

Кудряшова Ольга Михайловна

**Функциональная характеристика
микроРНК-мРНК интерактома человека**

Специальность 03.01.03 – Молекулярная биология

АВТОРЕФЕРАТ

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

Москва – 2021

Работа прошла апробацию в департаменте молекулярной и биологической физики
Федерального государственного автономного образовательного учреждения высшего
образования "Московский физико-технический институт (национальный исследовательский
университет)"

Научный руководитель: Скоблов Михаил Юрьевич, кандидат биологических наук,
Заведующий лабораторией функциональной геномики ФГБНУ «Медико-генетический научный
центр имени академика Н.П. Бочкова»

Ведущая организация: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения
Российской академии наук».

Защита состоится «21» сентября 2021 года в 13.00 на заседании диссертационного совета
ФБМФ03.01.03.004 по адресу: 141701, г. Долгопрудный Московской обл., Институтский
переулок, д. 9.

С диссертацией можно ознакомиться в библиотеке и на сайте Московского физико-технического
института (национальный исследовательский университет)

<https://mipt.ru/education/post-graduate/soiskateli-biologicheskie-nauki.php>.

Работа представлена «23» июня 2021 г. в Аттестационную комиссию федерального
государственного автономного образовательного учреждения высшего образования
«Московский физико-технический институт (национальный исследовательский университет)»
для рассмотрения советом по защите диссертаций на соискание ученой степени кандидата наук
в соответствии с п .3.1 ст. 4 Федерального закона «О науке и государственной научно -
технической политике»

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования и степень ее разработанности

Изучение механизмов регуляции экспрессии генов является одним из ключевых вопросов в молекулярной биологии. Различие в экспрессии генов является одним из основных двигателей эволюции молекулярных процессов, а изменения в регуляции широко исследуются в патогенезе широкого ряда заболеваний. Одним из таких эффективных механизмов регуляции экспрессии генов является «РНК-интерференция», благодаря которой происходит специфичное подавление экспрессии генов с помощью малых некодирующих РНК. За объяснение данного механизма подавления экспрессии генов по средствам образования двуцепочечного дуплекса РНК в 2006 году Эндрю Фаер и Крэйг Мелло получили Нобелевскую премию по физиологии и медицине.

микроРНК - класс малых некодирующих РНК, имеющих длину от 18 до 25нт и встречающихся в большинстве эукариотических организмов (Carthew R.W. 2009). Совместно с белком Ago2 микроРНК образует RISC комплекс, который связывается с мРНК и приводит либо к ингибированию трансляции, либо к деградации мРНК. Таким образом, микроРНК играет важную роль в регуляции экспрессии генов. Для человека в настоящий момент известно более 2'500 микроРНК [miRBase v.22, <http://www.mirbase.org>]. Широко известно, что микроРНК являются мощными регуляторами, которые принимают активное участие во многих молекулярных и клеточных процессах, как в норме, так и при различных патологиях. При этом показано, что один ген может регулироваться несколькими микроРНК и одна микроРНК может регулировать несколько сотен мишеней (Selbach 2008). Однако до сих пор отсутствует полноценное понимание всех образующихся взаимодействий - микроРНК – мРНК интерактома человека.

Распространённым способом поиска микроРНК-мРНК взаимодействий является использование предсказательных программ. Их алгоритмы основаны на различных наборах параметров, характеризующих потенциальные взаимодействия, таких как: комплементарность, каноничность, свободная энергия взаимодействия, консервативность и другие (Witkos T.M. 2011). К сожалению, полученные с помощью этих программ результаты часто сильно разнятся (Li H. 2014). В связи с этим распространена практика использования одновременно нескольких алгоритмов и дальнейшее изучение только тех результатов, которые являлись общими для нескольких программ.

Для экспериментального подтверждения микроРНК-мРНК взаимодействий часто используется репортерная люциферазная система. Однако такой подход способен провести лишь анализ единичных потенциальных взаимодействий (Vaccarini A. 2010). Ряд экспериментальных методов (CLIP методы) основанных на иммунопреципитации белка Ago2 с последующим

секвенированием всех регионов мРНК, входящих в комплекс RISC (Hafner M. 2010), позволяет определить лишь участки взаимодействия мРНК, но не конкретные микроРНК-мРНК взаимодействия. Недавно разработанный и апробированный на клеточной линии HEK293 метод CLASH (crosslinking, ligation and sequencing of hybrids) (Helwak A. 2013) и аналогичный ему метод CLEAR-CLIP (covalent ligation of endogenous Argonaute-bound RNAs) (Moore M.J. 2015), проведенный на клеточной линии Huh7.5, позволяют определять микроРНК-мРНК взаимодействия. Оба метода имеют дополнительный этап, заключающийся в лигировании концов мРНК и микроРНК, которые находятся в комплексе RISC. На данный момент эти две работы являются единственными по получению экспериментальных высокопроизводительных данных о микроРНК-мРНК взаимодействиях на клеточных линиях человека.

В настоящей работе был впервые проведен всесторонний анализ микроРНК-мРНК интерактома человека на основе экспериментальных данных. Результаты анализа позволяют расширить имеющиеся фундаментальные представления о микроРНК-мРНК взаимодействиях человека. На основе полученных данных была разработана программа “exp-miBR Annotator”, которая позволяет находить варианты нуклеотидной последовательности, способные нарушать связывание микроРНК с мРНК, что может приводить к нарушению регуляции экспрессии генов посредством микроРНК. С помощью разработанной программы “exp-miBR Annotator” были определены известные патогенные варианты нуклеотидной последовательности, находящиеся в сайтах связывания с микроРНК и проведена экспериментальная проверка *in-vitro* для нескольких найденных вариантов.

Цель и задачи настоящей работы

Целью данной работы является всесторонний анализ микроРНК-мРНК взаимодействий человека на основе экспериментальных данных. Для осуществления данной цели были поставлены следующие **задачи**:

1. Создание базы данных экспериментально установленных микроРНК-мРНК взаимодействий человека.
2. Проведение оценки эффективности биоинформатических алгоритмов по предсказанию сайтов связывания микроРНК с мРНК.
3. Проведение сравнительного анализа уровня экспрессии и количества взаимодействий для мРНК и микроРНК.

4. Разработка программы, позволяющей проводить поиск вариантов нуклеотидной последовательности ДНК в регионах мРНК человека, достоверно взаимодействующих с микроРНК.
5. Поиск функционально значимых вариантов нуклеотидной последовательности, нарушающих микроРНК-мРНК регуляцию.

Положения Диссертации, выносимые на защиту:

1. Впервые создана база данных достоверных, экспериментально идентифицированных сайтов связывания микроРНК (exp-miBRs) человека. Собранная база данных exp-miBRs позволяет проводить исследования регуляции экспрессии генов человека с помощью микроРНК.
2. Для микроРНК let-7 и онкогена CDK6 из базы данных exp-miBRs было экспериментально подтверждено микроРНК-мРНК взаимодействие. Таким образом дополнительно была показана высокая достоверность собранных взаимодействий в базе данных exp-miBRs.
3. Показан низкий уровень эффективности программ, предсказывающих микроРНК-мРНК взаимодействия (TargetScan, PicTar, PITA, RNA22 и miRanda) в сравнении с экспериментальными данными. Таким образом, показана необходимость использования экспериментально выявленных микроРНК-мРНК взаимодействий для разработки программ нового поколения.
4. Сравнительный анализ микроРНК-мРНК интерактома и профилей экспрессий мРНК и микроРНК позволил полноценно охарактеризовать данные взаимодействия в клеточных линиях человека HEK293 и Huh7.5. Были выявлены ткане-специфичные мРНК, проявляющие эффект «губки микроРНК» и группы микроРНК, имеющие разные свойства по уровню экспрессии и количеству взаимодействий.
5. Используя базу данных exp-miBRs, была разработана веб-программа “exp-miBR Annotator”, которая позволяет выявлять варианты нуклеотидной последовательности человека в регионах связывания с микроРНК (exp-miBRs). Таким образом, программа “exp-miBR Annotator” помогает находить новые механизмы патогенеза наследственных заболеваний, вызванных нарушением регуляции экспрессии генов посредством микроРНК. На примере генов TF и SNORD118 были найдены патогенные варианты, для которых было экспериментально подтверждено их влияние на микроРНК-мРНК взаимодействие.

Научная новизна

В настоящей работе был впервые проведен всесторонний анализ микроРНК-мРНК интерактома человека с использованием экспериментальных данных, полученных на клеточных линиях человека HEK293 и Huh7.5. По результатам была создана база данных достоверных, экспериментально идентифицированных сайтов связывания микроРНК человека. В неё вошли 46 тысяч регионов мРНК (exp-miBR). На основе базы данных exp-miBRs была впервые разработана программа “exp-miBR Annotator”, позволяющая определять варианты нуклеотидной последовательности генома человека в сайтах связывания микроРНК. С помощью разработанной программы “exp-miBR Annotator” были определены известные патогенные варианты нуклеотидной последовательности, находящиеся в сайтах связывания с микроРНК и проведена экспериментальная проверка *in-vitro* для нескольких найденных вариантов.

Теоретическое и практическое значение исследования

Научная и практическая значимость работы определяется поставленной фундаментальной научной задачей – исследование микроРНК-мРНК интерактома человека. микроРНК являются эффективными регуляторами экспрессии генов. Понимание природы мРНК-микроРНК взаимодействия позволяет не только определить фундаментальную роль в молекулярных процессах, но и определять механизмы патогенеза наследственных заболеваний, вызванных нарушением регуляции экспрессии генов по средствам микроРНК. Результаты работы были номинированы на 50-ой юбилейной конференции ESHG'2017 в Дании, организованной Европейской ассоциацией генетики человека (The European Society of Human Genetics - ESHG), став единственной номинированной работой за выдающиеся результаты молодых учёных Young Investigator Awards из России (<https://2017.eshg.org/index.php/abstracts-2/speaker/>).

По итогам работы была собрана база данных экспериментально установленных высокодостоверных микроРНК-мРНК взаимодействий (exp-miBRs), на основе которой разработана программа для анализа вариантов нуклеотидной последовательности. Как было показано, данный инструмент может играть важную роль при поиске новых механизмов и анализе вариантов у больных наследственными заболеваниями с неизвестным на данный момент механизмом патогенеза. База данных exp-miBRs была использована компанией geneXplain (<http://genexplain.com>) в коммерческом продукте TRANSFAC® 2020.1 – добавлено в релизе января 2020 года - http://genexplain.com/wp-content/uploads/2020/01/TRANSFAC_release.pdf

Работы, опубликованные по результатам данной диссертации в международных журналах активно цитируются, несмотря на то что были опубликованы совсем недавно. Так, например, публикация (Plotnikova 2019) была процитирована 11 раз за год: девять раз в международных

журналах: (Telford 2021), (Rohani 2021), (Vannan 2021), (Hussen 2021), (Gao 2021), (Fitriawan 2020), (Ueno 2020), (Najafipour 2021), (Annese 2020), в книжном издании (Schober 2020) и в диссертационной работе (Frazier 2021).

Степень достоверности и апробация диссертационной работы

Степень достоверности результатов данной работы подтверждается согласованностью результатов с опубликованными ведущими отечественными и зарубежными работами в области исследования микроРНК регуляции и роли микроРНК в патогенезе наследственных заболеваний. Основные результаты получены с использованием современных методов биоинформатического анализа и статистических методов.

По материалам диссертации опубликовано три статьи в рецензируемых научных журналах: две из них входят в основные библиометрические базы данных (РИНЦ, PubMed, WoS и Scopus) и одна индексируется базой данной РИНЦ:

1. **Plotnikova O.**, Baranova A., and Skoblov M. "Comprehensive analysis of human microRNA-mRNA interactome." *Frontiers in Genetics*, 2019, (10):933, doi.org/10.3389/fgene.2019.00933
2. **Плотникова О.М.**, Скоблов М.Ю. «Эффективность программ, предсказывающих микроРНК-мРНК взаимодействия» («efficiency of the miRNA- mRNA interaction prediction programs») *МОЛЕКУЛЯРНАЯ БИОЛОГИЯ*, 2018, 52(3):1–12
3. **Плотникова О.М.**, Скоблов М.Ю «Роль микроРНК в патогенезе наследственных заболеваний» *Медицинская генетика* 2020, 19(9):5-17

Положения и выводы данной работы были апробированы на международных конференциях, как в рамках представления полученных результатов в виде устного доклада, так и в рамках доклада на постерных сессиях. Основные результаты настоящей диссертации были представлены на 11 конференциях:

Устный доклад:

1. **Плотникова О.М.**, Скоблов М.Ю. «Анализ характеристик микроРНК-мРНК интерактома человека и оценка алгоритмов предсказания», 59 научная конференция МФТИ, г. Долгопрудный МФТИ, 21-26 ноября 2016г.
2. **Plotnikova O.M.**, Skoblov M.Y., "Expression insights into the human miRNA-mRNA interactome", C22.2, European Human Genetic Conference 2017, Denmark, Copenhagen, 24-31 May 2017.
3. **Plotnikova O.M.**, Skoblov M.Y., «Expression analysis of human miRNA - mRNA interactome», MCCMB 2017, Moscow, MSU 27-30.07.2017
4. **Плотникова О.М.**, Скоблов М.Ю. «Проблемы анализа микроРНК-мРНК интерактома человека», 60 научная конференция МФТИ, г. Долгопрудный МФТИ, 24 ноября 2017г.

5. **Плотникова** О.М., Скоблов М.Ю. “Идентификация качественного микроРНК-мРНК интерактома человека для анализа вариантов нуклеотидной последовательности полученных с помощью NGS”, международная конференция Ломоносов 2018, МГУ, Москва, 11 апреля 2018г.
6. **Plotnikova** O., Skoblov M., “Arranging The High-Confidence MicroRNA Binding Regions Into Web Tool For Analyzing Variants In Diagnostic Next Generation Sequencing” BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE/SYSTEMS BIOLOGY (BGRS\SB-2018); Symposium COGNITIVE SCIENCES, GENOMICS AND BIOINFORMATICS (CSGB-2018), Россия, Новосибирск, 24 августа 2018.
7. **Plotnikova** O.M., Skoblov M.Y., “Analysis of experimentally derived landscape of microRNA-mRNA human interactome”, MCCMB 2019, Moscow, MSU 27-30.07.2019

Постерный доклад:

1. **Плотникова** О.М., Скоблов М.Ю. «миРНК-мРНК ВЗАИМОДЕЙСТВИЯ – АНАЛИЗ ОСНОВНЫХ ХАРАКТЕРИСТИК И ПРЕДСКАЗАТЕЛЬНЫХ АЛГОРИТМОВ», Всероссийская конференция 50 лет ВОГиС: успехи и перспективы, Москва, 8-10 ноября 2016г.
2. **Плотникова** О.М., Скоблов М.Ю. «Анализ экспрессии и потенциала взаимодействия микроРНК-мРНК интерактома человека», Ломоносов 2017, Россия, Москва, 10-14 апреля 2017г.
3. **Plotnikova** O., Skoblov M., “Tool for pathogenesis analysis of nucleotide variants based on high-confidence human microRNA-mRNA interactions”, Bertinoro, Italy, 3-4 May, 2018
4. **Plotnikova** O.M., Skoblov M.Y., «Tool for pathogenesis analysis of nucleotide variants based on high-confidence human miRNA-mRNA interactions», European Human Genetic Conference 2018 (ESHG 2018), Италия, Милан, 16-19 июня, 2018

Работы вышли до смены фамилии (до августа 2020 – Плотникова О.М., после – Кудряшова О.М.).

Методология и методы исследования

Анализ, проведенный в рамках данной диссертационной работы, основывается на результатах двух единственных работ, опубликованных на данный момент и использующих высокопроизводительные методы для экспериментального определения микроРНК-мРНК взаимодействий (химерных прочтений) на клеточных линиях человека: метод CLASH (Helwak A. 2013) выполнен на клеточной линии HEK293 и метод CLEAR-CLIP (Moore M.J. 2015) - на клеточной линии Huh7.5. Также в анализе используются публичные экспериментальные данные о регионах взаимодействий микроРНК (CLIP-технологии) и об экспрессии микроРНК и мРНК в клеточных линиях HEK293 и Huh7.5. Основные результаты получены с использованием современных методов биоинформатического анализа, подробно описанные в главе 2 «Материалы и методы». Экспериментальная проверка, выявленных в данной работе микроРНК-мРНК взаимодействий, была проведена в лаборатории функциональной геномики в Медико-генетическом научном центре имени академика Н.П. Бочкова.

Личный вклад автора

Автор принимал участие в планировании работы, сборе, изучении и обработке публичных экспериментальных данных, проведении анализа данных, а также в написании и подготовке публикаций. Основные результаты диссертационной работы получены лично автором. Веб-интерфейс разработанной автором программы «ExpriBR Annotator» был реализован совместно с Корвиго Ильей. Экспериментальная проверка выявленных автором микроРНК-мРНК взаимодействий проводилась в лаборатории функциональной геномики в Медико-генетическом научном центре имени академика Н.П. Бочкова Филатовой Александрой.

Объем и структура диссертации.

Данная диссертационная работа содержит следующие разделы: обзор литературы (Глава 1), материалы и методы (Глава 2), результаты (Глава 3), заключение, список сокращений и список цитируемой литературы. Объем диссертации 84 страницы, включая 20 рисунков и 5 таблиц. Список цитируемой литературы содержит 127 источников.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Введение

Во **введении** описана актуальность темы диссертационной работы, поставлена цель и сформулированы задачи исследования. Описана научная новизна и практическая значимость, описана апробация полученных результатов, вклад автора, представлены положения к защите.

Глава 1

В **главе 1** содержится краткий обзор микроРНК – их открытия, биогенеза и функции. Представлены существующие подходы к изучению микроРНК регуляции и их ограничения. Раскрыты как экспериментальные, так и биоинформатические подходы определения микроРНК-мРНК взаимодействий. Собраны и представлены известные на данный момент механизмы наследственных заболеваний, патогенез которых связан с нарушением микроРНК регуляции. Освящен вопрос функциональности взаимодействия микроРНК с мРНК

Глава 2. Материалы и методы.

Во второй главе описаны методологические подходы, использованные для получения результатов. Схематически, проведенный биоинформатический анализ и экспериментальную валидацию можно поделить на модули, представленные на схеме (Рисунок 1):



Рисунок 1. Схема проведенных анализов в рамках диссертационной работы, состоящая из следующих разделов: сбор имеющихся экспериментальных данных высокопроизводительного секвенирования о взаимодействующих микроРНК-мРНК для клеточных линий человека

(данные CLASH и CLEAR-CLIP в центре); оценка работы программ по предсказанию микроРНК-мРНК взаимодействий (1); сравнение данных CLASH и CLEAR-CLIP (2); анализ связи уровня экспрессии и количества взаимодействий для микроРНК (3) и мРНК (4); дополнительный сбор экспериментальных данных о регионах взаимодействия микроРНК (технология CLIP) для идентификации достоверных сайтов связывания микроРНК (exp-miBRs) (5); разработка программы для поиска вариантов нуклеотидной последовательности в сайтах связывания с микроРНК (6); поиск функционально значимых вариантов нуклеотидной последовательности, нарушающих микроРНК-мРНК взаимодействие и их экспериментальная валидация (7).

Данный анализ основывается на двух единственных результатах работ, опубликованных на данный момент и использующих высокопроизводительные методы для экспериментального определения микроРНК-мРНК взаимодействий (химерных прочтений) на клеточных линиях человека: метод CLASH (Helwak A. 2013) выполнен на клеточной линии HEK293 и метод CLEAR-CLIP (Moore M.J. 2015) - на клеточной линии Huh7.5

Краткое описание методов приведенных на схеме:

1. Для оценки эффективности программ предсказания микроРНК-мРНК взаимодействий на регионах экспериментальных данных CLASH - была использована база данных Starbase v2.0. Анализ предсказанных взаимодействий и экспериментально выявленных регионов связывания микроРНК проводился с помощью скриптов на языке python.
2. Данные о микроРНК-мРНК химерных взаимодействиях CLASH и CLEAR-CLIP были биоинформатически обработаны единым образом и приведены к координатам референсного генома человека версии hg19. Сравнение результатов данных методов производилось между транскрибируемыми регионами. Для оценки ожидаемого пересечения вместо данных CLASH и CLEAR-CLIP были случайным образом искусственно сгенерированы аналогичные наборы данных.
3. Данные об уровне экспрессии мРНК были взяты из публичных экспериментов: GSE68611 для HEK293 и GSE64677 для Huh7.5. Каждый датасет содержал по две независимые биологические повторности. Оценка качества данных проводилась с помощью FastQC, уровень экспрессии мРНК был рассчитан программой kallisto.
4. Публичные данные об экспрессии микроРНК (GSE75136 два независимых эксперимента для HEK293 и GSE74014 – три биологические повторности для Huh7.5) были обработаны и нормализованы с помощью пакета DeSeq2 для сравнимости уровня экспрессии микроРНК между клеточными линиями. Консервативность нуклеотидных последовательностей, кодирующих микроРНК, были получены из геномного браузера UCSC. Использовалось значение phyloP, рассчитанное между 20 позвоночными.

5. Для определения высоко-достоверных, экспериментально выявленных регионов связывания микроРНК (exp-miBRs), дополнительно к данным CLASH и CLEAR-CLIP были использованы 79 публичных экспериментов CLIP с платформы POSTAR. Данные были обработаны одинаковыми алгоритмами: сырые прочтения – с помощью программы FASTX-Toolkit, далее специальными программами: PARalyzer для данных PAR-CLIP (N = 18) и CIMS для данных HITS-CLIP (N = 61). Для идентификации exp-miBRs был написан скрипт на языке python.
6. Выявленные exp-miBRs были использованы для создания веб-программы “exp-miBR Annotator” (<http://score.generesearch.ru/services/mirna/>). Данная программа позволяет по входным координатам нуклеотидной последовательности ДНК человека (в виде единичной позиции или VCF файла) определять их нахождение в регионах, достоверно взаимодействующих с микроРНК (в exp-miBR).
7. Для поиска функционально значимых вариантов нуклеотидной последовательности, нарушающих микроРНК-мРНК взаимодействие, были использованы данные о патогенных врожденных вариантах из базы данных ClinVar и проведена оценка их нахождения в регионах связывания с микроРНК (exp-miBRs) с помощью разработанной программы “exp-miBR Annotator” и скриптов на языке python. Консенсусный список онкогенов и генов супрессоров опухолевого роста (COSMIC) был использован для отбора микроРНК-мРНК взаимодействий и последующего экспериментального изучения влияния внесенных мутаций в exp-miBR на регуляцию экспрессии гена-мишени. Экспериментальная проверка проводилась на клеточной линии HEK293 с использованием плазмидной конструкции psiCHECK2, имеющей репортерные гены кодирующие люциферазы Firefly и Renilla.

Глава 3. Результаты.

В третьей главе описаны основные результаты диссертационной работы. А именно: описано создание базы данных экспериментально установленных микроРНК-мРНК взаимодействий; проведена оценка эффективности программ, предсказывающих микроРНК-мРНК взаимодействия; проведен экспрессионный анализ микроРНК и мРНК; описано создание программы «exp-miBR Annotator» для определения нахождения варианта нуклеотидной последовательности в сайте связывания с микроРНК; проведен поиск функционально значимых вариантов нуклеотидной последовательности, нарушающих микроРНК-мРНК регуляцию и экспериментально изучено влияние патогенных вариантов на нарушение сайта связывания микроРНК- мРНК на примере онкогена CDK6 и микроРНК let-7.

Создание базы данных экспериментально установленных микроРНК-мРНК

взаимодействий

Сравнение результатов высокопроизводительных методов определения микроРНК-мРНК взаимодействий

На данный момент опубликовано две работы по экспериментальному определению микроРНК-мРНК взаимодействий высокопроизводительными методами на клеточных линиях человека: метод CLASH (Helwak A. 2013) выполнен на клеточной линии HEK293 и метод CLEAR-CLIP (Moore M.J. 2015) - на клеточной линии Huh7.5. На первом этапе нами было проведено сравнение результатов данных методов CLASH и CLEAR-CLIP. Основным их техническим отличием является наличие дополнительного шага в методе CLASH (проведение обогащение комплексов белка Ago2), а также различная биоинформатическая обработка данных после секвенирования. Так, данные CLASH картировались на последовательность транскриптома человека, что привело к идентификации 18 тысяч микроРНК-мРНК взаимодействий в клеточной линии HEK293. Данные CLEAR-CLIP были откартированы на последовательность референсного генома hg19, в результате чего было найдено 32 тысячи микроРНК-мРНК взаимодействий в клеточной линии Huh7.5.

Для сравнения методик, на первом этапе мы сфокусировались на анализе микроРНК-мРНК взаимодействий, находящихся в транскриптах генов человека, таким образом для CLEAR-CLIP было оставлено 10 тысяч взаимодействий. Сравнение регионов мРНК, взаимодействующих с микроРНК, для данных полученных методами CLASH и CLEAR-CLIP, выявило схожее распределение: половина всех взаимодействий находилась в кодирующей области мРНК (CDS), ~40% в 3'-нетранслируемой области (UTR). Такое распределение совпадало со средним

значением длин данных регионов по всем белок кодирующим транскриптам (версии GENCODE) (Рисунок 2А).

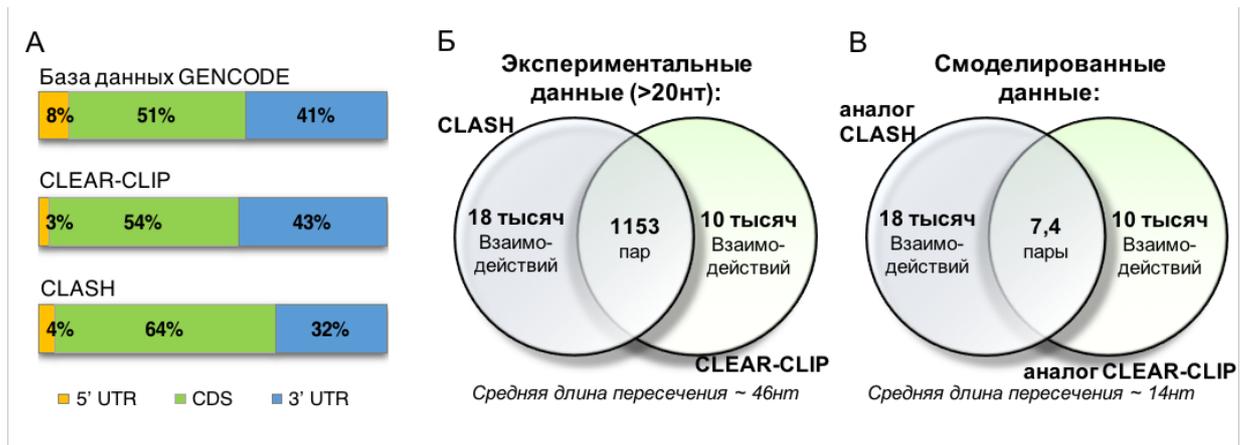


Рисунок 2. Анализ данных CLASH и CLEAR-CLIP. Распределение суммарных длин регионов мРНК в данных CLASH, CLEAR-CLIP и всем транскриптам согласно базе данных GENCODE (А). Диаграмма Венна показывает количество общих регионов между данными экспериментальных методов CLEAR-CLIP и CLASH (Б) и между случайно созданными искусственными аналогами данных CLASH и CLEAR-CLIP (В).

Пересечение результатов данных методов выявило 1'153 общих взаимодействий, которые были образованы 993 микроРНК-мРНК взаимодействиями в данных CLASH и 944 взаимодействиями в данных CLEAR-CLIP (Рисунок 2Б). Большая часть общих взаимодействий имели длину перекрытия более 20нт, при средней длине 45,8 нт +/- 13,9 нт. Так как данные были получены на различных клеточных линиях, в которых экспрессируются частично пересекающиеся гены и микроРНК, то такое небольшое пересечение может объясняться биологическим различием тканей. Для установления, не является ли данное пересечение случайным, нами было проведено сравнение пересечения экспериментальных данных с пересечением искусственно смоделированных взаимодействий. Было проведено создание нескольких независимых групп данных случайных взаимодействий и проведён анализ их пересечения. В среднем пересечение составляло 7,4 +/- 1,3 взаимодействия и имело среднюю длину перекрытия сегментов 14 нт +/- 6,7 нт (Рисунок 2В). Таким образом было показано, что экспериментальное пересечение данных CLASH и CLEAR-CLIP является статистически значимым ($p\text{-value} < 0,0001$). А наблюдаемое небольшое пересечение взаимодействий может объясняться ткане-специфичной экспрессией мРНК и микроРНК в данных клеточных линиях.

Выявление набора достоверных, экспериментально идентифицированных сайтов связывания микроРНК (*exp-miBRs*) с использованием данных CLIP экспериментов

Для поиска достоверных сайтов связывания микроРНК нами были дополнительно собраны и биоинформатически обработаны данные из 79 публичных CLIP экспериментов,

содержащих регионы взаимодействия микроРНК (CLIP методы не позволяют установить какая именно микроРНК взаимодействует в регионе). Собранные CLIP данные были получены на 9 различных клеточных линий, наибольшее количество экспериментов соответствовало клеточным линиям HEK293 (N=34) и Huh7.5 (N=19).

Комбинация всех микроРНК-взаимодействующих регионов (CLASH, CLEAR-CLIP, CLIP данные) выявила 156 тысяч уникальных микроРНК связывающих регионов. В дальнейшем из них были оставлены только высоко-достоверные экспериментальные области (exp-miBR), удовлетворяющие следующим критериям: (1) регион содержит непрерывную нуклеотидную последовательность, которая была подтверждена минимум двумя независимыми экспериментами или минимум двумя различными химерами из данных CLASH/CLEAR-CLIP и (2) длина такой последовательности составляет минимум 10нт (Рисунок 3).

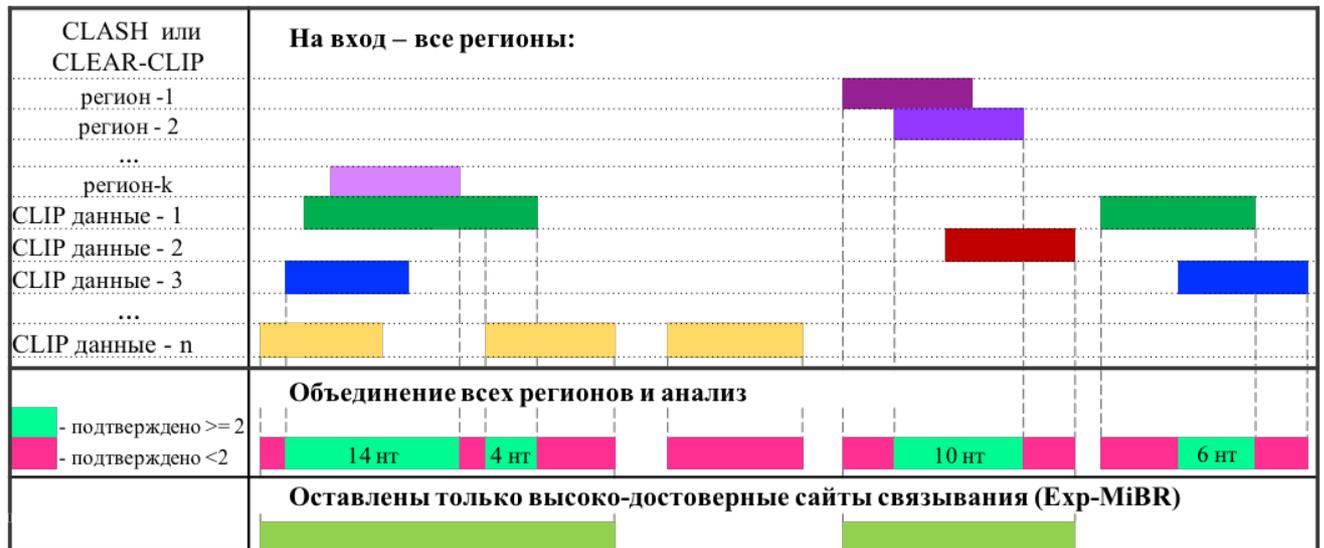


Рисунок 3. Схема определения высоко-достоверных экспериментальных микроРНК связывающих регионов (exp-miBR) из данных CLASH, CLEAR-CLIP и 79 CLIP экспериментов.

Таким образом, нами было получено 46'805 высоко-достоверных экспериментально подтвержденных регионов связывания микроРНК (Exp-MiBR – experimentally confirmed microRNA-binding regions), покрывающих 15 тысяч генов. Половина exp-miBR была расположена в 3'-нетранслируемой области генов, 24% - в кодирующей части гена, 10% в интронных и 6% в межгенных регионах (6%); остальные exp-miBR расположены в областях некодирующих РНК. Для 17% exp-miBR есть данные минимум об одной экспериментально определенной, взаимодействующей микроРНК согласно данным CLASH или CLEAR-CLIP.

Анализ характеристик выявленных exp-miBRs

Анализ выявленных exp-miBR показал, что часть регионов мРНК имеет специфичность для определенных клеточных линий, а некоторые регионы встречаются почти во всех клеточных линиях. Так, было выявлено ~ 9 тысяч уникальных exp-miBR, характерных только для клеточной линии HEK293 и около 700 регионов, уникальных для клеточной линии Huh7.5. Большое количество специфичных exp-miBR для клеточной линии HEK293 может объясняться большей представленностью экспериментов для данной клеточной линии. Также были выявлены exp-miBR, встречающиеся в большинстве клеточных линий и вероятно участвующие в общих процессах клеток (выполняющие “house-keeping” функции). Примерно 1% exp-miBR встречался в семи или более клеточных линиях. Для 351 гена, которые содержали данные «общие» exp-miBR, был проведен анализ обогащения по функциональной принадлежности с помощью программы Panther. Было выявлено, что данные гены были статистически значимо перепредставлены в следующих путях: процессы, связанные с жизненным циклом клетки (cellular process of cell cycle, FC 3,17 и p-value $1e10-8$); молекулярные процессы, связывания с ДНК (molecular function of nucleic acid binding, FC 1.75 и p-value $5e10-4$).

Большая часть exp-miBR (68%) имели длину 20-40нт, близкую к длине микроРНК (22нт), а 99% exp-miBR имели длину меньше 150нт. Был выявлен кластер взаимодействий микроРНК, имеющий наибольшую длину 631нт. Данный кластер был подтвержден в 54 различных экспериментах на 9 клеточных линиях. Также были выявлены exp-miBR, находящиеся близко друг к другу и формирующие кластер взаимодействий. Например, на второй хромосоме кластер exp-miBR охватывает область в примерно 1,5 тысячи нуклеотидов и находится между локусами генов *RNA5-8SP5* и *MIR663B*. Согласно данным CLASH и CLEAR-CLIP с данным кластером взаимодействует 52 различные микроРНК.

Нами было установлено, что exp-miBR покрывают 86% митохондриального генома, включая 35 из 37 известных генов в митохондриях человека. 37 таких exp-miBR были выявлены во всех девяти исследуемых клеточных линиях, в то время как все exp-miBR в митохондриальном геноме подтверждаются в среднем в 11 различных экспериментах. С данными регионами связываются 182 различные микроРНК. Это подтверждает описанное ранее мнение, что микроРНК-мРНК-взаимодействия происходят в различных клеточных компартментах, включая митохондрии (Ni 2015).

Таким образом, нами были собраны и охарактеризованы высоко-достоверные, экспериментально установленные сайты связывания с микроРНК для человека.

Оценка эффективности программ по предсказанию микроРНК-мРНК взаимодействий

Нами была проведена оценка эффективности программ по предсказанию микроРНК-мРНК взаимодействий в сравнении с экспериментальными данными на клеточной линии HEK293 (данные CLASH экспериментов).

Используя результаты высокопроизводительного метода CLASH (crosslinking, ligation and sequencing of hybrids), распознающего микроРНК и взаимодействующего с ней региона мРНК (Helwak A. 2013), была оценена эффективность наиболее часто используемых программ, предсказывающих микроРНК-мРНК взаимодействия: TargetScan (Agarwal V. 2015), PicTar (Lall S. 2006) (Krek A. 2005), PITA (Kertesz M. 2007), RNA22 (Miranda K.C. 2006) и miRanda (John B. 2004). Анализ эффективности предсказательных алгоритмов проводился по следующим параметрам: чувствительность, положительная прогностическая значимость, анализ взаимодействующих регионов мРНК (кодирующая область, 3'- и 5'-некодирующие области), каноничность и неканоничность взаимодействий, сравнение со случайно генерируемыми наборами данных.

Предсказания микроРНК в регионах мРНК, идентифицированных методом CLASH

Экспериментальные данные CLASH (Helwak A. 2013) установили 18 тысяч взаимодействий, образованных 6,5 тысячами мРНК и 386 микроРНК на 15'831 регионе мРНК (некоторые регионы содержат сразу несколько взаимодействий с различными микроРНК). Анализ предсказаний проводился только на регионах мРНК, экспериментально подтвержденных данными CLASH как взаимодействующие с микроРНК. Для каждого взаимодействия программа могла предсказать микроРНК, совпадающую с экспериментальными данными, тогда предсказание классифицировалось как «Верное предсказание», предсказать микроРНК, которая не была установлена экспериментом, тогда результат классифицировался как «Неверное предсказание» или не предсказать ни одной микроРНК.

По результатам предсказаний пяти программ было получено 19'398 предсказанных микроРНК-мРНК взаимодействий в экспериментально установленных регионах. Программы предсказывали разное количество взаимодействий: так больше всего взаимодействий было предсказано программами miRanda (9'758 взаимодействий) и PicTar (8'562 взаимодействий), а меньше всего взаимодействий было предсказано программой RNA22 (3'439) и PITA (3'199). При этом процент верно предсказанных взаимодействий для всех программ составлял ~3,5% (Рисунок 4). Суммарно всеми программами было предсказано только 624 верных микроРНК-мРНК взаимодействий.

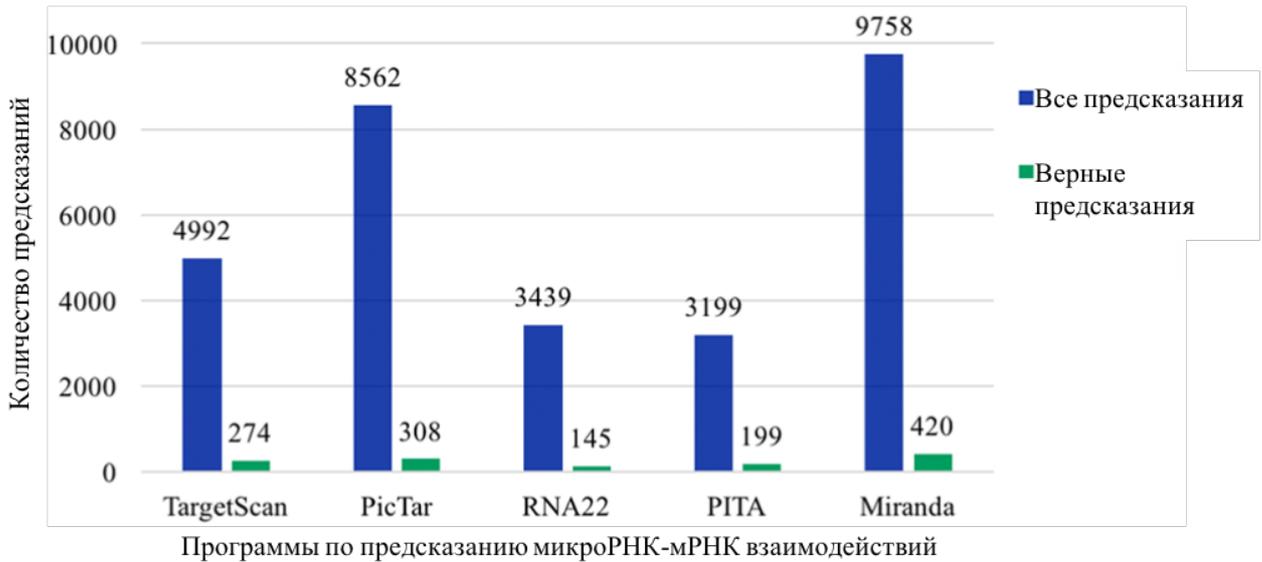


Рисунок 4. Распределение предсказанных и верно-предсказанных микроРНК-мРНК взаимодействий по программам.

Сравнение результатов предсказаний и экспериментальных данных показало, что программы предсказывают только в четверти (26%) экспериментальных регионов, а в остальных регионах мРНК не было ни одного предсказания. Также, было установлено, что программы предсказывают взаимодействия для 384 различных микроРНК, из которых только 168 совпадает с экспериментально установленными 386 микроРНК.

Анализ характеристик регионов взаимодействия мРНК

На следующем этапе нами были подробно изучены микроРНК-мРНК взаимодействия, полученные как экспериментально, так и предсказанные с помощью программ. Изучение распределения регионов мРНК, взаимодействующих с микроРНК выявило преимущественную локализацию предсказанных взаимодействий в 3'-нетранслируемой области, в то время как экспериментально установленные взаимодействия встречались во всех регионах мРНК: большая часть, 58%, взаимодействий лежит в CDS-области мРНК, 38% – в 3'-UTR и 4% – в 5'-UTR. При анализе средних длин всех известных для человека транскриптов (согласно данным GENCODE) выявлено, что микроРНК-мРНК взаимодействия, полученные методом CLASH, распределены равномерно в соответствии со средними длинами областей мРНК (Рисунок 5А).

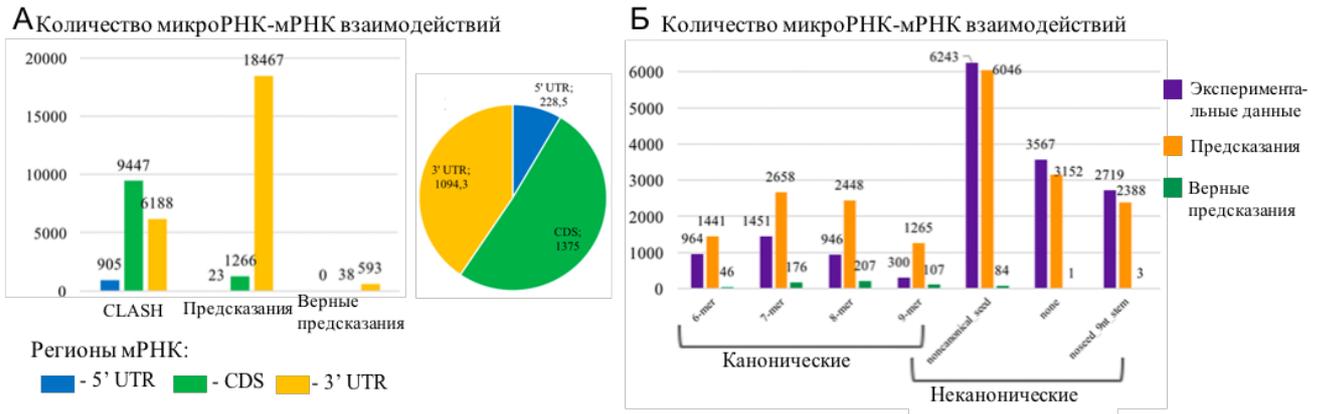


Рисунок 5. Сравнение предсказанных и экспериментально выявленных на клеточной линии HEK293 микроРНК-мРНК взаимодействий. (А) Анализ микроРНК-мРНК взаимодействий по регионам мРНК: 5'-нетранслируемой области (синий), кодирующей области (зеленый) и 3'-нетранслируемой области (желтый). Круговая диаграмма справа показывает средние длины данных регионов по всем транскриптам согласно GENCODE. (Б) Распределение микроРНК-мРНК взаимодействий по каноническим (4 вида) и неканоническим (3 вида) типам согласно данным CLASH и результатам предсказаний пяти программ.

Анализ первых экспериментальных данных выявил «канонические» правила взаимодействия микроРНК с мишенями, согласно которому главным фактором, определяющим взаимодействие является комплементарное взаимодействие между “seed”-регионом микроРНК (со 2 по 8 нуклеотид на 5'-конце) и регионом мРНК (Bartel 2009). Нами был проведен сравнительный анализ между экспериментальными микроРНК-мРНК взаимодействиями и предсказанными по типам взаимодействий (Рисунок 5Б). Все пять программ предсказывают больше (60%) неканонических взаимодействий. Однако процент верно предсказанных взаимодействий выше у канонических микроРНК-мРНК взаимодействий, причем чем длиннее “seed” регион комплементарного взаимодействия, тем выше процент верно предсказанных взаимодействий. При анализе данных CLASH выявлено, что микроРНК-мРНК взаимодействия представлены всеми известными типами, при этом большинство из них неканонические (77%).

Расчёт основных предсказательных характеристик программ

На последнем этапе для исследуемых пяти программ, предсказывающих микроРНК-мРНК взаимодействия, были оценены параметры чувствительности и положительной прогностической значимости (PPV). Результаты представлены в Таблице 1. Использование результатов всех программ одновременно позволяет увеличить чувствительность до 3.9%, но при этом PPV, наоборот, достигает минимального значения – 3.2%. Оценка предсказанных взаимодействий, сделанных одновременно минимум четырьмя из пяти программ, характеризуется чувствительностью в 0.9% и максимальным значением PPV – 9%.

Таблица 1. Основные предсказательные характеристики программ TargetScan, PicTar, PITA, RNA22 и miRanda, предсказывающих микроРНК–мРНК взаимодействия

Характеристики	По всем программам, %	Target-Scan, %	PicTar, %	RNA22, %	PITA, %	miRanda, %
Чувствительность	3.9	1.7	1.9	0.9	1.2	2.6
Положительная прогностическая значимость	3.2	5.5	3.6	4.2	6.2	4.3

Для дополнительной оценки предсказательных характеристик программ, результаты предсказаний каждой программы сравнивали со случайно сформированным набором данных. Всего было сформировано пять случайных наборов данных, в которых данные CLASH о микроРНК–мРНК были заменены случайным образом. В результате доля верно предсказанных микроРНК–мРНК взаимодействий, полученных при сравнении со случайными наборами данных, оказалась на порядок меньше, чем при сравнении с данными CLASH: чувствительность и PPV составили 0.41% вместо 3.9% и 0.34% вместо 3.2% соответственно.

Таким образом, на основании сравнения результатов программ по предсказанию микроРНК–мРНК взаимодействий с экспериментальными данными, нами была выявлена низкая эффективность программ. Была показана необходимость использования экспериментально выявленных микроРНК–мРНК взаимодействий для разработки программ нового поколения.

Экспрессионный анализ микроРНК–мРНК интерактома

На следующем этапе мы детально посмотрели на характер взаимодействующих мРНК и микроРНК, учитывая уровень их экспрессии. Для этого дополнительно были использованы публичные экспериментальные данные об уровне экспрессии мРНК и микроРНК в клеточных линиях HEK293 и Huh7.5. Был проведен комплексный анализ уровня экспрессий микроРНК и мРНК с количеством образующихся взаимодействий в каждой из клеточных линий: HEK293 (данные о взаимодействиях – CLASH) и Huh7.5 (данные о взаимодействиях – CLEAR-CLIP).

Анализ экспрессии мРНК

Для анализа уровня экспрессии мРНК нами были обработаны одинаковым образом публичные данные об уровне экспрессии в клеточной линии HEK293 (GSE68611) и Huh7.5 (GSE64677), содержащие по две биологические повторности. Было выявлено схожее количество

экспрессирующихся генов: 15,8 тысяч в клеточной линии HEK293 и 14,5 тысяч в клеточной линии Huh7.5. При этом, в обеих клеточных линиях примерно 6,9 тысяч генов взаимодействовали с одной или более микроРНК согласно данным CLASH и CLEAR-CLIP. Нами был выявлен схожий профиль распределения количества экспрессирующихся генов по количеству взаимодействий с микроРНК (Рисунок 6АБ). Так большая часть генов (57% и 59%) не взаимодействовали с микроРНК, и только 1-2% генов взаимодействовали с большим количеством микроРНК. Сравнение данных генов, образующих наибольшее количество взаимодействий с микроРНК (более девяти взаимодействий), между клеточными линиями выявило пересечение в 18 генов. Эти общие 18 генов образовывали в среднем $15,7 \pm 3,2$ и $14,1 \pm 2,4$ взаимодействий с микроРНК в клеточных линиях HEK293 и Huh7,5 соответственно. При этом взаимодействующие с данными генами микроРНК были различными в клеточных линиях. Анализ функциональной принадлежности данных генов с помощью программы Panther выявил обогащение только в одной генной онтологии – «молекулярная функция связывания РНК» («molecular function of RNA binding») FC 10,6 и p-value $9e10-11$.

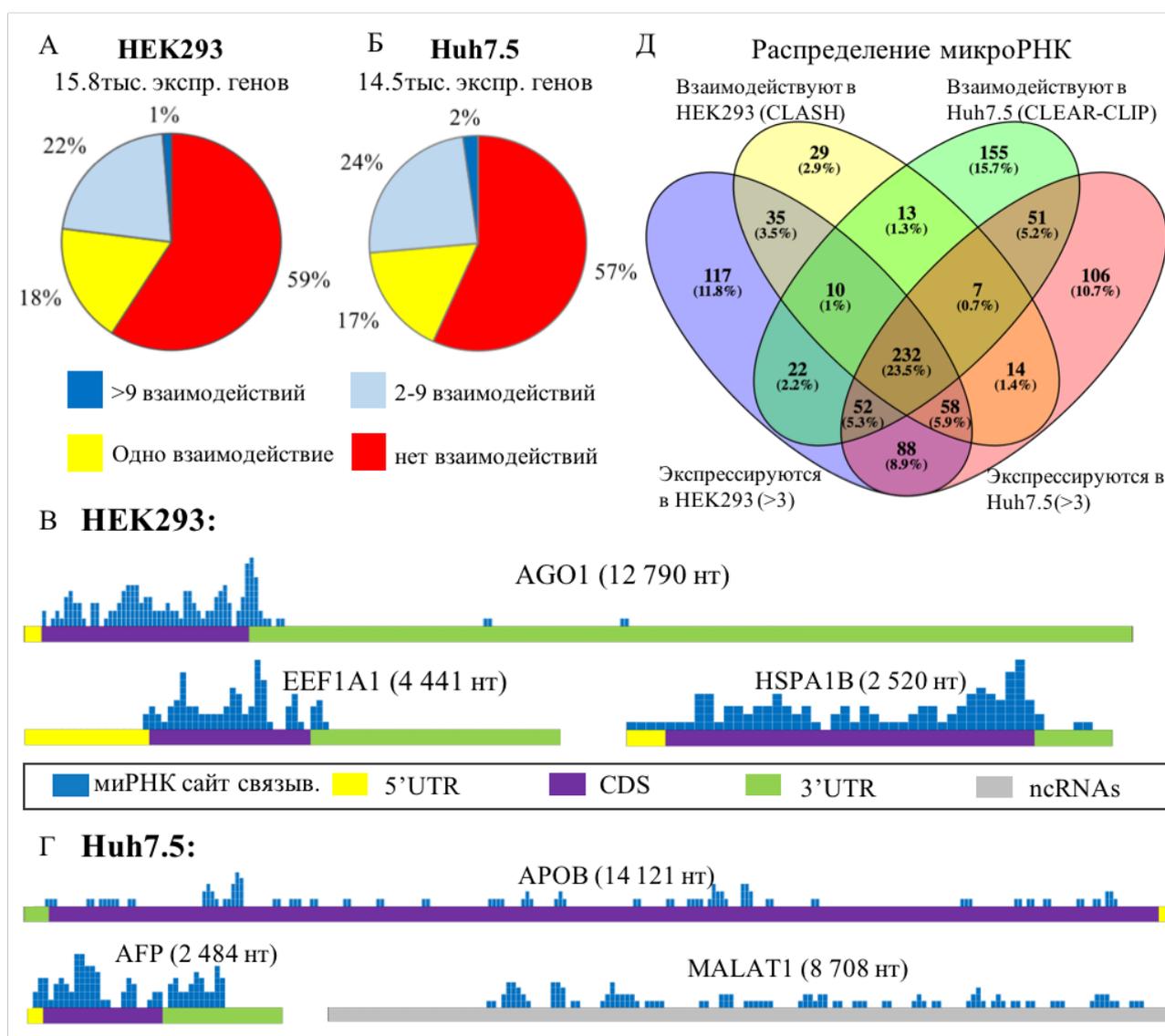


Рисунок 6. Анализ уровня экспрессии мРНК и их взаимодействий с микроРНК. Распределение экспрессирующихся генов по количеству их взаимодействий с микроРНК в клеточной линии НЕК293(А) и Huh7.5(Б). (В) и (Г): локализация сайтов связывания микроРНК (exр-miBRs) в схематически изображенных выявленных мРНК, проявляющих эффект губки в клеточной линии НЕК293 (В) и Huh7.5 (Г). Каждая мРНК сегментирована на регионы в 50нт, при наличии exр-miBR регион отмечен синим, количество отмеченных сегментов соответствует количеству exр-miBRs. Цветом указаны регионы мРНК, имя и длина указаны над мРНК. (Д) Диаграмма Вена, показывающая пересечение между экспрессирующимися микроРНК и взаимодействующими с мРНК микроРНК в клеточных линиях НЕК293 и Huh7.5.

Нами была выявлена группа мРНК, образующая взаимодействия с большим количеством микроРНК, при этом взаимодействия были образованы с различными микроРНК, без выраженного предпочтения к определенной микроРНК. Данные взаимодействия происходили по всей длине мРНК, включая 3'-нетранслируемые области, кодирующую и 5'-нетранслируемую. Ранее аналогичное невыраженное взаимодействие с большим количеством различных микроРНК было описано как эффект “губки микроРНК” (sponge properties) для циркулирующих РНК (circ-

RNA) и для длинных некодирующих РНК (lncRNA) (Hansen T.B. 2013). Для клеточной линии НЕК293 были выявлены мРНК, проявляющие эффект губки (имеющие более 50 взаимодействий с микроРНК), которые кодируются следующими генами: *AGO1*, *EEF1A1*, *HSPA1B* (Рисунок 6В). В то время как в клеточной линии Нuh7.5 были идентифицированы другие гены: *APOB*, *AFP*, *MALAT1*, *XIST* (Рисунок 6Г). Интересно, что в обеих клеточных линиях взаимодействия были идентифицированы во всех регионах мРНК, при этом большая часть взаимодействий была локализована в кодирующей области (CDS) (Рисунок 6ВГ).

Самая взаимодействующая мРНК в клеточной линии НЕК293 согласно данным CLASH, кодируется геном *AGO1* и образует 88 взаимодействий с 50 различными микроРНК. Ген *AGO1* из семейства аргонавтов участвует в синтезе микроРНК, однако для его генов-паралогов (*AGO2*, *AGO3*, *AGO4*) эффект губки не наблюдался, не смотря на высокий процент гомологии нуклеотидной и аминокислотной последовательностей. Это может объясняться тем, что согласно протоколу экспериментального метода CLASH – одним из первых этапов является искусственное повышение уровня экспрессии *AGO1*. Не специфичное свойство «губки» в клеточной линии также наблюдалось для мРНК, кодируемых генами *HSPA1B* (из семейства белков теплового шока) и *EEF1A1* (изоформа 1 фактора элонгации трансляции). Для *HSPA1B* было выявлено 77 взаимодействий с 41 различной микроРНК, а для *EEF1A1* – 50 взаимодействий с 42 микроРНК. При этом *EEF1A1* имел высокий уровень экспрессии в клеточной линии НЕК293 (более 19 тысяч TPM), в то время как *HSPA1B* показывал невысокий уровень экспрессии (775 tpm).

В клеточной линии Нuh7.5 наблюдался другой набор генов, чьи мРНК имели большое количество взаимодействий с микроРНК. Интересно, что помимо двух белок-кодирующих генов: *AFP* (47 взаимодействий с 32 различными микроРНК) и *APOB* (также 47 взаимодействий с 32 различными микроРНК), были идентифицированы две известные длинные некодирующие РНК: *MALAT1* (47 взаимодействий с 27 микроРНК) и *XIST* (55 взаимодействий с 31 микроРНК). Их уровень экспрессии в клеточной линии Нuh7.5 также был различным: *AFP* – более 19 тысяч tpm, *APOB* – 358 tpm, *XIST* – 202 tpm, *MALAT1* – 80 tpm, в то время как средний уровень экспрессии генов составил 69 tpm. Стоит отметить, что данные белок-кодирующие гены являются важными для ткани печени, из которой происходит клеточная линия Нuh7.5. Так, гены альфа-фетопротейн *AFP* (Parpart S. 2014) и аполипротеин *APOB* (Vi Y. 2014) были ранее описаны как биомаркеры онкогенеза печени.

Сравнительный анализ уровней экспрессии микроРНК и их участия в мРНК-взаимодействиях

Для оценки роли микроРНК в регуляции мРНК, были использованы данные об уровне экспрессии микроРНК в клеточных линиях HEK293 (два эксперимента из GSE75136) и Huh7.5 (три эксперимента из GSE74014). Была показана высокая корреляция Пирсона ($r \gg 0,99$) данных уровней экспрессии микроРНК между экспериментами внутри каждой клеточной линии, что говорит о хорошем качестве экспериментальных данных. Данные об уровне экспрессий были проанализированы по единому схеме и отнормированы для сравнения между собой. Суммарно было выявлено 989 экспрессирующихся микроРНК, при этом только четверть из них (23%) присутствовали в обеих клеточных линиях (Рисунок 6Д) Были выявлены микроРНК, которые экспрессировались в клеточных линиях (205 микроРНК в HEK293 и 194 микроРНК в Huh7.5), но их взаимодействия не были выявлены в данных CLASH или CLEAR-CLIP, соответственно.

С другой стороны, часть микроРНК (4-17%) была выявлена в данных CLASH и CLEAR-CLIP и не была идентифицирована как экспрессирующаяся согласно данным RNA-Seq. Эти микроРНК образовывали малое количество взаимодействий: 197 таких микроРНК в данных CLASH имели в среднем $2,2 \pm 0,6$ взаимодействия, 168 микроРНК в данных CLEAR-CLIP в среднем имели $5,1 \pm 2,2$ взаимодействия. В то время, как среднее количество взаимодействий на микроРНК составляет $55,8 \pm 12,7$ для данных CLASH в клеточной линии HEK293 и $143,5 \pm 28,5$ для данных CLEAR-CLIP в клеточной линии Huh7.5. Вероятно, данные микроРНК не были идентифицированы как экспрессирующиеся в виду низкого уровня экспрессии или ввиду наличия защиты от деградации в виде, например, комплекса RISC.

На следующем этапе, мы проанализировали связь между уровнем экспрессии микроРНК и количеством взаимодействий с мРНК. Для этого для каждой клеточной линии были оставлены микроРНК, которые были идентифицированы как экспрессирующиеся в данной клеточной линии и как взаимодействующие: 335 микроРНК в HEK293 и 342 микроРНК в Huh7.5. Для обеих клеточных линий коэффициент корреляции Спирмена был низким – 0,18 и 0,29.

Далее, для каждой микроРНК была вычислена удельная характеристика (R), равная отношению уровня экспрессии к количеству взаимодействий в каждой из клеточных линий. Это позволило охарактеризовать две группы микроРНК, которые наблюдались независимо для каждой клеточной линии. Первая группа включала в себя микроРНК с высоким уровнем экспрессии и относительно небольшим количеством взаимодействий, соответствующие уровню R выше 90% перцентиля, а именно для клеточной линии HEK293 16 микроРНК, у которых экспрессия выше 4 418 и соотношение больше 252; для клеточной линии Huh7.5 12 микроРНК с экспрессией больше 6 941 и с коэффициентом больше 209. Восемь микроРНК первого типа были

общими, таким образом в обеих клеточных линиях данные микроРНК имели высокий уровень экспрессии и почти не взаимодействовали с мРНК, согласно данным CLASH и CLEAR-CLIP.

Другая группа микроРНК, тип 2, характеризовалась низким соотношением уровня экспрессии к количеству взаимодействий (R) и при этом большим количеством взаимодействий. В данную группу, определяемую по нижнему 10% перцентилю по R и верхнему 90% перцентилю по количеству взаимодействий вошло 11 микроРНК для клеточной линии HEK293 (абсолютное количество взаимодействий больше 150) и 6 микроРНК для клеточной линии Huh7.5 (абсолютное количество взаимодействий больше 165). Интересно, что в отличие от микроРНК первого типа, которые сильно пересекались между клеточными линиями, микроРНК второго типа были различными для клеточных линий.

Для определения возможной связи данных двух групп микроРНК с эволюционным отбором, для каждой микроРНК был подсчитан коэффициент консервативности (phyloP score) по 20 позвоночным организмам. Коэффициенты консервативности микроРНК входящих в группу 1 и в группу 2 были схожими: 0,99 и 0,95 соответственно. Интересно, что консервативность данных микроРНК была сильно выше среднего уровня консервативности по всем известным микроРНК (0,24). При этом средний коэффициент консервативности микроРНК, идентифицированных как экспрессирующиеся или взаимодействующие с мРНК равнялся 0,74 для клеточной линии HEK293 и 0,71 – для Huh7.5. Таким образом, было показано, что последовательности микроРНК экспрессирующиеся или связывающиеся с мРНК, были в среднем в два раза более консервативными, что может косвенно указывать на функциональную важность данных микроРНК.

Сравнение характеристик микроРНК между типами клеточных линий

На следующем этапе было проведено сравнение характеристик микроРНК между клеточными линиями. Большинство микроРНК имели прямую зависимость уровня экспрессии и количества взаимодействий между клеточными линиями: т.е. в случае, если уровень экспрессии микроРНК был выше в одной клеточной линии, то данная микроРНК образовывала больше взаимодействий с мРНК в этой клеточной линии, по сравнению с другой клеточной линией (Рисунок 7А).

Типы микроРНК		Кл. линия HEK293		Кл. линия Huh7.5	
		Экспр. микроРНК	Кол-во взаим.	Экспр. микроРНК	Кол-во взаим.
А микроРНК со схожим уровнем экспрессии и кол-вом взаимодействий между кл. линиями					
Б 30 микроРНК, сохраняющие отношение уровня экспр. и кол-ва взаимодействий	N=18				
	N=12				
В Обратное соотношение уровня экспр. и кол-ва взаим.	N=2				
	N=2				

Рисунок 7. Классификация микроРНК согласно изменению характеристик в контексте клеточных линий. Большинство микроРНК имели согласованное изменение уровня экспрессии и количества взаимодействий между клеточными линиями (А). 30 микроРНК показывали явную специфичность для одной из клеточных линий (Б) и 4 микроРНК имели не согласующиеся характеристики между клеточными линиями (В). Обозначения: В колонке «экспр. микроРНК» схематически изображен уровень экспрессии микроРНК, а в колонке «Кол-во взаим.» схематически показано количество взаимодействий.

Однако, были выявлены две небольшие группы микроРНК, демонстрирующие клеточную специфичность. Так 18 микроРНК имели высокий уровень экспрессии и большое количество взаимодействий в клеточной линии Huh7.5 по сравнению с данными характеристиками в клеточной линии HEK293, и наоборот, 12 микроРНК имели повышенный уровень экспрессии и количество взаимодействий в HEK293 по сравнению с Huh7.5 (Рисунок 7Б). Например, микроРНК hsa-miR-194-5p, описанная ранее, как репрессор MAPK1 пути (Kong Q. 2018) имела уровень экспрессии в клеточной линии Huh7.5 в 89 раз выше, чем в HEK293, а согласно данным CLEAR-CLIP количество взаимодействий равнялось 336, в то время как в клеточной линии HEK293 – было выявлено всего лишь 7 взаимодействий для данной микроРНК. С другой стороны, микроРНК hsa-miR-10a-5p, подавляющая активность ланостеролсинтазы (Kim J.E. 2018), экспрессировалась в 450 раз выше в клеточной линии HEK293, чем в Huh7.5, образуя 267 взаимодействий в HEK293 и лишь 8 взаимодействий в клеточной линии Huh7.5. Такие микроРНК полностью согласуются с предположением, что чем выше уровень экспрессии микроРНК, тем большее количество взаимодействий образует данная микроРНК.

В каждой клеточной линии было выявлено по две микроРНК, которые не подчинялись данному правилу, и даже наоборот - такие микроРНК образовывали на порядок больше взаимодействий в клеточной линии, где уровень экспрессии был намного меньше (Рисунок 7В). Например, микроРНК hsa-miR-331-3p при высоком уровне экспрессии в клеточной линии Huh7.5, почти не образует в ней взаимодействий, а в клеточной линии НЕК293, где данная микроРНК почти не экспрессируется – образует больше 300 взаимодействий (Huh7.5: уровень экспрессии 1030, количество взаимодействий – 47; НЕК293: уровень экспрессии 65, количество взаимодействий – 342). Другим примером может быть микроРНК hsa-miR-100-5p: в Huh7.5 – уровень экспрессии 916, количество взаимодействий – 1; НЕК293 – уровень экспрессии 41, количество взаимодействий – 30.

Таким образом, нами был проведен сравнительный анализ уровня экспрессий и количества взаимодействий для мРНК и для микроРНК на клеточных линиях человека НЕК293 и Huh7.5. Для обеих клеточных линий показано, что только 1-2% экспрессирующихся генов взаимодействовали с большим количеством микроРНК, в то время как 57-59% не имели взаимодействий с микроРНК. Были выявлены ткане-специфичные гены, проявляющие эффект губки – мРНК данных генов взаимодействуют с большим количеством различных микроРНК. Экспрессионный анализ микроРНК выявил различие по клеточным линиям, а также показал, что только 24% микроРНК экспрессируются и связываются с мРНК в обеих изучаемых клеточных линиях. Были выявлены две интересные группы микроРНК, различающиеся по уровню экспрессии и количеству взаимодействий.

Создание программы для поиска вариантов нуклеотидной последовательности в сайтах связывания с микроРНК

Программа для анализа вариантов нуклеотидной последовательности ДНК (exp-miBR Annotator)

Для анализа нуклеотидных вариантов генома человека, выявленные регионы высоко-достоверных экспериментально подтвержденных сайтов связывания с микроРНК (exp-miBR), были оформлены в веб-программу “exp-miBR Annotator”, доступную по адресу: <http://score.generesearch.ru/services/mirna/>. Данная программа позволяет по входным координатам нуклеотидной последовательности ДНК человека определять их нахождение в регионах, достоверно взаимодействующих с микроРНК (в exp-miBR).

Для каждого exp-miBR в базе данных содержатся все собранные характеристики. Названия параметров указаны в скобках:

- **Информация о регионе ДНК:** направление нити ДНК +/- (**MIRNA_STRAND**), имя гена (**MIRNA_GENES**), тип региона РНК (**MIRNA_REGION**).
- **Информация об экспериментах,** где выявлен данный exp-miBR: количество экспериментов (**MIRNA_NDATASET**), их перечень в виде идентификаторов с базы данных NCBI GEO (**MIRNA_DATASET**), а также перечень клеточных линий (**MIRNA_LINES**) и их количество (**MIRNA_NLINES**).
- **Информация о взаимодействующих микроРНК,** в случае если они были идентифицированы: перечень взаимодействующих микроРНК (**MIRNA**).

Программа может работать в двух модулях: анализировать единичную замену по координате (модуль – “Process a single position”) или анализировать файл, содержащий набор позиций (модуль – “Process a VCF query”). В обоих модулях программа поддерживает работу с вариантами, проаннотированными на референсный геном человека версии hg19 и версии 38, используемую версию необходимо указать перед запуском программы. В случае анализа файла, программе на вход необходимо подать файл формата VCF – “variant calling format”. Данный формат был разработан международным консорциумом по итогам проекта 1000 геномов и широко используется для описания и аннотирования вариантов нуклеотидной последовательности. На входной файл установлены ограничения: поддерживается файл последней версии VCF формата (4.0-4.2), размер - не более 20Мб (файл может быть сжат с помощью gzip или bgzip), не более 500 тысяч строк (не включая строки заголовка (“header”) VCF файла. На выход программа выдает сжатый bgzipped VCF файл.

Так как генетические данные в медицинской практике считаются персональными данными, нами дополнительно была выложена база данных exp-miBR в виде отдельного трека (раздел “Download Data”). В данном разделе пользователи могут скачать данные exp-miBR для референсной сборки генома человека версии hg19 или версии 38 и имплементировать в свой анализ для аннотации вариантов нуклеотидной последовательности. Так например, в январе 2020 года данный трек был скачен и использован в коммерческой программе TRANSFAC® 2020.1 компании geneXplain (<http://genexplain.com>) - http://genexplain.com/wp-content/uploads/2020/01/TRANSFAC_release.pdf

Таким образом, нами была разработана программа для поиска вариантов нуклеотидной последовательности в сайтах связывания с микроРНК.

Поиск функционально значимых вариантов нуклеотидной последовательности, нарушающих микроРНК-мРНК взаимодействие.

Роль микроРНК в патогенезе наследственных заболеваний

Созданный в начале 1960-х годов каталог наследственных заболеваний и расстройств человека “ОМIM” (<https://www.omim.org/>) на данный момент содержит описание около 7000 генетических заболеваний и около 15 тысяч генов. Однако, до сих пор примерно 60% случаев диагностики наследственных заболеваний остаются без установленной патогенетической причины, что отчасти объясняется сложными механизмами регуляции экспрессии генов. Одним из механизмов патогенеза наследственных заболеваний может являться нарушение экспрессии генов за счёт микроРНК. Ранее были опубликованы работы, описывающие роль микроРНК в патогенезе онкологических заболеваний (Lin S. 2015), аутоиммунных заболеваний (de Almeida R.C. 2018), нейродегенеративных болезней (Ferrante M. 2017) и многих других. Изучение роли микроРНК в патогенезе заболеваний важно не только для понимания механизмов развития заболеваний, но и помогает в разработке терапевтических подходов.

Нами был проведен анализ литературных источников и показано как нарушение регуляции микроРНК может приводить и играть роль в развитии таких наследственных заболеваниях, как: муковисцидоз, миодистрофия Дюшена, бета-талассемия, глаукома, лице-лопаточно-плечевая миодистрофия Ландузи-Дежерина, болезнь Гиршпрунга, синдром Ретта, синдром Туретта, Пемфигус (болезнь хейли-хейли). В описанных случаях патогенез заболеваний связан либо с потерей сайта связывания микроРНК, либо с его усилением, либо с изменением последовательности самой микроРНК.

Но случаев известных механизмов патогенеза наследственных заболеваний с участием микроРНК до сих пор мало ввиду сложности их определения. Для нахождения таких новых механизмов мы применили разработанную нами программу exp-miBR Annotator на публичных данных связи фенотипов и вариантов нуклеотидной последовательности, а также выбрали дополнительные функционально значимые гены и провели экспериментальную проверку влияния внесения мутаций в сайт связывания микроРНК на регуляцию экспрессии мРНК.

Анализ описанных патогенных вариантов способных нарушать микроРНК-мРНК взаимодействие

На первом этапе нами была проведена оценка количества вариантов нуклеотидной последовательности человека, находящихся в регионах связывания микроРНК (exp-miBR) и описанных как патогенные варианты при наследственных заболеваниях, но при этом имеющие неизвестный механизм патогенеза. Для этого была использована база данных ClinVar

(<http://www.clinvar.com>), содержащая описанные связи вариантов нуклеотидной последовательности и их фенотипических проявлений. Из 429 тысяч вариантов, описанных как патогенные в ClinVar для одного или нескольких фенотипов, 961 вариант был выявлен в высокодостоверных сайтах связывания микроРНК (exp-miBRs). 95% выявленных вариантов (916 из 961) были расположены в кодирующей области генов и только 6 вариантов были локализованы в 3' нетранслируемой области. При этом только 24% из 46 805 exp-miBRs находятся в кодирующей области и примерно половина exp-miBRs локализована в 3' нетранслируемых областях.

Для 3 из 6 вариантов, идентифицированных в 3' нетранслируемой области, - rs387906334, rs35717904, rs878854355 не было описано взаимодействующей микроРНК. При этом для патогенного варианта rs35717904, находящегося в 3' нетранслируемой области гена *FOXC1* (NM_001453:c.*734A>T) был ранее описан механизм патогенеза глаукомы с участием микроРНК miR-548l (Medina-Trillo C. 2015). Данный вариант был выявлен у двух пробандов: у одного пробанда дополнительно к данному варианту был описан вариант p.I126S в гене *FOXC1* в цис-положении, другой неродственный ему пробанд дополнительно имел инсерцию p.G447_G448insDG также в цис-положении. С помощью люциферазных конструкций на клеточной линии НЕК-293Т было показано, что экспрессия miR-548l приводит к снижению активности люциферазы в конструкции дикого типа. Наличие патогенного варианта с.*734A>T в дополнении к каждому из двух вариантов в кодирующей области (p.I126S и p.G447_G448insDG) приводило к увеличению уровня экспрессии мутантной РНК и белка. Таким образом было показано, что данный некодирующий патогенный вариант повышает стабильность мРНК *FOXC1* за счёт разрушения сайта связывания с miR-548l (Medina-Trillo C. 2015).

Среди трёх оставшихся вариантов, находящиеся в 3' нетранслируемой области с известной микроРНК, у одного варианта rs150586939 в гене *IER3IP1*, частота встречаемости в здоровой популяции была равна 2%. Скорее всего, на основании этого, данный вариант позднее был дополнительно проаннотирован в ClinVar как вероятно не патогенный. Оставшиеся два варианта были расположены рядом друг с другом в 3' нетранслируемой области гена *TMEM107* (на расстоянии 4 нт) и вне кодирующей области гена *SNORD118*: вариант rs201787275 и rs75008470 (g.8076766G>C). Ген *SNORD118* был описан как единственный известный ген (согласно базе данных OMIM) ассоциированный с лейкоэнцефалопатией с церебральной кальцификацией и кистами (известного также как синдром Лабруна, OMIM614561). Синдром Лабруна является редким наследственным заболеванием, аутосомно-рецессивного типа наследования. При низкой популяционной частоте, данные два варианта были выявлены и описаны у 6 и 8 пробандов с синдромом Лабруна, часть пробандов были неродственными (Jenkinson 2016). Варианты rs201787275 и rs75008470 представляют особый интерес, так как потенциально могут нарушать

сайт связывания с микроРНК. Эти два варианта в гене SNORD118 были отобраны нами для дальнейшей экспериментальной проверки *in-vitro*.

Из 916 выявленных кодирующих вариантов, локализованных в *exr-miBR*, большая часть (47%) приводила к образованию стоп-кодона, за счет нонсенс-вариантов или однонуклеотидной инсерции-делеции (Рисунок 8). Для другой половины исследуемых вариантов была определена как минимум одна связывающаяся микроРНК с *exr-miBR*, содержащим данный патогенный вариант, что выше ожидаемого - только для 17% из 46 тысяч *exr-miBRs* есть экспериментально установленная микроРНК (или несколько микроРНК).



Рисунок 8. Характеристики выявленных патогенных вариантов ClinVar в *exr-miBRs*.
Распределение 916 патогенных вариантов, выявленных в кодирующих регионах.

Дальнейший анализ был сфокусирован на несинонимичных однонуклеотидных вариантах и делециях, не меняющих рамку считывания, для которых была выявлена микроРНК, связывающаяся с *exr-miBR*. Исследуемые 169 вариантов были расположены в 86 различных генах. Интересно, что 13 патогенных вариантов были ко-локализованы друг с другом - расстояние между ними было меньше длины микроРНК (22 нуклеотидов).

Для изучения возможного участия патогенных вариантов в патогенезе наследственных заболеваний, нами были вручную проанализированы выявленные варианты и ассоциированные с ними заболевания. Во внимание принимались следующие факторы: уровень экспрессии гена, содержащего патогенный вариант в ткани, где проявляется данное заболевание и уровень экспрессии данного гена в других тканях. Для 25 самых интересных несинонимичных вариантов были вручную проверены известные публикации, описывающие связь варианта с патогенезом заболевания. Было найдено, что для большинства данных вариантов были опубликованы эксперименты, результаты которых исключают роль микроРНК в патогенезе, а именно: показано сохранение уровня экспрессии мутантного белка в сравнении с уровнем экспрессии белка дикого типа; показана возможная роль варианта при сплайсинге; наличие варианта в важном участке

связывания белка. Дополнительным фактором при анализе вариантов также являлся высокий уровень консервативности варианта между различными позвоночными.

Был выявлен один интересный вариант rs121918680, приводящий к замене 394 глутаминовой кислоты на лизин в кодирующей области гена трансферина (*TF*). Трансферин преимущественно синтезируется в печени, где он играет важную роль при транспорте ионов железа. Данный вариант rs121918680 описан как патогенный для атрансферринемии – редкого наследственного гематологического заболевания, характеризующегося недостатком трансферина и микроцитарной гипохромной анемией. Вариант rs121918680 был описан в 2002 году, но экспериментальной проверки функциональной значимости данного варианта у пациентов проведено не было, из-за трудностей с получением образцов РНК из тканей, где экспрессируется *TF*, таких как печень или мозг (Asada-Senju M. 2002). Согласно данным *exp-miBR*, с данным вариантом взаимодействует одна из трех микроРНК miR-33b-5p, miR-194-5p, miR-93-3p выявленных экспериментальным методом CLEAR-CLIP на клеточной линии Huh7.5, также полученной из ткани печени. Интересно, что микроРНК miR-33b-5p преимущественно экспрессируется в тканях мозга, а miR-194-5p – исключительно в тканях толстой кишки и в печени. Данный вариант в гене *TF* также был отобран для дальнейшей экспериментальной проверки *in-vitro*.

Экспериментальное изучение влияния патогенных вариантов в *exp-miBR* на регуляцию экспрессии гена-мишени

На последнем этапе нами была проведена экспериментальная проверка влияния патогенных вариантов на уровень экспрессии РНК на модели *in-vitro*. Для этого сначала было отобрано микроРНК-мРНК взаимодействие из базы данных *exp-miBRs* для экспериментальной проверки влияния нарушения сайта связывания с микроРНК на регуляцию экспрессии мРНК. Особый интерес представляет изучение регуляции генов, играющих важную роль в ключевых процессах клетки, нарушение которых может приводить к онкогенезу. Было предположено, что в норме уровень экспрессии онкогена регулируется посредством микроРНК, а внесение мутации в сайт связывания может приводить к увеличению уровня экспрессии мутантного онкогена, и следовательно вносить свой вклад в развитие опухолевого клона.

Для выявления таких взаимодействий между микроРНК и онкогенами, нами был взят консенсусный список онкогенов и генов супрессоров опухолевого роста (COSMIC), опубликованный в *Nature* в 2018 году (Sondka Z. 2018). Из 576 таких генов, половина (280) проявляет онкогенные свойства. Для данных онкогенов был проведен поиск *exp-miBR*, для последующей экспериментальной проверки. Так как экспериментальная проверка планировалась

на клеточной линии HEK293, были оставлены *exp-miBRs*, встречающиеся в клеточной линии HEK293, а длина выбираемых *exp-miBR* была ограничена в 100нт. По данным требованиям было выявлено 148 *exp-miBRs*. Из них были оставлены только *exp-miBRs*, взаимодействующие с высоко-экспрессируемыми в клеточной линии HEK293 микроРНК, это позволило провести эксперимент без искусственного увеличения уровня экспрессии микроРНК. Таким образом, было отобрано взаимодействие между геном CDK6 и микроРНК *let-7a-5p*, образующееся в 3' нетранслируемой области мРНК (согласно 38 сборке генома человека *exp-miBR* расположен в регионе chr7:92606406-92606489). Согласно базе данных *exp-miBR* данный регион мРНК гена CDK6 взаимодействовал с микроРНК в нескольких клеточных линиях человека (GSE32113; GSE41437). Также экспериментально проверялось влияние описанных в ClinVar патогенных вариантов rs201787275 и rs75008470 на регион гена SNORD118 и варианта rs121918680 в гене TF.

В лаборатории Функциональной геномики в Медико-генетическом научном центре имени академика Н.П. Бочкова был проведен эксперимент по изучению влияния данных вариантов нуклеотидной последовательности на уровень экспрессии мРНК. Для экспериментальной проверки использовалась плазмидная конструкция psiCHECK2, содержащая репортерные гены, кодирующие люциферазы Firefly и Renilla. Встраивание изучаемого локуса происходит под промотор второго репортерного гена. Таким образом, уровень активности первого репортерного гена говорит об уровне эффективности трансфекции, а второй репортерный ген – об активности изучаемого региона. Полученный уровень экспрессии второго репортерного гена нормируется на уровень экспрессии первого репортерного гена, таким образом учитывается эффективность трансфекции в разных экспериментах. Для каждого эксперимента создавалась отдельно конструкция psiCHECK2 с регионом гена дикого типа (wild type – WT) и с регионом гена, содержащим изучаемые варианты. Для гена SNORD118 было сделано две конструкции с мутациями (одна конструкция содержала rs201787275 и другая - rs75008470).

Структура взаимодействия мРНК CDK6 и микроРНК *let-7* была предсказана с помощью программы RNAhybrid (Рисунок 20А). Внесенная динуклеотидная мутация CT → GA находилась в середине «seed»-региона, отвечающего за связывание мРНК и микроРНК. Введение такой динуклеотидной замены должно приводить к полному или частичному нарушению сайта связывания микроРНК *hsa-let7-5p* с 3' нетранслируемой областью гена CDK6. Для генов TF и SNORD118 сайт связывания с микроРНК был также определен с помощью программы RNAhybrid, показавшей локализацию изучаемых патогенных вариантов в местах связывания с микроРНК (Рисунок 20ББ).

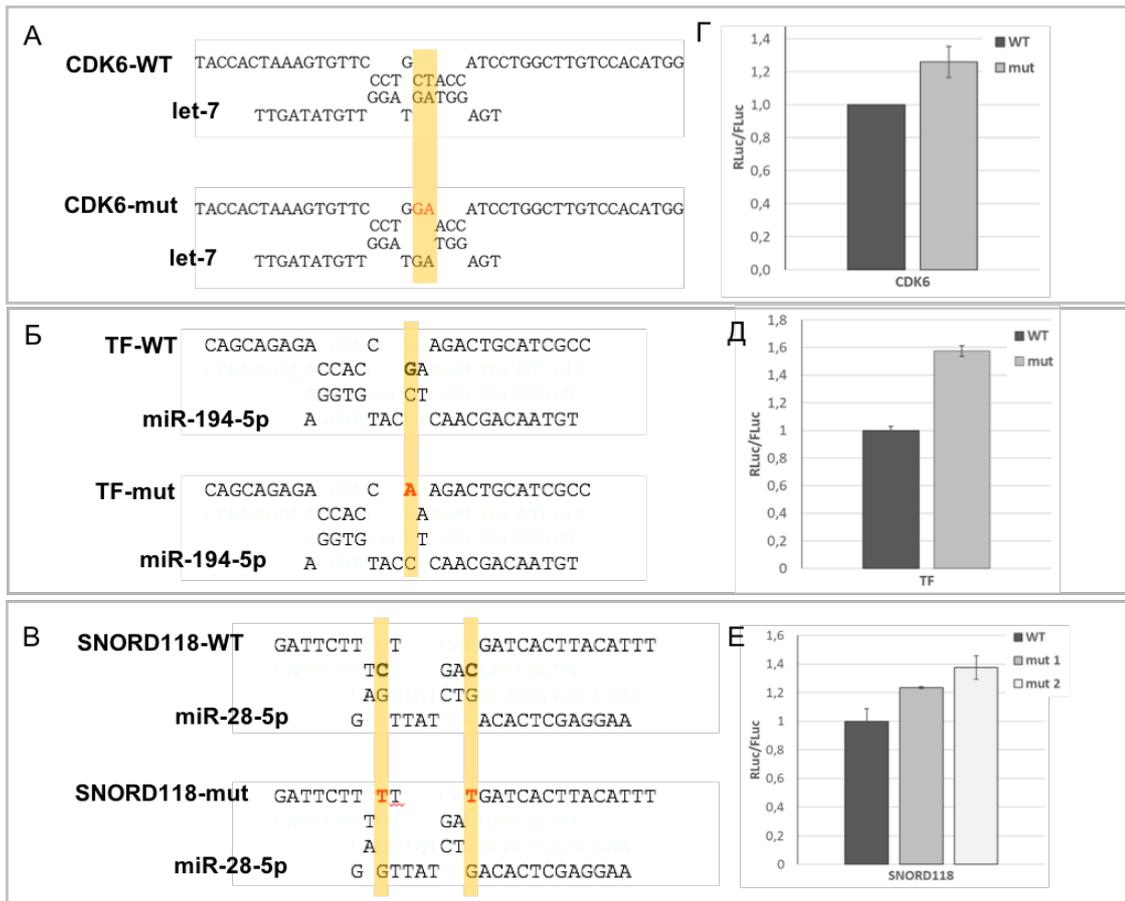


Рисунок 20. *In-vitro* анализ влияния внесения мутаций в сайт связывания микроРНК на уровень экспрессии мРНК. Схема внесенных мутаций и их локализация в сайте связывания CDK6 и микроРНК let-7-5p. (А); сайте связывания TF – miR-194-5p (Б) и SNORD118-miR-28-5p (В). Уровень экспрессии конструкции с мРНК гена CDK6 дикого типа (черный) и с внесенными мутациями (серый) (Г). Уровень экспрессии конструкции с мРНК гена TF дикого типа (черный) и с внесенной мутацией rs121918680 (серый) (Д). Уровень экспрессии конструкции с мРНК гена SNORD118 дикого типа (wt - черный) и с внесенными мутациями: rs75008470 (mut1- серый) и rs201787275 (mut2- белый) (Е).

Проведенные два независимых эксперимента, каждый по 4 повторности, показал, что данная динуклеотидная мутация значимо (p -value 0,0006) увеличивает экспрессию конструкции с мутантным регионом гена CDK6 в сравнении с конструкцией с диким вариантом гена CDK6 (Рисунок 20Г). Это подтверждает, что внесение данной мутации нарушает сайт связывания и приводит к увеличению экспрессии конструкции с мутантной мРНК гена CDK6.

Интересно, что ранее был экспериментально подтвержден другой сайт связывания CDK6 с данной микроРНК (Zhang Z. 2014). Данный эксперимент проводился с искусственным увеличением экспрессии микроРНК. Кроме того, в описанной ранее публикации удалялся весь сайт связывания. Нами было показано, что даже внесение динуклеотидной мутации способно частично нарушить сайт связывания и привести к значительному увеличению уровня экспрессии гена CDK6.

Внесение патогенного варианта rs121918680, описанного в ClinVar для атрансферринемии, показало увеличение экспрессии конструкции с мутантным регионом гена TF со средним значением 1,58 (Рисунок 20Д). Согласно данным exp-miBR, данный вариант rs121918680 в гене TF находится в области взаимодействия одной из трех микроРНК: miR-33b-5p, miR-194-5p, miR-93-3p выявленных экспериментальным методом CLEAR-CLIP на клеточной линии Huh7.5, также полученной из ткани печени. Интересно, что микроРНК miR-33b-5p преимущественно экспрессируется в тканях мозга, а miR-194-5p – исключительно в тканях толстой кишки и в печени. Показано, что выявленный патогенный вариант rs121918680 находится в регионе связывания с микроРНК miR-194-5p (Рисунок 20Б), а внесение данной мутации в конструкцию (TF-mut) приводит к увеличению экспрессии конструкции.

Два других варианта, описанных в ClinVar в гене SNORD118, были изучены экспериментально в двух разных конструкциях. Было показано, что каждый из двух описанных патогенных вариантов приводит к увеличению уровня экспрессии конструкции с данными регионами гена SNORD118: до значения 1,24 для варианта rs75008470 и до значения 1,37 для варианта rs201787275 (Рисунок 20Е). Согласно данным exp-miBR, изучаемые патогенные варианты входят в регион, взаимодействующий с микроРНК hsa-miR-28-5p и hsa-miR-122-5p.

Таким образом, нами была изучена роль нарушения сайта связывания микроРНК в патогенезе наследственных заболеваний и применимость разработанной программы для поиска вариантов нуклеотидной последовательности в сайтах связывания с микроРНК. Также была показано возможность нарушения регуляции микроРНК-мРНК взаимодействий при внесении мутации в сайт их связывания.

ЗАКЛЮЧЕНИЕ

1. Создана база данных достоверных, экспериментально идентифицированных сайтов связывания микроРНК человека. В неё вошли 46 тысяч регионов мРНК (exp-miBR), полученных в девяти клеточных линиях. Данные регионы были всесторонне охарактеризованы.
2. Сравнительный анализ экспериментально идентифицированных сайтов связывания микроРНК с предсказанными взаимодействиями показал низкий уровень эффективности работы пяти программ, предсказывающих микроРНК-мРНК взаимодействия (TargetScan, PicTar, PITA, RNA22 и Miranda).
3. Проведен сравнительный анализ уровня экспрессий и количества взаимодействий для мРНК и для микроРНК. Выявлены ткане-специфичные мРНК, проявляющие эффект «губки микроРНК». Охарактеризованы две группы микроРНК, отличающиеся по свойствам экспрессии и активности взаимодействий с мРНК.
4. Разработана веб-программа “exp-miBR Annotator”, позволяющая проводить поиск вариантов нуклеотидной последовательности генома человека в сайтах связывания с микроРНК (exp-miBR).
5. Для микроРНК let-7 и онкогена CDK6 из базы данных exp-miBRs было экспериментально подтверждено микроРНК-мРНК взаимодействие. Показано, что внесение мутации в сайт взаимодействия микроРНК с мРНК приводит к изменению экспрессии гена CDK6.
6. С помощью разработанной программы “exp-miBR Annotator” были выявлены известные патогенные варианты нуклеотидной последовательности, находящиеся в сайтах связывания с микроРНК. Экспериментально показано что выявленные патогенные варианты (rs121918680, rs201787275 и rs75008470) в генах TF и SNORD118 приводят к разрушению сайтов связывания микроРНК и к увеличению экспрессии мРНК.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

По материалам диссертации опубликовано две статьи в рецензируемых научных журналах, входящих в основные библиометрические базы данных (РИНЦ, PubMed, WoS и Scopus):

1. **Plotnikova O.**, Baranova A., and Skoblov M. "Comprehensive analysis of human microRNA-mRNA interactome." *Frontiers in Genetics*, 2019, (10):933
2. **Плотникова О.М.**, Скоблов М.Ю. «Эффективность программ, предсказывающих микроРНК-мРНК взаимодействия» («efficiency of the miRNA- mRNA interaction prediction programs») *МОЛЕКУЛЯРНАЯ БИОЛОГИЯ*, 2018, 52(3);1–12
3. **Плотникова О.М.**, Скоблов М.Ю. «Роль микроРНК в патогенезе наследственных заболеваний» *Медицинская генетика* 2020, 19(9):5-17

Работы вышли до смены фамилии (до августа 2020 – Плотникова О.М., после – Кудряшова О.М.)

СПИСОК ЛИТЕРАТУРЫ

Климентова, Е.А., Гилязова, И.Р., Кунсбаева, Г.Б., Измайлов, А.А., Султанов, И. М., Павлов, В. Н., Хуснутдинова, Э. К. «Ассоциация полиморфного варианта сайта связывания микроРНК rs10491534 гена TSC1 с тяжестью течения рака почки.» *Медицинская генетика* 15(4) (2016): 50-52.

Ørom, Ulf Andersson, Finn Cilius Nielsen, and Anders H. Lund. «MicroRNA-10a binds the 5' UTR of ribosomal protein mRNAs and enhances their translation.» *Molecular cell* 30.4 (2008): 460-471.

Agarwal V., Bell G.W., Nam J.W., Bartel D.P. «Predicting effective microRNA target sites in mammalian mRNAs.» *eLife* 4 (2015): e05005.

Annese, T., Tamma, R., De Giorgis, M., Ribatti, D. «microRNAs Biogenesis, Functions and Role in Tumor Angiogenesis.» *Frontiers in Oncology* 10 (2020).

Asada-Senju M., Maeda T., Sakata T., et. al. «Molecular analysis of the transferrin gene in a patient with hereditary hypotransferrinemia.» *Journal of human genetics* 47(7) (2002): 355-9.

Baccarini A., Brown B.D. «Monitoring microRNA activity and validating microRNA targets by reporter-based approaches.» *Methods Mol. Biol.* 667 (2010): 215–233.

Bartel, D.P., «MicroRNAs: target recognition and regulatory functions.» *Cell* 136 (2009): 215–233.

Bi Y., He Y., Huang J., et al. «Functional characteristics of reversibly immortalized hepatic progenitor cells derived from mouse embryonic liver.» *Cell. Physiol. Biochem.* 34(4) (2014): 1318–1338.

Carthew R.W., Sontheimer E.J. «Origins and mechanisms of miRNAs and siRNAs.» *Cell* 136(4) (2009): 642-655.

de Almeida R.C., Chagas V.S., Castro M.A., et al. «Integrative analysis identifies genetic variants associated with autoimmune diseases affecting putative microRNA binding sites.» *Front. Genet.* 9 (2018): 139.

de Almeida, R. C., Chagas, V. S., Castro, M. A., Petzl-Erler, M. L. «Integrative analysis identifies genetic variants associated with autoimmune diseases affecting putative microRNA binding sites.» *Frontiers in Genetics* 9 (2018): 139.

Ferrante M., Conti G. O. «Environment and neurodegenerative diseases: an update on miRNA role.» *Microna* 6(3) (2017): 157-165.

Fitriawan, A. S., Kartika, A. I., Chasanah, S. N., Aryandono, T., Haryana, S. M. «Expression of Circulating MicroRNA-141 in Epithelial Ovarian Cancer.» *The Malaysian Journal of Medical Sciences: MJMS* 27(6) (2020): 27.

Frazier, S. E. «Investigating placental microRNAs in preeclampsia and optimising a gene therapy strategy for targeting the placenta.» *University of Glasgow*, 2021: Diss.

Gao, L., Zhao, Y., Ma, X., Zhang, L. «Integrated analysis of lncRNA–miRNA–mRNA ceRNA network and the potential prognosis indicators in sarcomas.» *BMC Medical Genomics* 14(1) (2021): 1-11.

Hafner M., Landthaler M., Burger L., et al. «PAR-CLIP - A Method to Identify Transcriptome-wide the Binding Sites of RNA Binding Proteins.» *Journal of Visualized Experiments* 41 (2010): e2034.

Hansen T.B., Jensen T.I., Clausen B.H., et al. «Natural RNA circles function as efficient microRNA sponges.» *Nature* 495(7441) (2013): 384-388.

Helwak A., Kudla G., Dudnakova T., Tollervey D. «Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding .» *Cell* 153(3) (2013): 654-665.

Hussen, B.M., Hidayat, H.J., Salihi, A., Sabir, D.K., Taheri, M., Ghafouri-Fard, S. «MicroRNA: A signature for cancer progression.» *Biomedicine & Pharmacotherapy* 138 (2021): 111528.

Jenkinson, E. M., Rodero, M. P., Kasher, P. R., et. al. «Mutations in SNORD118 cause the cerebral microangiopathy leukoencephalopathy with calcifications and cysts.» *Nature genetics* 48(10) (2016): 1185-1192.

John B., Enright A.J., Aravin A., et al. «Human microRNA targets .» *PLoS Biol.* 2 (2004): e363.

Kertesz M., Iovino N., Unnerstall U., et al. «The role of site accessibility in microRNA target recognition .» *Nat. Genet.* 39 (2007): 1278–1284.

Kim J.E., Hong J.W., Lee H.S., et al. «Hsa-miR-10a-5p downregulation in mutant UQCRB-expressing cells promotes the cholesterol biosynthesis pathway.» *Sci. Rep.* 8 (1) (2018): 12407.

Kong Q., Zhang S., Liang C., et al. «LncRNA XIST functions as a molecular sponge of miR-194-5p to regulate MAPK1 expression in hepatocellular carcinoma cell.» *J. Cell. Biochem* 119(6) (2018): 4458–4468.

Krek A., Grün D., Poy M.N., et al. «Combinatorial microRNA target predictions.» *Nat. Genet.* 37 (2005): 495–500.

Lall S., Grün D., Krek A., et al. «A genome-wide map of conserved microRNA targets in *C. elegans*.» *Curr. Biol.* 16 (2006): 460–471.

Li H., Xie S., Liu X., et al. «Matrine alters microRNA expression profiles in SGC-7901 human gastric cancer cells.» *Oncol. Rep.* 32 (2014): 2118–2126.

Lin S., Gregory R.I. «MicroRNA biogenesis pathways in cancer.» *Nature reviews cancer* 15(6) (2015): 321-333.

Medina-Trillo C., Aroca-Aguilar J.D., Ferre-Fernández J.J., et al. «The Role of hsa-miR-548l Dysregulation as a Putative Modifier Factor for Glaucoma-Associated FOXC1 Mutations.» *Microna* 4(1) (2015): 50-56.

Miranda K.C., Huynh T., Tay Y., et al. «A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes.» *Cell* 126 (2006): 1203–1217.

Moore M.J., Scheel T.K., Luna J.M., et al. «miRNA–target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity.» *Nature communications* 6(1) (2015): 1-17.

Najafipour, R., Momeni, A., Yousefipour, F., Mousavi, S., Moghbelinejad, S. «Underexpression of hsa-miR-449 family and their promoter hypermethylation in infertile men: A case-control study.» *International Journal of Reproductive BioMedicine* 19(1) (2021): 23-34.

- Ni, W. J., Leng, X. M. «Dynamic miRNA-mRNA paradigms: new faces of miRNAs.» *Biochem. Biophys. Rep.* 4 (2015): 337–341.
- Parpart S., Roessler S., Dong F., et al. «Modulation of miR-29 expression by α -fetoprotein is linked to the hepatocellular carcinoma epigenome.» *Hepatology* 60(3) (2014): 872–883.
- Plotnikova, O., Baranova, A., Skoblov, M. «Comprehensive analysis of human microRNA–mRNA interactome.» *Frontiers in genetics* 10 (2019): 933.
- Rohani, N., Moughari, F. A., Eslahchi, C. «DisCoVering potential candidates of RNAi-based therapy for COVID-19 using computational methods.» *PeerJ* 9 (2021): e10505.
- Schober, A., Maleki, S. S., Nazari-Jahantigh, M. «Regulatory Non-coding RNAs in Atherosclerosis.» *Handbook of Experimental Pharmacology*, 2020.
- Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., Rajewsky, N. «Widespread changes in protein synthesis induced by microRNAs.» *Nature* 455(7209) (2008): 58-63.
- Sondka Z., Bamford S., Cole C.G., et. al. «The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers.» *Nature Reviews Cancer* 18(11) (2018): 696-705.
- Telford, B.J., Yahyanejad, S., de Gunst, T., et. al. «Multi-modal effects of 1B3, a novel synthetic miR-193a-3p mimic, support strong potential for therapeutic intervention in oncology.» *Oncotarget* 12(5) (2021): 422.
- Ueno, D., Mikami, M., Yamasaki, S., Kaneko, M., Mukuta, T., Demura, T., Kato, K. «Changes in mRNA degradation efficiencies under varying conditions are regulated by multiple determinants in *Arabidopsis thaliana*.» *Plant and Cell Physiology*, 2020.
- Vannan, A., Powell, G. L., Dell'Orco, M., et. al. «microRNA regulation related to the protective effects of environmental enrichment against cocaine-seeking behavior.» *Drug and Alcohol Dependence* 221 (2021): 108585.
- Wang, K., Li, M., and Hakonarson, H. «ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.» *Nucleic Acids Res* 38 (16) (2010): e164.
- Wang, L., Zhou, L., Jiang, P., Lu, L., Chen, X., Lan, H., Guttridge, DC., Sun, H., Wang, H. «Loss of miR-29 in myoblasts contributes to dystrophic muscle pathogenesis.» *Molecular therapy* 20(6) (2012): 1222-1233.
- Wang, X. «Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies.» *Bioinformatics* 32(9) (2016): 1316-1322.
- Wang, X. «Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies.» *Bioinformatics* 32(9) (2016): 1316-1322.
- Wang, Y., Jiang, Q., Chakravarti, A., Cai, H., Xu, Z., Wu, W., Gu, B., Li, L., Cai, W. «MicroRNA-4516-mediated regulation of MAPK10 relies on 3' UTR cis-acting variants and contributes to the altered risk of Hirschsprung disease.» *Journal of Medical Genetics* 0 (Feb 2020): 1-9.
- Wen, M., Cong, P., Zhang, Z., Lu, H., & Li, T. (2018). «DeepMirTar: a deep-learning approach for predicting human miRNA targets.» *Bioinformatics* 34(22) (2018): 3781-3787.
- Wissink, E. M., Fogarty, E. A., and Grimson, A. «High-throughput discovery of post-transcriptional cis-regulatory elements.» *BMC Genomics* 17 (2016): 177.
- Witkos T.M., Koscianska E., Krzyzosiak W.J. «Practical aspects of microRNA target prediction.» *Curr. Mol. Med.* 11 (2011): 93–109.
- Wong, L. L., Wang, J., Liew, O. W., Richards, A. M., Chen, Y. T. «MicroRNA and heart failure.» *International journal of molecular sciences* 17(4) (2016): 502.
- Yang, H., and Wang, K. «Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR.» *Nat. Protoc.* 10(10) (2015): 1556–1566.

Yates, A., Beal, K., Keenan, S., McLaren, W., Pignatelli, M., Ritchie, G. R., et al. «Ensembl REST API: Ensembl data for any language.» *Bioinformatics* 31(1) (2014): 143–145.

Yung-Chun Chang, Cen-Chieh Chen, Yu-Lun Hsieh, Chien Chin Chen, Wen-Lian Hsu. «Linguistic Template Extraction for Recognizing Reader-Emotion and Emotional Resonance Writing Assistance.» *ACL-IJCNLP*, 2015: 775-780.

Zanotti S., Gibertini S., Curcio M., Savadori P., Pisanisi B., Morandi L., Cornelio F., Mantegazza R., Mora M. «Opposing roles of miR-21 and miR-29 in the progression of fibrosis in Duchenne muscular dystrophy.» *Biochim Biophys Acta*. 1852(7) (2015): 1451-1464.

Zhang Z., Huang L., Yu Z., et. al. «Let-7a functions as a tumor suppressor in Ewing's sarcoma cell lines partly by targeting cyclin-dependent kinase 6.» *DNA and cell biology* 33(3) (2014): 136-47.

Плотникова О.М., Скоблов М.Ю. «Эффективность программ, предсказывающих микроРНК-мРНК взаимодействия.» *Молекулярная биология* 52(3) (2018): 543-554.

Плотникова, О.М., Скоблов, М.Ю. «Роль микроРНК в патогенезе наследственных заболеваний.» *Медицинская генетика* 19(9) (2020): 5-17.