадиол, тестостерон), также контролируется фактором SF1.

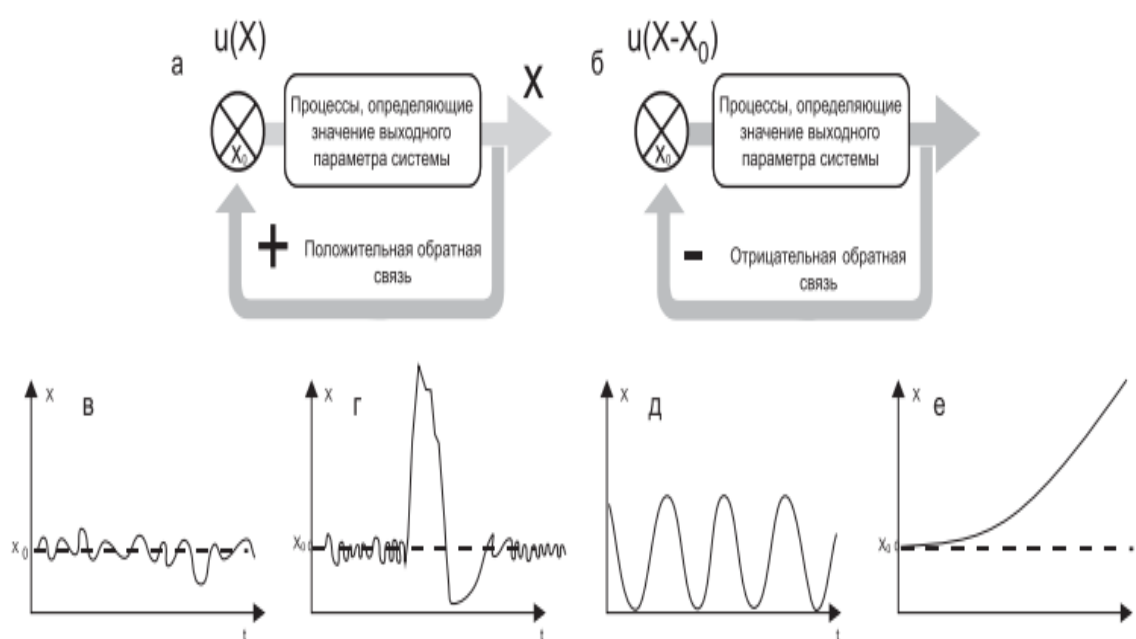


Рисунок 3.7. Регуляторные контуры с обратными связями и типы динамики критических параметров, контролируемых генными сетями. а – регуляторный контур с отрицательной обратной связью; б – регуляторный контур с положительной обратной связью; в – постоянство контролируемых параметров для генных сетей гомеостаза; г – выраженное отклонение контролируемых параметров с последующим возвращением

к норме для генных сетей стрессового ответа; д – осцилляция контролируемых параметров для генных сетей циклических процессов; е – монотонное отклонение параметров от текущего состояния для некоторых генных сетей морфогенеза.

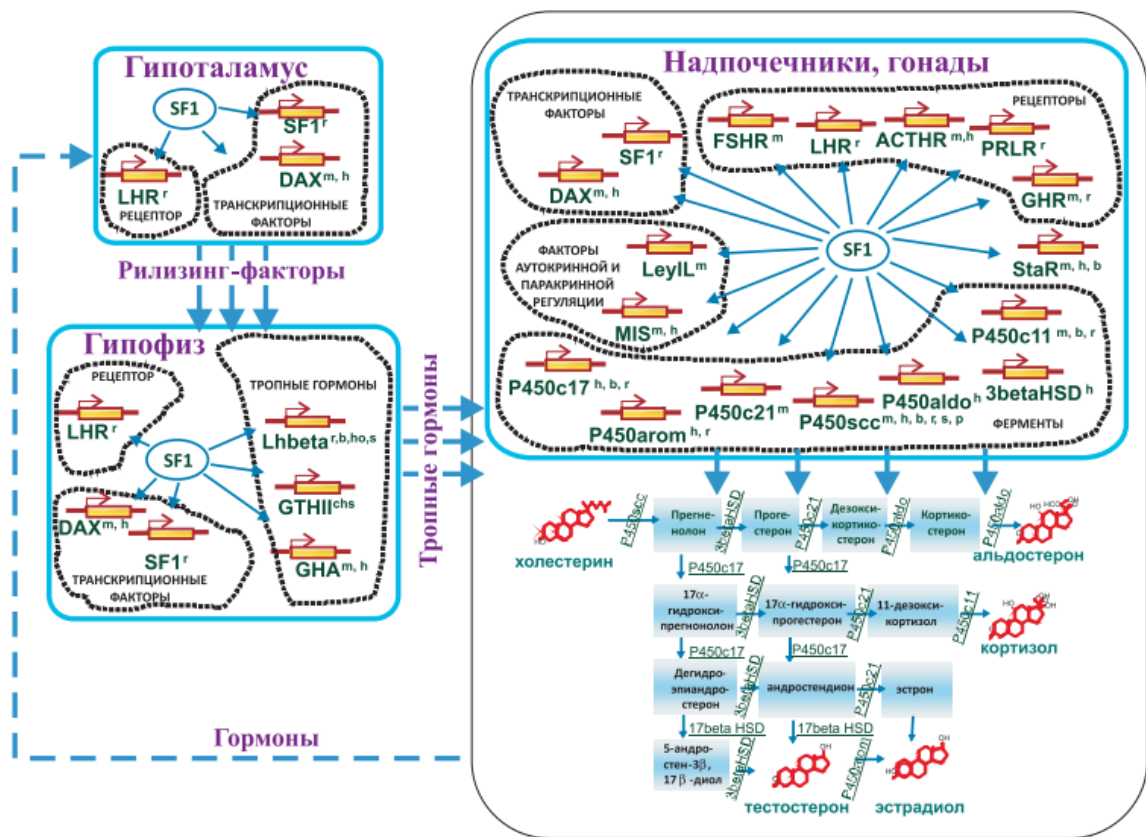


Рисунок 3.8. Генные сети регуляции стероидогенеза у млекопитающих.

Среди множества подходов к реконструкции генных сетей можно выделить следующие основные группы.

1. Ручная реконструкция генных сетей экспертами с использованием специальных программных средств – редакторов генных сетей.
2. Извлечение знаний о молекулярно-генетических взаимодействиях и семантических ассоциациях различного типа из текстов научных публикаций компьютерными методами (методы text-mining).
3. Методы автоматической интеграции гетерогенных данных о молекулярно-генетических взаимодействиях (ДНК-белковых и белок-белковых и т. д.), метаболических и сигнальных путях, генных сетях и других из различных

источников данных, основанные на использовании онтологических подходов. Заметим, что современные системы для реконструкции и анализа генных сетей, как правило, построены на сочетании нескольких подходов и используют, кроме того, различные подходы, основанные на многомерном анализе данных и компьютерных методов предсказания (Taylor, Singhal, 2009; Haibe-Kains et al., 2012; Wang et al., 2013)

Алгоритм функционирования генной сети приведен на рисунке 3.10.

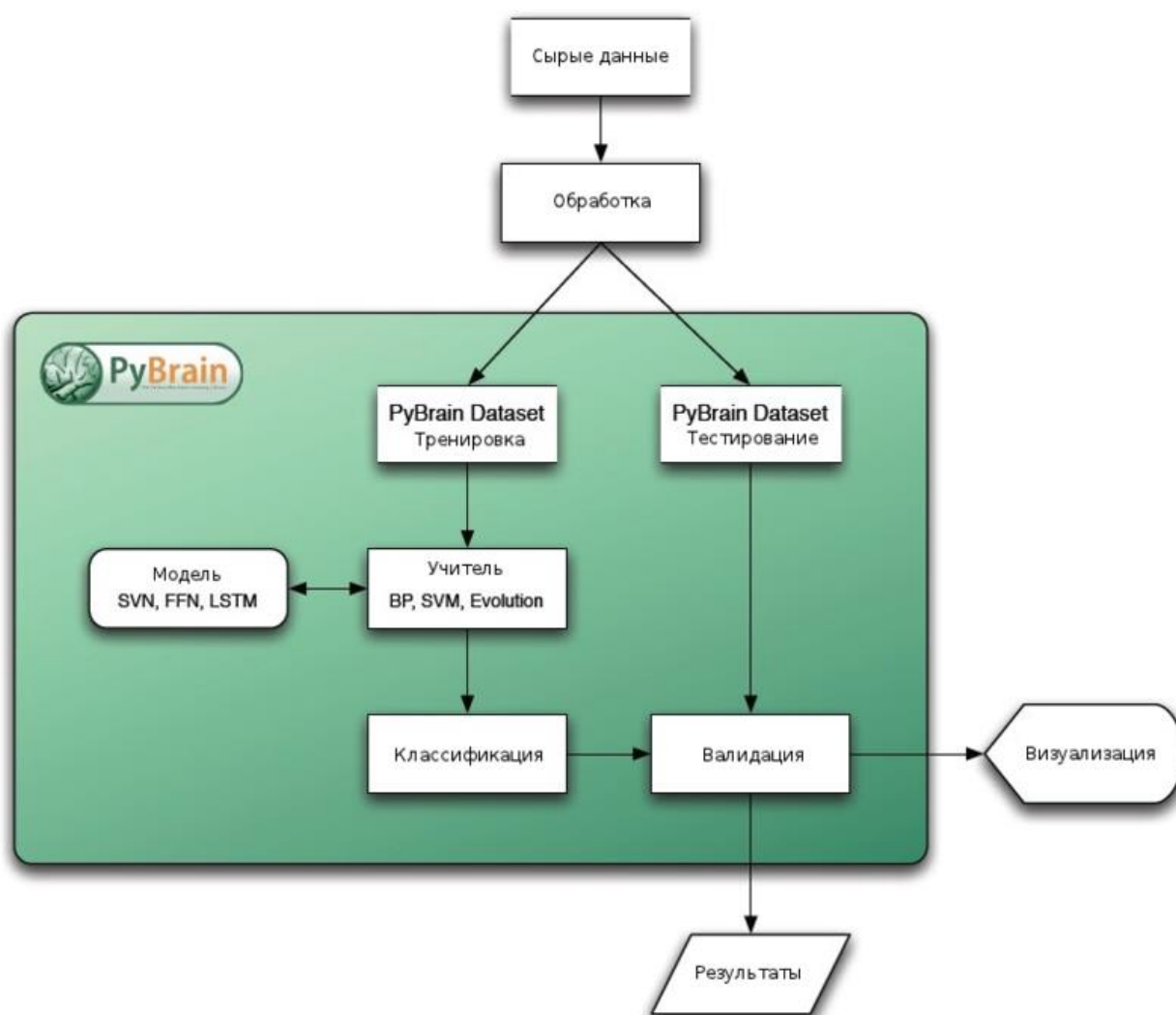


Рисунок 3.10. Алгоритм функционирования генной сети
3.4. Разработка программы для реконструкции генной сети

Код программы

```

dnacode = str(input('Введите последовательность кода ДНК: \n'))
print('Введен ДНК-код: ', dna_code)
dna_code = dna_code.lower()
  
```

```

dna_a = int(0)
dna_g = int(0)
dna_c = int(0)
dna_t = int(0)
for symbol in dna_code:
    if symbol == 'a':
        dna_a = dna_a + 1
    elif symbol == 'g':
        dna_g = dna_g + 1
    elif symbol == 'c':
        dna_c = dna_c + 1
    elif symbol == 't':
        dna_t = dna_t + 1
    print('Во введенном коде: \n Аденина: ', dna_a, '\n Гуанина: ', dna_g, '\n
Цитозина: ', dna_c, '\n Тимина: ', dna_t)

import numpy
def cal_pop_fitness(equation_inputs, pop):
    # Calculating the fitness value of each solution in the current population.
    # The fitness function calculates the sum of products between each input
and its corresponding weight.
    fitness = numpy.sum(pop*equation_inputs, axis=1)
    return fitness
def select_mating_pool(pop, fitness, num_parents):
    # Selecting the best individuals in the current generation as parents for pro-
ducing the offspring of the next generation.
    parents = numpy.empty((num_parents, pop.shape[1]))
    for parent_num in range(num_parents):
        max_fitness_idx = numpy.where(fitness == numpy.max(fitness))
        max_fitness_idx = max_fitness_idx[0][0]
        parents[parent_num, :] = pop[max_fitness_idx, :]

```

```

    fitness[max_fitness_idx] = -99999999999
    return parents
def crossover(parents, offspring_size):
    offspring = numpy.empty(offspring_size)
    # The point at which crossover takes place between two parents. Usually,
it is at the center.
    crossover_point = numpy.uint8(offspring_size[1]/2)
    for k in range(offspring_size[0]):
        # Index of the first parent to mate.
        parent1_idx = k%parents.shape[0]
        # Index of the second parent to mate.
        parent2_idx = (k+1)%parents.shape[0]
        # The new offspring will have its first half of its genes taken from the
first parent.
        offspring[k, 0:crossover_point] = parents[parent1_idx,
0:crossover_point]
        # The new offspring will have its second half of its genes taken from
the second parent.
        offspring[k, crossover_point:] = parents[parent2_idx, crossover_point:]
    return offspring
def mutation(offspring_crossover, num_mutations=1):
    mutations_counter = numpy.uint8(offspring_crossover.shape[1] /
num_mutations)
    # Mutation changes a number of genes as defined by the num_mutations
argument. The changes are random.
    for idx in range(offspring_crossover.shape[0]):
        gene_idx = mutations_counter - 1
        for mutation_num in range(num_mutations):
            # The random value to be added to the gene.
            random_value = numpy.random.uniform(-1.0, 1.0, 1)

```



```
    offspring_crossover[idx, gene_idx] = offspring_crossover[idx,  
gene_idx] + random_value  
    gene_idx = gene_idx + mutations_counter  
return offspring_crossover
```

ВЫВОДЫ ПО 3-Й ГЛАВЕ

К настоящему времени разработан специальный язык моделирования ОХКММ и программное обеспечение, позволяющее строить имитационные модели сложных многокомпарментных биологических систем, вплоть до организменного уровня.

В третьей главе выполнено сравнение методов моделирования генных сетей.

ГЛАВА 4. ОЦЕНКА ЭКОНОМИЧЕСКОЙ ЭФФЕКТИВНОСТИ ПРОГРАММЫ

4.1. Сравнение разработанной сети с существующими сетями

Генные сети — это группы координированно функционирующих генов, обеспечивающих формирование фенотипических признаков организмов (молекулярных, биохимических, физиологических, морфологических, поведенческих и т.д.) Обязательными компонентами любой генной сети являются: гены, кодируемые ими РНК, белки, метаболиты, пути передачи сигналов, метаболические пути, регуляторные контуры с положительными и отрицательными обратными связями.

Система GeneNet позволяет учитывать, что компоненты генной сети могут быть разнесены по различным органам, тканям, клеткам и клеточным компартментам. Кроме того, GeneNet позволяет описывать различные уровни организации генной сети: молекулярный, клеточный и организменный.

Большим достоинством созданной технологии является возможность автоматической визуализации генных сетей [2]. Формализованные данные обрабатываются с помощью специальной программы (GeneNet viewer) и представляются пользователю в виде графической схемы. Каждый компонент генной сети имеет свое изображение, отражающее его особенности. Например, форма изображения белка отражает степень его мультимеризации, цвет — функциональное состояние.

Анализ информации из базы данных GeneNet, а также из имеющихся литературных данных позволяет выделить несколько основных типов генных сетей. 1. Генные сети, обеспечивающие осуществление циклических процессов, например, клеточного цикла, цикла сокращения сердечной мышцы и т.д. 2. Генные сети, обеспечивающие процессы роста и дифференцировки клеток, морфогенеза тканей и органов, роста и развития организмов. 3. Генные сети, обеспечивающие гомеостаз биохимических и физиологических параметров организма. 4. Генные сети, обеспечивающие реакции организмов на изменение состояния внешней среды, например, стрессовый ответ.

Генная сеть, регулирующая клеточный цикл, представляет собой циклограмму – последовательность событий, результатом которых является возвращение генной сети в исходное состояние. В функционирование генной сети клеточного цикла вовлечено большое количество генов и их белковых продуктов. Основную роль здесь играют транскрипционные факторы семейства E2F [5–7]. На начальных этапах функционирования циклической сети преобладают регуляторные воздействия преимущественно положительного типа (рис. 3). Это обеспечивает интеграцию пролиферативных сигналов, запускающих деление клетки и репликацию ДНК (G1 и G1/S фазы клеточного цикла) [8]. На последующих этапах запускаются отрицательные регуляторы, блокирующие действие пролиферативных сигналов и приводящие к завершению программы клеточного цикла [9, 10, 7]. Подробно генная сеть, регулирующая клеточный цикл, описана в статье О.В.Кель [11].

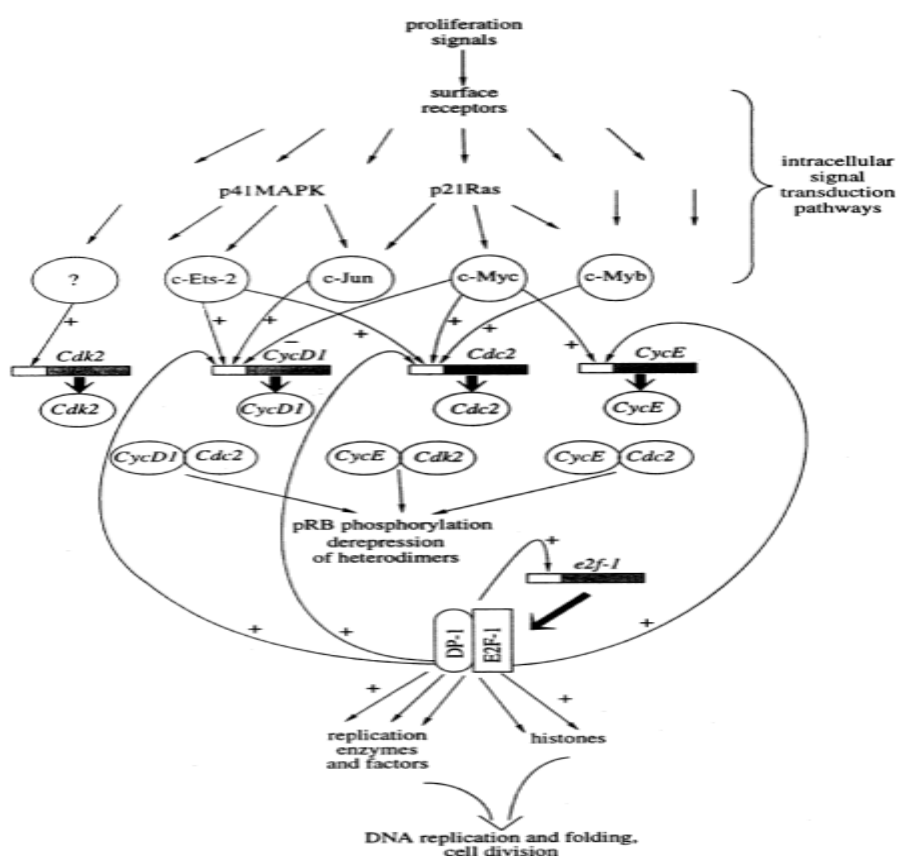


Рисунок 4.1. Интеграция пролиферативных сигналов, запускающих деление клетки и репликацию ДНК. [Цит. по: [11]].

Характерной особенностью генных сетей, контролирующих дифференцировку клеток, морфогенез тканей и органов, рост и развитие организмов, является наличие регуляторных контуров с положительными обратными связями. Известно, что функция регуляторного контура с положительной обратной связью состоит в максимально эффективном отклонении контролируемого параметра X от его текущего значения. Положительные обратные связи играют ключевую роль в регуляции процессов морфогенеза, роста и развития организмов, по своему смыслу представляющих быстрый уход от начального состояния.

Изучение организации генных сетей, описанных в базе данных GeneNet, позволяет выделить следующие принципы их организации:

- 1) наличие в каждой локальной генной сети «центральных» генов, обеспечивающих координацию функций остальных генов этой сети;
- 2) наличие в каждой генной сети регуляторных контуров типа положительной и отрицательной обратных связей, обеспечивающих ее авторегуляцию;
- 3) существование большого разнообразия молекулярных механизмов, обеспечивающих функционирование обратных связей, в том числе за счет изменения интенсивности процессов транскрипции, сплайсинга, трансляции, фосфорилирования и т.д.;
- 4) иерархическая организация локальных генных сетей и их интеграция в глобальную генную сеть организма.

В целях более полного описания генных сетей и механизмов их функционирования в настоящее время ведется развитие системы GeneNet по следующим направлениям:

- 1) разрабатываются способы динамического представления генной сети;
- 2) совершенствуется концепция иерархического описания генных сетей на уровне различных компартментов организма.

В каждой генной сети можно выделить несколько обязательных типов структурных и функциональных компонентов: 1) группа координированно экспрессирующихся генов, составляющая ядро сети; 2) белки, кодируемые этими генами; 3) низкомолекулярные компоненты (гормоны и другие сигнальные молекулы, энергетические компоненты, метаболиты); 4) отрицательные и положительные обратные связи, стабилизирующие параметры генной сети на определенном уровне или, напротив, отклоняющие их от исходного значения, обеспечивая переход к новому функциональному состоянию. Генные сети могут включать такие крупные функциональные модули, как: а) метаболические пути; б) пути передачи сигналов от клеточных мембран в ядра клеток, обеспечивающие активацию или подавление транскрипции генов в ответ на внешние по отношению к клетке стимулы.

Качественный анализ работы генных сетей позволяет выделить несколько их классов.

1) Генные сети, обеспечивающие гомеостаз организма, обычно содержат контуры отрицательной обратной связи (ОС), обеспечивающие возвращение контролируемых (критических) переменных к норме. Пример: Генные сети регуляции уровня холестерина в клетке. Если уровень слишком высок, то активация его предшественника (preSREBP) замедляется самими избыточными молекулами холестерина. Уровень внутриклеточного холестерина постепенно возвращается к норме.

2) Генные сети, обеспечивающие адаптивный ответ организма на резкие изменения внешних условий. Подобные Генные сети часто имеют двухэтапную схему работы: сначала происходит быстрая активация процессов за счет положительной ОС, затем включается отрицательная ОС, которая подавляет ответ. В результате контролируемые переменные Генные сети (концентрации какихлибо метаболитов) поддерживаются на определенном уровне. Примеры: Генные сети противовирусного ответа и Генные сети теплового шока. Так, при действии теплового шока усиливается экспрессия многих генов по принципу положительной ОС. Ключевой транскрипционный

фактор ответа на тепловой шок HSF1 (heat shock factor 1) в неактивном состоянии находится в цитоплазме в комплексе с молекулярными шаперонами, например, HSP90 (heat shock protein 90). При действии теплового шока появляются частично денатурированные, неправильно свернутые белки. Комплекс распадается. HSP90 связывается с белками и способствует их правильному свертыванию. Фактор HSF1 активируется, затем транспортируется в ядро, где связывается с HSE-сайтами и активирует транскрипцию генов теплового шока, являющимися в большинстве молекулярными шаперонами. При избытке шаперонов, что регулируется 15 концентрацией в клетке неправильно свернутых белков, HSP70 связывается с HSF1, который транспортируется в цитоплазму. То есть срабатывает отрицательная ОС и ответ на тепловой шок завершается.

3) Генные сети, обеспечивающие циклические процессы. Пример: клеточный цикл. В подобных генных сетях наблюдается четкий баланс между положительными и отрицательными обратными связями.

4) Генные сети необратимых процессов, контролирующие рост и развитие организмов, дифференцировку клеток, морфогенез органов и отмирание тканей. Они содержат контуры положительной ОС, обеспечивающие эффективное отклонение основной переменной от исходного значения. Примеры: Генные сети дифференцировки эритроцитов, Генные сети, регулирующая прорастание семян и Генные сети апоптоза. Последний пример будет разобран ниже для иллюстрации каскадного принципа усиления сигнала.

5) Генные сети, обеспечивающие передачу наследственной информации и воспроизведение организма.

1) в неактивном состоянии находится в цитоплазме в комплексе с молекулярными шаперонами, например, HSP90 (heat shock protein 90). При действии теплового шока появляются частично денатурированные, неправильно свернутые белки. Комплекс распадается. HSP90 связывается с белками и способствует их правильному свертыванию. Фактор HSF1 активируется, затем транспортируется в ядро, где связывается с HSE-сайтами и активирует

транскрипцию генов теплового шока, являющимися в большинстве молекулярными шаперонами. При избытке шаперонов, что регулируется 15 концентрацией в клетке неправильно свернутых белков, HSP70 связывается с HSF1, который транспортируется в цитоплазму. То есть срабатывает отрицательная ОС и ответ на тепловой шок завершается. 3) Генные сети, обеспечивающие циклические процессы. Пример: клеточный цикл. В подобных генных сетях наблюдается четкий баланс между положительными и отрицательными обратными связями. 4) Генные сети необратимых процессов, контролирующие рост и развитие организмов, дифференцировку клеток, морфогенез органов и отмирание тканей. Они содержат контуры положительной ОС, обеспечивающие эффективное отклонение основной переменной от исходного значения. Примеры: Генные сети дифференцировки эритроцитов, генные сети, регулирующие прорастание семян и Генные сети апоптоза. Последний пример будет разобран ниже для иллюстрации каскадного принципа усиления сигнала.

5) Генные сети, обеспечивающие передачу наследственной информации и воспроизведение организма.

Для построения модели любой ГС используют информацию об ее элементах и связях между ними. Это означает, что необходимо знать какие гены участвуют в формировании сети, их регуляторные элементы (промоторы и др.), синтезируемые ими белки, пути передачи сигналов в клетке. Привлечение этой информации идет путем обобщения литературных данных и получения новых данных в ходе натурных экспериментов с конкретными генными сетями. Составляют уравнения, характеризующие динамику генных сетей, подбирают значения параметров – констант, входящих в эти уравнения.

Генные сети в упрощенных вариантах моделей описывают на языке генетических элементов. Каждый из них содержит ген, соответствующие ему мРНК и белки. Функционирование генетического элемента в простейшем случае – процесс синтеза белка. В генетический элемент неявно включаются все процессы, обеспечивающие синтез мРНК, белка, формирование мультимера белка (мономерная форма белка есть частный случай мультимерной).

Белки обеспечивают регулирующие взаимодействия между генетических элементов.

Адекватные математические модели генных сетей и их компьютерные реализации являются незаменимым инструментом для анализа функционирования сложных биологических систем.

С помощью моделей генных сетей возможно:

- предсказывать динамику тех или иных процессов в генных сетях;
- исследовать механизмы их протекания;
- выявлять ключевые звенья в генных сетях;
- анализировать влияние различных воздействий на живые системы;
- подбирать оптимальные воздействия;
- изучать эволюцию ГС.

4.2. Определение затрат на разработку программы

Затраты на разработку программы включают в себя:

затраты на оплату труда;

отчисления на социальное страхование;

затраты на электроэнергию;

иные затраты.

Затраты на оплату труда;

Общая сумма затрат на оплату труда ($Z_{тр}$) рассчитывается по следующей формуле:

$$Z_{тр} = \sum_{i=1}^n ЧС_i \times T_i \quad (4.1)$$

$ЧС_i$ - часовая ставка i -го работника (руб./ч.);

T_i - время разработки проекта (ч);

i - категория работника;

n - количество работников, занятых разработкой проекта.

Общее время работы разработчика программы $T_i = 300$ ч.

Среднечасовая заработная плата разработчика рассчитывается по формуле:

$$ЧС_i = \frac{ЗП_i}{ФРВ_i} \quad (4.2)$$

ЗП_і - среднемесячная заработная плата разработчика проекта.

ФРВ_і - среднемесячный фонд рабочего времени.

Среднемесячная заработная плата начинающего программиста составляет 15 000 рублей.

Среднемесячный фонд рабочего времени берется = 100 часов.

Таким образом, стоимость одного часа работы программиста, равна:

$$ЧС = 15\,000 \text{руб.} / 100 \text{ч.} = 150 \text{руб.}$$

Затраты на оплату труда каждого работника приведены в таблице 21.

Таблица 4.1. Затраты на оплату труда

Категория работника	Квалификация работника	Время разработки программы (час)	Часовая ставка (руб/ч.)	Сумма (руб.)
Разработчик программы	Программист	200	150.0	30 000
Итого	30 000			

Общая сумма затрат на оплату труда, рассчитанная по формуле, равна: $З_{тр} = 30\,000$

Затраты на электроэнергию

Таблица 4.2. Затраты на электроэнергию.

Наименование оборудования	Паспортная мощность, кВт	Количество, шт	Время работы оборудования на разработку программы, ч	Цена электроэнергии, руб./кВт*ч	Сумма, руб.
Компьютер	0,6	1	200	2,45	294
Принтер	0,3	1	1	2,45	0,735
Итого:	294,74				

Затраты на разработку программы составили 30 294,74 рубля.

4.3. Оценка сроков окупаемости программы

Расчет срока окупаемости программы показывает величину годового прироста прибыли, образующуюся в результате производства или эксплуатации нового программного продукта.

Срок окупаемости капитальных вложений в новом варианте, по сравнению с прежним составит:

$$t = KB / ДЕ \quad (2.18)$$

Экономия эксплуатационных расходов ДЕ равна:

$$ДЕ = (1 + w_c) \cdot (1 + w_o) \cdot Z_i \cdot 12 - E \quad (2.19)$$

где Z_i - среднемесячная зарплата исполнителя:

$$Z_i = 30000,66 \text{ руб.}$$

12 - число месяцев в году;

w_c - коэффициент, учитывающий ЕСН (0,34);

w_o - коэффициент, учитывающий доплату за гос. отпуск (0,12).

$$ДЕ = (1 + 0,34) \cdot (1 + 0,12) \cdot 30000 \cdot 12 - 1722,65 = 538\,565,35 \text{ тогда}$$

$$t = 46208,85 / 538\,565,35 = 0,085 \text{ года,}$$

$t \approx 1$ мес.

Значение расчетного коэффициента эффективности капитальных вложений сопоставления с нормативным значением коэффициента эффективности капитальных вложений вычислительной техники, установленным решением директивных органов.

В случае, когда расчетный коэффициент равен нормативному или ниже него, произвольно и внедрение программного продукта является эффективным.

1 мес. < 3 лет, значит, применение разработанного программного продукта является экономически эффективным.

Расчет денежного годового эффекта

Денежный годовой эффект, полученный одним потребителем новой программы:

$$\mathcal{E} = DE - E_n \cdot KB, (2.20)$$

где E_n - 0,15 – нормативный коэффициент экономической эффективности капитальных вложений:

$$\mathcal{E} = 538\,565,35 - 0,15 \cdot 46208,85 = 533020,288 \text{ руб.}$$

ВЫВОДЫ ПО 4-Й ГЛАВЕ

Опыт разработки подобных баз данных указывает на необходимость создания такой универсальной компьютерной технологии, которая позволяет описывать любые элементарные структуры, события и процессы, значимые для функционирования генных сетей.

При этом существенными требованиями являются: 1) возможность описания в рамках единого подхода как механизмов регуляции экспрессии генов, так и различных метаболических процессов; 2) возможность описания генных сетей как про-, так и эукариот, а также их совместного функционирования; 3) наличие понятной и удобной в работе системы для описания генных сетей и ввода информации в базу; 4) возможность автоматической визуализации структуры генных сетей на основе накопленных данных. Нерешенность этих задач тормозит развитие подходов к изучению генных сетей, по-

сколько любые исследования подобного рода должны основываться на анализе большой совокупности разнородных экспериментальных данных.

Предложенная компьютерная технология GeneNet [2, 3] является, по-видимому, одной из первых попыток успешного решения задачи, которая открывает путь к быстрому накоплению в компьютерных базах данных экспериментальной информации о структурно-функциональной организации генных сетей и возможности для систематизации и анализа этой информации

ГЛАВА 5. ТЕСТИРОВАНИЕ ПРОГРАММЫ

5.1. Формирование набора тестов

Накопление генетических данных в области изучения болезни Паркинсона в настоящее время сильно прогрессирует от определения факторов риска до уверенного прогнозирования возникновения заболевания. Для поиска новых мишеней терапии необходима реконструкция генной сети заболевания, кластеризация генов в сети, выявление ключевых генов, обладающих наибольшим числом контактов в сети. С помощью онлайн-инструментов биоинформатики OMIM, PANTHER, g:Profiler, GeneMANIA и STRING-DB мы проанализировали актуальный на данный момент массив данных, связанных с болезнью Паркинсона, рассчитали категории генных онтологий для большого списка генов, визуализировали их и построили генные сети, содержащие выявленные ключевые объекты и их взаимосвязи. Биологическая интерпретация полученных результатов всё еще остается сложной задачей. Анализ генов, связанных с болезнью Паркинсона, определение их положения в генной сети (связанности) позволяет оценить их перспективность в качестве генов-мишеней для лекарственных воздействий.

Приведённые в таблице 5.1 данные показывают, что наиболее значимыми для генов болезни Паркинсона являются категории общей регуляции клеточной смерти, регуляции клеточной смерти нейронов, регуляции апоптоза и программируемой клеточной смерти, негативная регуляция клеточной смерти. Эти данные подтверждают ключевые этиологические признаки заболевания, среди которых центральным аспектом патофизиологии болезни Паркинсона является прогрессирующая гибель дофаминовых нейронов среднего мозга и их аксональных проекций. Процессы аутофагии являются одним из основных путей внутриклеточной деградации α -синуклеина и актуальные исследования показывают, что дисфункциональная аутофагия при болезни Паркинсона является одним из основных факторов риска развития заболевания [5].

Таблица 5.1. Категории генных онтологий для генов болезни Паркинсона по PANTHER
(биологические процессы), $p=1,0E-20$

Категории генных онтологий для биологических процессов	Класс онтологий GO	Число генов	FC*	<i>p</i> (Бонферрони)
регуляция гибели клетки (regulation of cell death)	GO:0010941	107	4,85	1,59E-40
позитивная регуляция клеточного метаболизма (positive regulation of cellular metabolic process)	GO:0031325	140	3,24	5,10E-37
позитивная регуляция метаболических процессов (positive regulation of metabolic process)	GO:0009893	149	3,00	1,36E-36
регуляция гибели нейронов (regulation of neuron death)	GO:1901214	52	12,52	1,20E-34
регуляция метаболизма белков (regulation of protein metabolic process)	GO:0051246	121	3,49	3,12E-33
регуляция клеточного белкового обмена (regulation of cellular protein metabolic process)	GO:0032268	116	3,56	3,37E-32
регуляция программируемой гибели клеток (regulation of programmed cell death)	GO:0043067	93	4,59	3,63E-32
позитивная регуляция метаболизма азотистых соединений (positive regulation of nitrogen compound metabolic process)	GO:0051173	129	3,16	7,67E-32
регуляция транспортировки (regulation of transport)	GO:0051049	99	4,22	8,09E-32
регуляция процессов апоптоза (regulation of apoptotic process)	GO:0042981	91	4,61	2,03E-31
регуляция организации клеточных компонентов (regulation of cellular component organization)	GO:0051128	110	3,49	3,45E-29
регуляция клеточных катаболических процессов (regulation of cellular catabolic process)	GO:0031329	69	6,04	5,84E-29
регуляция локализации (regulation of localization)	GO:0032879	117	3,20	2,53E-28
позитивная регуляция процессов белкового обмена (positive regulation of protein metabolic process)	GO:0051247	88	4,29	6,94E-28
негативная регуляция гибели клеток (negative regulation of cell death)	GO:0060548	72	5,40	1,48E-27
позитивная регуляция процессов клеточного белкового обмена (positive regulation of cellular protein metabolic process)	GO:0032270	85	4,37	2,73E-27
регуляция катаболических процессов (regulation of catabolic process)	GO:0009894	72	5,28	5,73E-27
регуляция процессов модификации белков (regulation of protein modification process)	GO:0031399	89	4,08	8,85E-27
поведение (behavior)	GO:0007610	55	7,52	1.80E-26

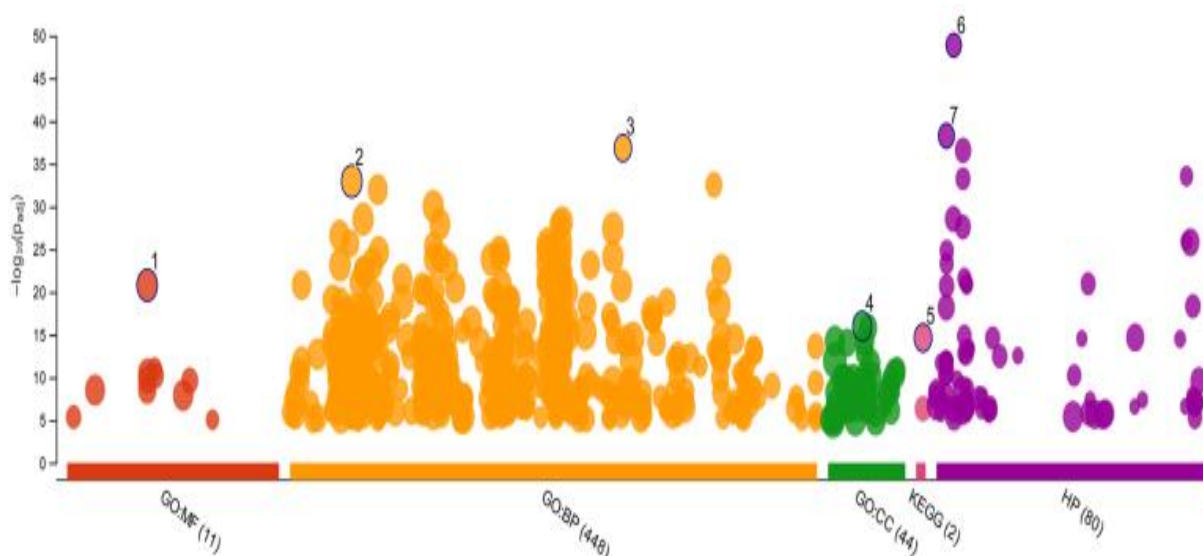


Рисунок 5.1. График поточечных значений значимости категорий генных онтологий генов болезни Паркинсона, рассчитанный с помощью программы g:Profiler. По оси Y — значения p-value в логарифмической шкале, по оси X — термины генных онтологий (по группам, выделены цветом). Цветной вариант рисунка доступен в электронной версии журнала.

Генетические тесты: поисковые и подтверждающие

Тест	Тип	Цена
Анализ повтора в гене <i>C9orf72</i>	Подтверждающий	~5 т.р.
Поиск мутаций в гене <i>SOD1</i>	Комбинированный	~5 т.р.
Панели «БАС», «Нейродегенеративные заболевания» и т.п. (десятки или сотни генов)	Поисковый	~35 т.р.
Секвенирование экзона (~20 000 генов)	Поисковый	~50 т.р.
Секвенирование генома	Поисковый	~100 т.р.
Подтверждение одной мутации по Сэнгеру		Подтверждающий
		~5 т.р.

Результаты «поисковых» тестов на основе NGS требуют подтверждения по Сэнгеру и часто бывают сложны в интерпретации

Рисунок 5.2. Тесты генов

Болезнь Паркинсона: генетика

Моногенные формы болезни Паркинсона

Локус	Хромосома	Белок	Мутации	Тип наследования	Авторы
PARK1	4q21-22	Альфа-синуклеин	A53T, A30P, мультимпликация	Аутосомно-доминантный	Италия, Греция, Германия
PARK2 Ювенил.	6q25.2-27	Паркин	Делеция, точковые мутации, изменение порядка чередования нуклеотидов	Аутосомно-рецессивный	Повсеместно
PARK3	2p13			Аутосомно-доминантный	Европа
PARK4	4p15			Аутосомно-доминантный	Семьи, повсеместно
PARK5	4p14-15	UCH-L1			Германия
PARK6	1p35-36	PINK1	G309D, E414G, S414S	Аутосомно-рецессивный	Повсеместно
PARK7	1p36	DJ	Ala167Ala (c501A>G, rs71653621)	Аутосомно-рецессивный	18,19
PARK8	12p12	LRRK2	G2019S	Аутосомно-доминантный	14,15,16

Н.А. Шнайдер, М.Р. Сапронова, 2012 с дополнениями

Рисунок 5.3. Тесты генов



Рисунок 5.4. Технология тестирования

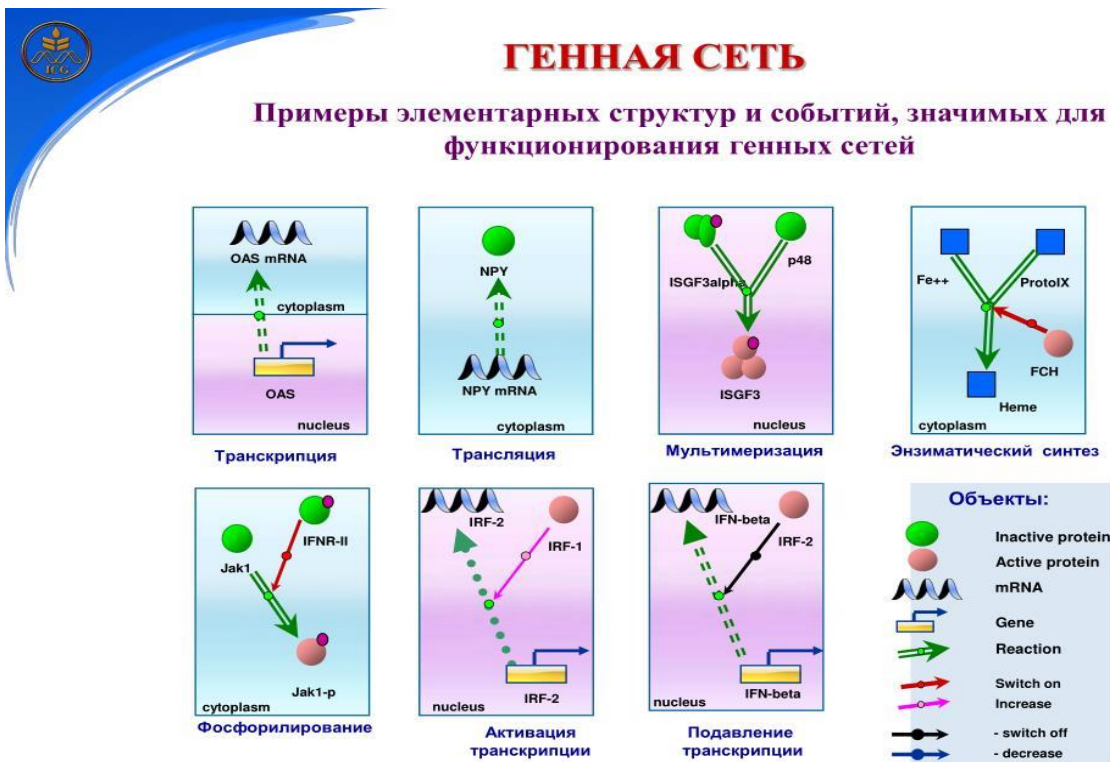


Рисунок 5.5. Элементы генной сети

5.2. Выполнение тестов в программе

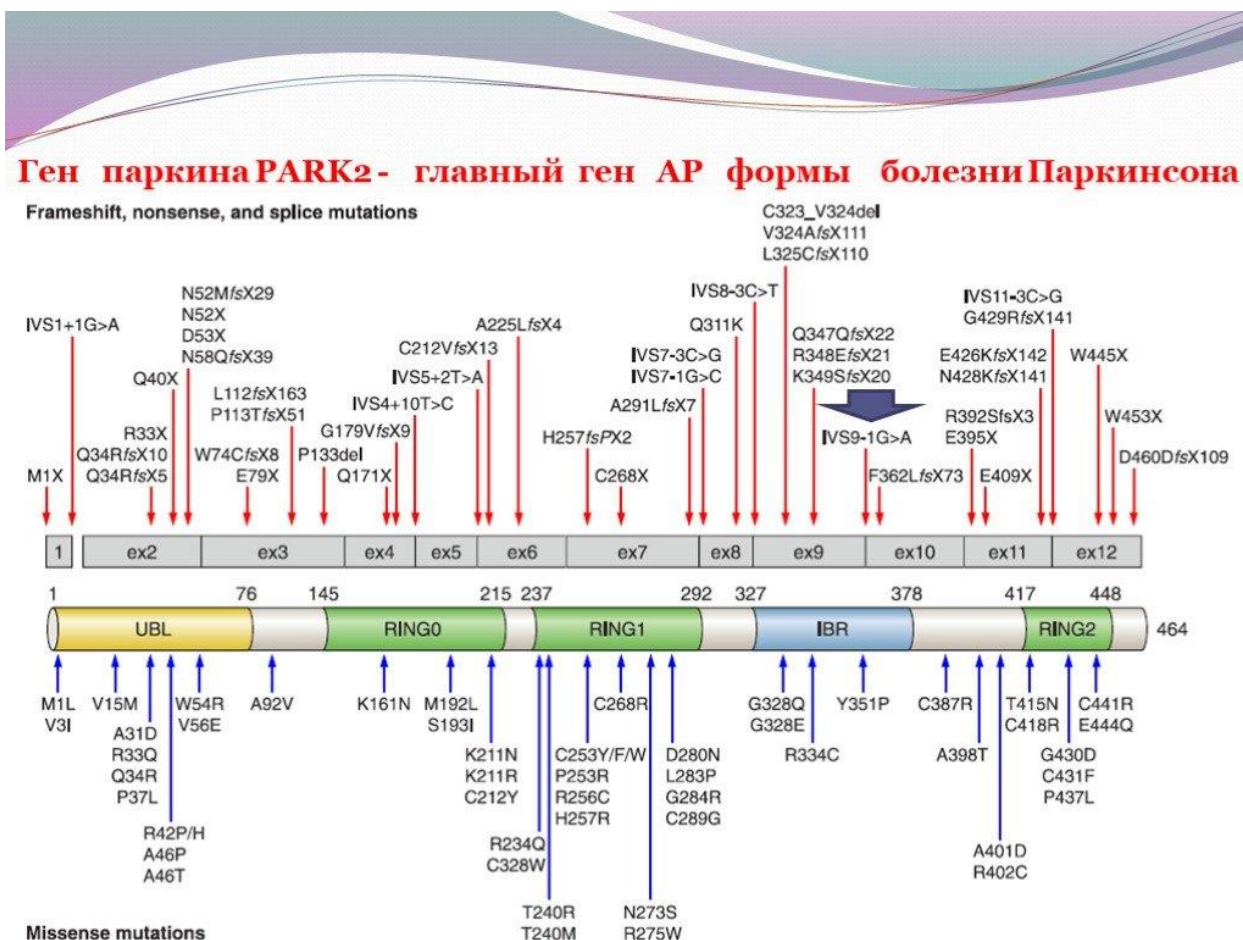


Рисунок 5.6. Оценка параметров мутации генов

ГЕН ЛОКАЛИЗОВАН	ВАРИАНТ	ТИП НАСЛЕДОВА НИЯ	НОМЕР ОМIM	МУТАЦИЯ ГЕНА
1p36.23	Болезнь Паркинсона 7, аутосомно-рецессивный раннее начало	A-P	606324	DJ1
1p36.13	Синдром Kufor-Rakeb	A-P	606693	ATP13A2
1p36.12	болезнь Паркинсона 6, раннее начало	A-P	605909	PINK1
1p32	Болезнь Паркинсона 10		606852	PARK10
1p31.3	Parkinson 19b болезнь, раннее начало	A-P	615528	DNAJC6
1p31.3	Паркинсона болезнь 19a, ювенильный	A-P	615528	DNAJC6
1q22	Болезнь Паркинсона, позднее начало	Многофакторный	168600	GBA
1q32	Болезнь Паркинсона 16		613164	PARK16
2p13	Болезнь Паркинсона 3		602404	PARK3
2p13.1	Болезнь Паркинсона 13		610297	HtrA2
2q37.1	Болезнь Паркинсона 11		607688	GIGYF2
3q22	Болезнь Паркинсона 21	A-Д	616361	PARK21
3q27.1	Болезнь Паркинсона 18	A-Д	614251	EIF4G1
4p13	Болезнь Паркинсона 5		613643	UCHL1
4q22.1	Болезнь Паркинсона 1	A-Д	168601	SNCA
4q22.1	Болезнь Паркинсона 4	A-Д	605543	SNCA
4q23	Болезнь Паркинсона,	Многофакторный	168600	ADH1C
6q26	болезнь Паркинсона, юношеский, тип 2	A-P	600116	PRKN

Рисунок 5.7. Локализация генов при болезни Паркинсона

ГЕН	ВАРИАНТ	ТИП НАСЛЕДОВА НИЯ	НОМЕР ОМIM	МУТАЦИЯ ГЕНА
7p11.2	Болезнь Паркинсона 22, аутосомно-доминантное		616710	CHCHD2
12q12	Болезнь Паркинсона 8	A-Д	607060	LRRK2
12q24.12	Болезнь Паркинсона, позднее начало, восприимчивость к	Многофакторный	168600	ATXN2
15q22.2	Болезнь Паркинсона 23, аутосомно-рецессивный, раннее начало	A-P	616840	VPS13C
16q11.2	Болезнь Паркинсона 17	A-Д	614203	VPS35
17q21.31	Болезнь Паркинсона	Многофакторный	168600	MAPT
21q22.11	Болезнь Паркинсона 20, раннее начало	A-P	615530	SYNJ1
22q12.3	Болезнь Паркинсона 15, аутосомно-рецессивный	A-P	260300	FBXO7
22q13.1	Болезнь Паркинсона 14, аутосомно-рецессивное	A-P	612953	PLA2G6
Xq21-q25	Болезнь Паркинсона 12		300557	PARK12
Xq24	Болезнь Паркинсона, возраст начала, модификатор	Многофакторный	168600	GLUD2

Рисунок 5.8. Локализация генов при болезни Паркинсона

Широко распространено предположение, что белки, кодируемые рецессивными генами, ассоциированными с ранним началом болезни Паркинсона (PINK1, PARK2, DJ-1 и FBXO7) участвуют в системе контроля качества митохондрий и её регуляции [17]; считается, что митохондриальная дисфункция играет ключевую роль в ранних событиях начала болезни Паркинсона.

5.3. Оценка быстродействия программы и точности результатов

На рисунке 5.9. показано, что по мере роста генов в моделируемой сети растёт время обработки результатов в программе, но на современных компьютерах и при использовании языка программирования Python оценка параметров генов в программе выполняется достаточно быстро.

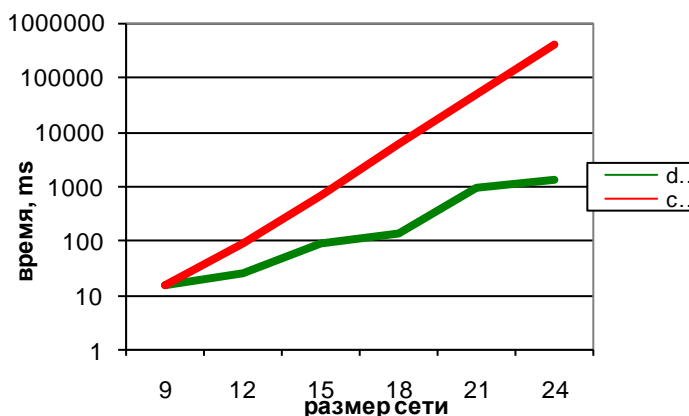


Рисунок 5.9. Оценка времени работы программы

Точность результатов составляет 93-95%.

ВЫВОДЫ ПО 5-Й ГЛАВЕ

Реконструкция генных сетей, основанная на анализе актуальных данных генов, ассоциированных с болезнью Паркинсона, приводит к выявлению сетевых структур, функционально связанных с генами риска заболевания. Обнаружение функциональных связей открывает путь к созданию новых лекарственных препаратов с точным направленным действием и минимумом побочных эффектов, выявляет новые стратегические пути к терапевтической стимуляции гаснущих нейронных процессов.

ЗАКЛЮЧЕНИЕ

Определение ключевых генов болезни Паркинсона, полученных с помощью анализа структуры геной сети, необходимо для дальнейшего подбора и изучения потенциальных вариантов лечения заболевания, выбора лекарств, способных взаимодействовать с белками, что в перспективе приведёт к созданию лекарственных средств на их основе.

Решение указанных задач необходимо начинать с выбора и построения актуального списка генов, ассоциированных с развитием болезни Паркинсона, с анализа категорий геной онтологий и реконструкции геной сети.

Этиология генетически сложной болезни Паркинсона представляет огромный набор возможных взаимодействий генетических рисков и окружающих факторов. Составление списка ключевых генов, ассоциированных с развитием болезни Паркинсона, проведенный анализ категории геной онтологий для этого списка и реконструкция геной сети позволили выбрать гены-мишени для лекарственных воздействий, модифицирующих степень выраженности болезни.

Существующие методы терапии позволяют продлить время жизни пациента и улучшить её качество, но пока не направлены непосредственно на причину заболевания, а носят скорее заместительный характер, что говорит о необходимости дальнейшего исследования генетических основ болезни. Появились работы, показывающие, что геной терапия становится перспективным подходом для лечения неврологических расстройств, включая болезнь Паркинсона.

Накопление генетических данных в области изучения болезни Паркинсона в настоящее время сильно прогрессирует, продолжается рост числа публикаций по данной теме от определения факторов риска до уверенного прогнозирования возникновения самого заболевания.

Компьютерные методы применяются не только для функционального анализа генов, их классификации и установления взаимосвязей, но и широко

используются при экспериментальном изучении структуры и взаимодействия генов и регуляции их экспрессии в клетках мозга.

Разработанная программа позволяет оценить параметры болезни Паркинсона и выполнить прогноз развития болезни, определить этапы ее лечения.

Список литературы

1. Колчанов Н.А. Регуляция транскрипции генов эукариот: базы данных и компьютерный анализ // Mol. Biol. 1997. Т. 31. С. 581–583.
2. Kolpakov F.A., Ananko E.A., Kolesov G.B., Kolchanov N.A. GeneNet: a database for gene networks and its automated visualization // Bioinformatics. 1998. V. 14. P. 529– 537.
3. Kolpakov F.A., Ananko E.A. Interactive data input into the GeneNet database // Bioinformatics. 1999. V. 15. P. 713–714.
4. Ananko E.A., Kolpakov F.A., Kolesov G.B., Kolchanov N.A. Gene networks: a database and its automated visualization through the Internet in the GeneNet computing system // Proc. I Intern. conf. on bioinformatics of genome regulation and structure. 1998. Novosibirsk, Russia. P. 82–85.
5. Farnham P.J., Slansky J.E., Kollmar R. The role of E2F in the mammalian cell cycle // Biochim. Biophys. Acta. 1993. V. 1155. P. 125–131.
6. DeGregori J., Kowalik T., Nevins J.R. Cellular targets for activation by the E2F1 transcription factor include DNA synthesis- and G1/S-regulatory genes // Mol. Cell. Biol. 1995. V. 15. P. 4215–4224.
7. O'Connor D.J., Lam E.W.-F., Griffin S. et al. Physical and functional interactions between p53 and cell cycle co-operating transcription factors, E2F1 and DP1 // EMBO J. 1995. V. 14. P. 6184–6192.
8. Johnson D.G., Ohtani K., Nevins J.R. Autoregulatory control of E2F1 expression in response to positive and negative regulators of cell cycle progression // Genes Dev. 1994. V. 8. P. 1514–1525.
9. Di Leonardo A., Linke S.P., Clarkin K., Wahl G.M. DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts // Genes Dev. 1994. V. 8. P. 2540–2551.
10. Agarwal M.L., Agarwal A., Taylor W.R., Stark G.R. p53 controls both the G2/M and the G1 cell cycle checkpoints and mediates reversible growth arrest in human fibroblasts // Proc. Natl. Acad. Sci. USA. 1995. V. 92. P. 8493–8497.

11. Кель О.В., Кель А.Э. Межгенные взаимодействия в регуляции клеточного цикла: ключевая роль транскрипционных факторов семейства E2F // Мол. биология. 1997. Т. 31, № 4. С. 656–670.
12. Matthews D.J., Topping R.S., Cass R.T., Giebel L.B. A sequential dimerization mechanism for erythropoietin receptor activation // Proc. Natl. Acad. Sci. USA. 1996. V. 93. P. 9471–9476.
13. Elliott S., Lorenzini T., Yanagihara D. et al. Activation of the erythropoietin (EPO) receptor by bivalent anti-EPO receptor antibodies // J. Biol. Chem. 1996. V. 271. P. 24691–24697.
14. Penta K., Sawyer S.T. Erythropoietin induces the tyrosine phosphorylation, nuclear translocation, and DNA binding of STAT1 and STAT5 in erythroid cells // J. Biol. Chem. 1995. V. 270. P. 31282–31287.
15. Briegel K., Bartunek P., Stengl G. et al. Regulation and function of transcription factor GATA-1 during red blood cell differentiation // Development. 1996. V. 122. P. 3839–3850.
16. Boyes J., Byfield P., Nakatani Y., Ogryzko V. Regulation of activity of the transcription factor GATA-1 by acetylation // Nature. 1998. V. 396. P. 594–598.
17. Zhang W., Bieker J.J. Acetylation and modulation of erythroid Kruppel-like factor (EKLF) activity by interaction with histone acetyltransferases // Proc. Natl. Acad. Sci. USA. 1998. V. 95. P. 9855–9860.
18. Prasad K.S., Jordan J.E., Koury M.J. et al. Erythropoietin stimulates transcription of the TAL1/SCL gene and phosphorylation of its protein products // J. Biol. Chem. 1995. V. 270. P. 11603–11611.
19. Dalyot N., Fibach E., Ronchi A. et al. Erythropoietin triggers a burst of GATA-1 in normal human erythroid cells differentiating in tissue culture // Nucl. Acids Res. 1993. V. 21. P. 4031–4037.
20. Hannon R., Evans T., Felsenfeld G., Gould H. Structure and promoter activity of the gene for the erythroid transcription factor GATA-1 // Proc. Natl. Acad. Sci. USA. 1991. V. 88. P. 3004–3008.

21. Tsai S.-F., Strauss E., Orkin S.H. Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter // *Genes Dev.* 1991. V. 5. P. 919–931.
22. Chin K., Oda N., Shen K., Noguchi C.T. Regulation of transcription of the human erythropoietin receptor gene by proteins binding to GATA-1 and Sp1 motifs // *Nucl. Acids Res.* 1995. V. 23. P. 3041–3049.
23. Crossley M., Tsang A.P., Bieker J.J., Orkin S.H. Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1 // *J. Biol. Chem.* 1994. V. 269. P. 15440–15444.
24. Lecointe N., Bernard O., Naert K. et al. GATA-and SP1-binding sites are required for the full activity of the tissue-specific promoter of the tal-1 gene // *Oncogene.* 1994. V. 9. P. 2623–2632.
25. Royer-Pokora B., Rogers M., Zhu T.H. et al. The TTG-2/RBTN2 T cell oncogene encodes two alternative transcripts from two promoters: the distal promoter is removed by most t(11p13) translocations in acute T cell leukaemia's (T-ALL) // *Oncogene.* 1995. V. 10. P. 1353–1360.
26. Vieille-Grosjean I., Huber P. Transcription factor GATA-1 regulates human HOXB2 gene expression in erythroid cells // *J. Biol. Chem.* 1995. V. 270. P. 4544–4550.
27. Mikkonen A., Porali I., Cercos M., Ho T.D. A major cysteine proteinase, EPB, in germinating barley seeds: structure of two intronless genes and regulation of expression // *Plant Mol. Biol.* 1996. V. 31. P. 239–254.
28. Gubler F., Kalla R., Roberts J.K., Jacobsen J.V. Gibberellin-regulated expression of a myb gene in barley aleurone cells: evidence for Myb transactivation of a high-pI alphaamylase gene promoter // *Plant Cell.* 1995. V. 7. P. 1879–1891.
29. Rogers J.C., Rogers S.W. Definition and functional implications of gibberellin and abscisic acid cis-acting hormone response complexes // *Plant Cell.* 1992. V. 4. P. 1443–1451.

30. Jackson S.M., Ericsson J., Osborne T.F., Edwards P.A. MF-Y has a novel role in sterol-dependent transcription of two cholesterol genes // *J. of Biol. Chemistry*. 1995. V. 270. P. 21445–21448.
31. Lloyd D.B., Thompson J.F. Transcriptional modulators affect in vivo protein binding to the low density lipoprotein receptor and 3-hydroxy-3-methylglutaryl coenzyme A reductase promoters // *J. Biol. Chem*. 1995. V. 270. P. 25812–25818.
32. Ericsson J., Jackson S.M., Lee B.C., Edwards P.A. Sterol regulatory element binding protein binds to a cis element in the promoter of the farnesyl diphosphate synthase gene // *Proc. Natl. Acad. Sci. USA*. 1996. V. 93. P. 945–950.
33. Guan G., Jiang G., Koch R.L., Shechter I. Molecular cloning and functional analysis of the promoter of the human squalene synthase gene // *J. Biol. Chem*. 1995. V. 270. P. 21958–21965.
34. Wang X., Sato R., Brown M. et al. SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis // *Cell*. 1994. V. 77. P. 53–62.
35. Wang X., Pai J.-T., Wiedenfeld E.A. et al. Purification of an interleukin-1 beta converting enzyme-related cysteine protease that cleaves sterol regulatory element binding proteins between the leucine zipper and transmembrane domains // *J. Biol. Chem*. 1995. V. 270. P. 18044–18050.
36. Briggs M.R., Yokoyama C., Wang X. et al. Nuclear protein that binds sterol regulatory element of low density lipoprotein receptor promoter. I. Identification of the protein and delineation of its target nucleotide sequence // *J. Biol. Chem*. 1993. V. 268. P. 14490–14496.
37. Fisher R.F., Long S.R. Interactions of NodD at the nod Box: NodD binds to two distinct sites on the same face of the helix and induces a bend in the DNA // *J. Mol. Biol*. 1993. V. 233. P. 336–348.
38. Gottfert M. Regulation and function of rhizobial nodulation genes // *FEMS Microbiol. Rev*. 1993. V. 10. P. 39–63.

39. Cook D., Dreyer D., Bonnet D. et al. Transient induction of a peroxidase gene in *Medicago truncatula* precedes infection by *Rhizobium meliloti* // *Plant Cell*. 1995. V. 7. P. 43–55.
40. Vijn I., Christiansen H., Lauridsen P. et al. A 200 bp region of the pea ENOD12 promoter is sufficient for nodule specific and nod factor induced expression // *Plant Mol. Biol.* 1995. V. 28. P. 1103–1110.
41. Horvath B., Heidstra R., Lados M. et al. Lipo oligosaccharides of *Rhizobium* induce infection related early nodulin gene expression in pea root hairs // *Plant J.* 1993. V. 4. P. 727–733.
42. Yang W.C., de Blank C., Meskiene I. et al. *Rhizobium* nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation // *Plant Cell*. 1994. V. 6. P. 1415–1426.
43. Savoure A., Magyar Z., Pierre M. et al. Activation of the cell cycle machinery and the isoflavonoid biosynthesis pathway by active *Rhizobium meliloti* Nod signal molecules in *Medicago microcallus* suspensions // *EMBO J.* 1994. V. 13. P. 1093–1102.
44. Minami E., Kouchi H., Cohn J.R. et al. Expression of the early nodulin, ENOD40, in soybean roots in response to various lipo-chitin signal molecules // *Plant J.* 1996. V. 10. P. 23–32.
45. Cheon C.I., Lee N.G., Siddique A.B. et al. Roles of plant homologs of Rab1p and Rab7p in the biogenesis of the peribacteroid membrane, a subcellular compartment formed de novo during root nodule symbiosis // *EMBO J.* 1993. V. 12. P. 4125–4135.
46. Nap J.P., Bisseling T. Developmental biology of a plant-prokaryote symbiosis: the legume root nodule // *Science*. 1990. V. 250. P. 948–954.
47. Kardailsky I., Yang W.C., Zalensky A. et al. The pea late nodulin gene PsNOD6 is homologous to the early nodulin genes PsENOD3/14 and is expressed after the leghaemoglobin genes // *Plant Mol. Biol.* 1993. V. 23. P. 1029–1037.

48. Kouchi H., Hata S. GmN56, a novel nodule-specific cDNA from soybean root nodules encodes a protein homologous to isopropylmalate synthase and homocitrate synthase // *Mol. Plant. Microbe Interact.* 1995. V. 8. P. 172–176.
49. Fu X.-Y., Schindler C., Improta T. et al. The proteins of ISGF-3, the interferon alpha-induced transcriptional activator, define a gene family involved in signal transduction // *Proc. Natl. Acad. Sci. USA.* 1992. V. 89. P. 7840–7843.
50. Qureshi S.A., Salditt-Georgieff M., Darnell J.E., Jr. Tyrosine-phosphorylated Stat1 and Stat2 plus a 48-kDa protein all contact DNA in forming interferon-stimulated-gene factor 3 // *Proc. Natl. Acad. Sci. USA.* 1995. V. 92. P. 3829–3833.
51. Pine R., Canova A., Schindler C. Tyrosine phosphorylated p91 binds to a single element in the ISGF2/IRF-1 promoter to mediate induction by IFN alpha and IFN gamma, and is likely to autoregulate the p91 gene // *EMBO J.* 1994. V. 13. P. 158–167.
52. David M., Petricoin E.F., Igarashi K. et al. Prolactin activates the interferon-regulated p91 transcription factor and the Jak2 kinase by tyrosine phosphorylation // *Proc. Natl. Acad. Sci. USA.* 1994. V. 91. P. 7174–7178.
53. Thanos D., Maniatis T. Identification of the rel family members required for virus induction of the human beta interferon gene // *Mol. Cell. Biol.* 1995. V. 15. P. 152–164.
54. Garoufalis E., Kwan I., Lin R. et al. Viral induction of the human beta interferon promoter: modulation of transcription by NF-kappa B/rel proteins and interferon regulatory factors // *J. Virol.* 1994. V. 68. P. 4707–4715.
55. Harada H., Takahashi E.-I., Itoh S. et al. Structure and regulation of the human interferon regulatory factor 1 (IRF-1) and IRF-2 genes: implications for a gene network in the interferon system // *Mol. Cell. Biol.* 1994. V. 14. P. 1500–1509.
56. Harada H., Fujita T., Miyamoto M. et al. Structurally similar but functionally distinct factors, IRF-1 and IRF-2, bind to the same regulatory elements of IFN and IFN-inducible genes // *Cell.* 1989. V. 58. P. 729–739.

57. Morimoto R. Regulation of heat shock transcription response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators // *Genes Dev.* 1998. V. 12. P. 3788–3796.

58. Kim D., Ouyang H., Li G.C. Heat shock protein hsp70 accelerates the recovery of heat-shocked mammalian cells through its modulation of heat shock transcription factor HSF1 // *Proc. Natl. Acad. Sci. USA.* 1995. V. 92. P. 2126–2130.

59. Li G.C., Yang S.H., Kim D. et al. Suppression of heat-induced hsp70 expression by the 70-kDa subunit of the human Ku autoantigen // *Proc. Natl. Acad. Sci. USA.* 1995. V. 92. P. 4512–4516.

60. Ghoshal K., Jacob S.T. Heat shock selectively inhibits ribosomal RNA gene transcription and down-regulates E1BF/Ku in lymphosarcoma cells // *Biochem. J.* 1996. V. 317. P. 689–695.

61. Ратнер В.А. Концепция молекулярно-генетических систем управления. Новосибирск, НГУ, 1993. 117 с.

62. Ратнер В.А., Жарких А.А., Колчанов Н.А. и др. Проблемы теории молекулярной эволюции. М.: Наука, 1985. 263 с.

63. Igarashi T., Kaminuma T. Development of a cell signaling Network Database // *Proc. Symp. Biocomput.* 1997. P. 187–197