

ՀԱՅԱՍՏԱՆԻ ՀԱՆՐԱՊԵՏՈՒԹՅԱՆ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԶԳԱՅԻՆ ԱԿԱԴԵՄԻԱ
ՄՈԼԵԿՈՒԼԱՅԻՆ ԿԵՆՍԱԲԱՆՈՒԹՅԱՆ ԻՆՍՏԻՏՈՒՏ

ՆԵՐՍԻՍՅԱՆ ՍՏԵՓԱՆ ԱՇՈՏԻ

ՄԻԿՐՈՌՆԹ-ՆԵՐԻ ԵՎ ԴՐԱՆՑ ԻԶՈՉԵՎԵՐԻ ՖՈՒՆԿՑԻՈՆԱԼ ԱԿՏԻՎՈՒԹՅԱՆ ԴԵՐՇ
ՄԱՐԴՈՒ ՉԱՐՈՐԱԿ ՈՒՌՈՒՅՔՆԵՐԻ ՊԱԹՈԳԵՆԵՏԻԿ

Գ.00.03 - «Մոլեկուլային և բջջային կենսաբանություն» մասնագիտությամբ
կենսաբանական գիտությունների թեկնածուի գիտական աստիճանի
հայցման ատենախոսության

ՍԵՂՄԱԳԻՐ

ԵՐԵՎԱՆ - 2022

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК РЕСПУБЛИКИ АРМЕНИЯ
ИНСТИТУТ МОЛЕКУЛЯРНОЙ БИОЛОГИИ

НЕРСИСЯН СТЕПАН АПОТОВИЧ

РОЛЬ ФУНКЦИОНАЛЬНОЙ АКТИВНОСТИ МИКРОРНК И ИХ ИЗОФОРМ В
ПАТОГЕНЕЗЕ ЗЛОКАЧЕСТВЕННЫХ ОПУХОЛЕЙ ЧЕЛОВЕКА

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата биологических наук по специальности
03.00.03 - «Молекулярная и клеточная биология»

ЕРЕВАН - 2022

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ մոլեկուլային կենսաբանության
ինստիտուտի գիտական խորհրդում:

Գիտական ղեկավար՝ կ.գ.դ., պրոֆ., ՌԳԱ թղթ. անդամ Ալեքսանդր
Գրիգորիի Տոնեվիցկի
Պաշտոնական ընդդիմախոսներ՝ կ.գ.դ. Կարեն Բաբկենի Նազարյան
կ.գ.դ., պրոֆ. Յուրի Լևի Օրլով
Առաջատար կազմակերպություն՝ Ռուս-Հայկական համալսարան

Ատենախոսության պաշտպանությունը տեղի կունենա 2022 թ. սեպտեմբերի 23-ին, ժամը
14:00-ին, ՀՀ ԳԱԱ մոլեկուլային կենսաբանության ինստիտուտում, 042 մասնագիտական
խորհրդի նիստում (ՀՀ 0014, ք. Երևան, Հասարայան 7):

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ մոլեկուլային կենսաբանության
ինստիտուտի գրադարանում և <http://molbiol.sci.am/> կայքում:

Ատենախոսության սեղմագիրն առաքվել է 2022 թ. օգոստոսի 12-ին:

042 մասնագիտական խորհրդի գիտական քարտուղար,
կենս. գիտ. թեկնածու

Զ.Ա. Խաչատրյան

Тема диссертации утверждена на заседании ученого совета Института молекулярной
биологии НАН РА.

Научный руководитель:՝ ժ.Բ.Ն., профессор, член-корреспондент РАН
Тоневицкий Александр Григорьевич

Официальные оппоненты:՝ ժ.Բ.Ն. Назарян Карен Бабкенович
ժ.Բ.Ն., профессор, Орлов Юрий Львович

Ведущая организация:՝ Российско-Армянский университет

Защита диссертации состоится 23 сентября 2022 г. в 14:00 часов на заседании
специализированного совета 042, в Институте молекулярной биологии НАН РА (РА, 0014,
г. Ереван, ул. Асратяна 7).

С диссертацией можно ознакомиться в библиотеке Института молекулярной биологии НАН
РА и на сайте <http://molbiol.sci.am/>.

Автореферат разослан 12 августа 2022 г.

Ученый секретарь специализированного совета 042
кандидат биол. наук.

Խачատրյան Յ.Ա.



ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. МикроРНК – короткие некодирующие РНК (в среднем 22 нуклеотида), основной функцией которых является регуляция экспрессии генов (Bartel, 2009). Комплементарное связывание микроРНК с мРНК-мишенью, преимущественно происходящее в 3'-нетранслируемой области (3'-НТО) мРНК, приводит к деградации мРНК или остановке ее трансляции (Bartel, 2009). Наиболее важным участком микроРНК с функциональной точки зрения являются нуклеотиды 2–7 с 5'-конца молекулы, которые принято называть seed регионом: для успешного связывания микроРНК и мРНК необходима почти полная комплементарность seed последовательности микроРНК с целевой мРНК (допускается максимум одна пара некомплементарных нуклеотидов или один неспаренный выступающий нуклеотид) (Miranda et al., 2006). Количество комплементарных пар нуклеотидов за пределами seed региона положительно коррелирует с эффективностью подавления экспрессии гена (Miranda et al., 2006).

Большинство первичных микроРНК (при-микроРНК) в клетках человека транскрибируются РНК-полимеразой II из независимых генов микроРНК или из интронов белок-кодирующих генов (Treiber, Treiber & Meister, 2019). При-микроРНК имеют шпильчатую структуру и сразу после транскрипции обрабатываются рибонуклеазой III Drosha в ядре клетки (Treiber, Treiber & Meister, 2019). В результате разрезания шпильки при-микроРНК ферментом Drosha (отрезаются свободные концы молекулы) получается шпилька пре-микроРНК, переносимая белком экспортин 5 в цитоплазму, где рибонуклеаза III Dicer освобождает пре-микроРНК от петли (Treiber, Treiber & Meister, 2019). Одна из цепей полученного дуплекса загружается в РНК-индуцируемый комплекс выключения гена RISC, основным участником которого является один из белков семейства Argonaute (преимущественно Ago2) (Treiber, Treiber & Meister, 2019). Зрелая одноцепочечная микроРНК в составе комплекса RISC далее связывается с мРНК-мишенями, осуществляя регуляцию экспрессии генов.

Известно, что шпильки при- и пре-микроРНК неточно обрабатываются ферментами Drosha и Dicer, в результате чего появляются изоформы микроРНК – варианты одной зрелой микроРНК, отличающиеся друг от друга наличием или отсутствием 1–3 крайних нуклеотидов (Tomasello et al., 2021). Особую важность имеет вариация длины микроРНК с 5'-конца (соответствующие варианты называют 5'-изоформами микроРНК), так как добавление или удаление даже одного нуклеотида изменяет seed регион микроРНК, а, значит, изменяется и множество генов-мишеней молекулы (Tomasello et al., 2021).

Причины неточной обработки при-/пре-микроРНК рибонуклеазами III Drosha и Dicer на сегодняшний день не до конца известны и активно изучаются. При этом большая часть исследований проводится на единичных модельных шпильках в небольшом числе клеточных линий (Starega-Roslan, Galka-Marciniak & Krzyzosiak, 2015; Bofill-De Ros et al., 2022), в связи с чем, по нашим данным, средняя точность работы данных ферментов в различных видах клеток (в том числе опухолевых) неизвестна.

Аналогично, лишь в единичных работах представлены результаты экспериментов, доказывающих, что у двух 5'-изоформ одной микроРНК могут быть специфические гены-мишени (Tomasello et al., 2021; Zelli et al., 2021). Нами не было найдено систематических массовых исследований множеств мишеней 5'-изоформ микроРНК с использованием методов биоинформатики. Одной из возможных причин отсутствия таких работ является трудность применения программ и ресурсов для предсказания мишеней микроРНК к изоформам микроРНК. А именно, запуск широко известных программ, таких как miRDB или

TargetScan (Agarwal et al., 2015; Chen & Wang, 2020), требует серьезных навыков программирования и существенных вычислительных ресурсов.

На сегодняшний день известно более двух тысяч молекул микроРНК человека (Alles et al., 2019), и по существующим данным биоинформатических предсказаний на основе нуклеотидных последовательностей одна микроРНК имеет в среднем около 200 мишеней (Lewis, Burge & Bartel, 2005), что порождает сотни тысяч возможных регуляторных взаимодействий микроРНК-мишень. Не менее важную роль могут играть и другие молекулы, способные регулировать экспрессию как микроРНК, так и генов. Ярким примером таких молекул служат транскрипционные факторы (ТФ), напрямую контролирующие количество транскрибируемых микроРНК и мРНК. Анализ такого числа разнородных взаимодействий требует применения специальных биоинформатических и математических методов. Примером такого метода является анализ регуляторных сетей взаимодействий: рассматриваемым молекулам (микроРНК, гены, ТФ) ставят в соответствие вершины сети, а ориентированные ребра соответствуют взаимодействующим вершинам (Galatenko et al., 2018). Построение и анализ регуляторных сетей микроРНК, генов и ТФ успешно применялись для выявления молекулярных биомаркеров при различных патологиях (Mallik & Maulik, 2015; Li et al., 2019; Hao et al., 2019).

Одним из возможных подходов к построению регуляторных сетей является использование биоинформатических предсказаний взаимодействий и информации из литературных баз данных экспериментально валидированных взаимодействий (Guo et al., 2010; Sun et al., 2012; Fan et al., 2016), тогда как остальные методы основаны на проведении анализа ко-экспрессии по данным профилей экспрессии генов в некоторой выборке образцов (Lu et al., 2011; Do et al., 2019; Liu et al., 2020). Оба подхода имеют как достоинства, так и недостатки. Анализ, основанный на биоинформатических предсказаниях и базах данных, обычно не является тканеспецифичным. Корреляционные исследования в большинстве случаев приводят к общирному количеству ложноположительных взаимодействий, так как высокая корреляция значений экспрессии не гарантирует наличия взаимодействия между двумя молекулами. Например, отрицательная корреляция не взаимодействующих микроРНК и гена может быть обусловлена наличием ТФ, активирующего микроРНК и ингибирующего транскрипцию гена. Существующие программы для построения регуляторных сетей взаимодействий микроРНК, генов и ТФ не позволяют соединить в одну модель два описанных источника данных.

Онкологические заболевания являются одной из ведущих причин смертей в мире (Sung et al., 2021). В число лидеров по заболеваемости и летальности входят колоректальный рак (КРР, третье место по частоте заболеваемости, второе место по частоте летальных исходов) и рак молочной железы (РМЖ, первое место по заболеваемости и по летальности у женщин) (Sung et al., 2021). Широко известны примеры участия микроРНК в механизмах возникновения, прогрессии и метастазирования различных видов онкологических заболеваний (Garzon, Calin & Croce, 2009). В качестве примера часто приводятся микроРНК семейства miR-200, играющие роль опухолевых супрессоров, подавляя экспрессию ТФ ZEB1 и ZEB2, являющихся драйверами эпителиально-мезенхимального перехода (ЭМП) – ключевого процесса при метастазировании злокачественных опухолей (Hill, Browne & Tulchinsky, 2013). Однако, роль активности большинства микроРНК и их изоформ в патогенезе данных заболеваний не раскрыта, а точность прогностических биомаркеров на основе экспрессии микроРНК и их генов-мишеней далека от идеальной.

Целью диссертационной работы является определение роли микроРНК и их изоформ в патогенезе злокачественных опухолей человека с помощью методов биоинформатического анализа.

Для достижения указанной цели были поставлены следующие задачи:

- Оценить точность разрезания шпилек при-/пре-микроРНК ферментами Drosha и Dicer при различных видах рака.
- Определить множество изоформ микроРНК, экспрессированных в различных злокачественных опухолях человека, предсказать гены-мишени и оценить влияние вариации одного нуклеотида с 5'-конца молекулы на множество мишеней.
- Создать веб-портал, содержащий информацию об экспрессии и активности изоформ микроРНК при различных видах рака.
- Разработать, программно реализовать и апробировать алгоритм для построения сетей взаимодействий изоформ микроРНК, генов и транскрипционных факторов по данным профилей экспрессии РНК в выборке образцов и базам данных взаимодействий.
- Выявить ключевые взаимодействия изоформ микроРНК и их мишеней, играющие роль в патогенезе раков молочной железы и кишечника; использовать найденные взаимодействия для нахождения прогностических маркеров.

Научная новизна:

- Впервые рассчитана точность работы ферментов Drosha и Dicer на 5'-концах молекул микроРНК, показано, что полученные оценки одинаковы для различных видов рака.
- Впервые показано, что среднее количество общих мишеней изоформ микроРНК, различающихся одним нуклеотидом с 5'-конца, составляет лишь 6%.
- Создан единственный на сегодняшний день веб-портал isomiRTar, содержащий информацию об активности изоформ микроРНК при 31 виде рака у человека.
- Разработан, программно реализован и апробирован на данных ER⁺ РМЖ новый алгоритм, названный miRGTF-net, для построения регуляторных сетей взаимодействий микроРНК, генов и ТФ по данным транскриптомного анализа и базам данных взаимодействий.
- Впервые исследована активность неканонических изоформ микроРНК miR-203a и miR-192, предсказана их противоопухолевая роль при КРР.

Практическая значимость работы. Полученные результаты о предсказании течения заболевания по данным транскриптомного анализа первичной опухоли и регуляции экспрессии генов в злокачественных опухолях открывают новые возможности для подбора персонализированной терапии КРР и ER⁺ РМЖ, а также для поиска новых видов терапии.

Апробация работы. Материалы диссертационного исследования были представлены на конференции “International Conference on Research in Life-Sciences & Healthcare (ICRLSH),” 11 - 12 декабря, 2020; на конференции “25th International Conference on Research in Computational Molecular Biology (RECOMB 2021),” 29 августа – 2 сентября, 2021; на совместном семинаре Armenian Bioinformatics Institute (ABI) и факультета биологии и биотехнологии НИУ ВШЭ, 20 апреля, 2022, а также на заседаниях Ученого совета Института молекулярной биологии РАН (2022).

Объем и структура работы. Диссертация состоит из списка использованных сокращений и обозначений, введения, литературного обзора, материалов и методов, экспериментальной части, включая результаты исследований и их обсуждение, выводов, списка литературы, включающего 178 источников. Работа оформлена на русском языке, изложена на 118 страницах, иллюстрирована 30 рисунками и 9 таблицами.

Публикации. Основные результаты диссертации изложены в 7 научных работах, включая 7 научных статей, опубликованных в республиканских и международных научных изданиях.

МАТЕРИАЛЫ И МЕТОДЫ

Предсказание мишеней микроРНК и их изоформ. Нуклеотидные последовательности шпильчатых молекул при-микроРНК извлекали из портала miRBase версии 21 (Kozomara, Birgaonu & Griffiths-Jones, 2019). Аннотацию 5'-изоформ микроРНК проводили в соответствии со стандартной номенклатурой, введенной в работе (Telonis et al., 2015): строка до символа “|” определяла микроРНК, целое число после символа “|” определяло смещение позиции начала 5'-конца изоформы микроРНК относительно канонической позиции в направлении 5'-3' (для канонических микроРНК данное число равнялось нулю).

Предсказание мишеней 5'-изоформ на основании их последовательностей проводили с помощью двух широко используемых программ: miRDB версии 6.0 (Chen & Wang, 2020) и TargetScan версии 7.2 (Agarwal et al., 2015). Для каждой изоформы микроРНК уравнивали число предсказанных мишеней, выбирая соответствующее число предсказаний TargetScan на основании качества связывания. Медиана числа общих предсказаний miRDB и TargetScan равнялась 26.5%. Для пары 5'-изоформ микроРНК схожесть множеств определяли с помощью меры Жаккара.

Сбор и обработка данных секвенирования микроРНК и мРНК. В клетках HT-29 (ATCC, Манассас, США) и Сасо-2 (Институт цитологии РАН, Санкт-Петербург, Россия) вызывали гипоксический ответ с помощью воздействия двумя химическими агентами. В ходе первого эксперимента клетки инкубировали по 24 часа вместе с 5 мкМ производного оксихинолина 4896–3212 (Исследовательский Институт Химического Разнообразия, Химки, Россия), в ходе второго – с 300 мкМ хлорида кобальта(II). Библиотеки для секвенирования мРНК и микроРНК готовили с помощью наборов Illumina Stranded mRNA Library Prep Kit и NEBNext Multiplex Small RNA Library Prep Kit for Illumina (Illumina, Сан-Диего, США), соответственно, образцы секвенировали с помощью Illumina NextSeq 550. Непосредственно экспериментальную работу выполняли сотрудники факультета биологии и биотехнологии НИУ ВШЭ.

В случае данных секвенирования мРНК риды картировали на последовательность генома человека GENCODE GRCh38.p13 версии 34 с помощью STAR версии 2.7.5b (Dobin et al., 2013). В результате картирования получали матрицу, состоящую из количества ридов, соответствующих каждому гену. Картирование ридов секвенирования микроРНК проводили с помощью программы BCGSC miRNA (Chu et al., 2016). Ввиду отсутствия исходных файлов секвенирования РНК проекта The Cancer Genome Atlas (TCGA) в публичном доступе, скачивали таблицы с числами ридов, соответствующих каждой 5'-изоформе микроРНК и каждому гену, из портала GDC (<https://portal.gdc.cancer.gov>).

Нормализацию глубины секвенирования проводили с помощью алгоритма trimmed mean of M-values (TMM), доступного в пакете edgeR версии 3.30.3 (Robinson, McCarthy & Smyth, 2010). Данные секвенирования мРНК нормировали аналогично данным микроРНК (алгоритм TMM), при этом экспрессию дополнительно нормировали на эффективную длину гена. Анализ дифференциальной экспрессии микроРНК и мРНК проводили с помощью пакета DESeq2 версии 1.28.1 (Love, Huber & Anders, 2014). Для функциональной аннотации и анализа обогащения по функциональной принадлежности множеств генов использовали веб-сервис DAVID. Поправку на множественность проводили с помощью процедуры Бенджамини-Хохберга, вычисляя false discovery rate (FDR). Статистически достоверными считались изменения с $FDR < 0.05$.

Построение и анализ регуляторных сетей. Парные выборки профилей экспрессии изоформ микроРНК и их генов-мишеней использовали для подсчета коэффициента корреляции

Спирмена. В случае необходимости бинарной фильтрации взаимодействий микроРНК-мишень выбирали пары с корреляцией меньше -0.3 , $FDR < 0.05$.

Базу данных miRTarBase версии 7.0 (Chou et al., 2018) использовали для нахождения экспериментально подтвержденных взаимодействий микроРНК и мРНК (люциферазные репортерные конструкции). Базы данных TRRUST версии 2 (Han et al., 2018) и TransmiR версии 2 (Tong et al., 2019) использовали для поиска валидированных взаимодействий ТФ-ген и ТФ-микроРНК, соответственно.

На первом шаге алгоритма miRGTF-net строили ориентированную сеть взаимодействий, используя информацию о ребрах из баз данных взаимодействий. На следующем шаге алгоритма встраивали данные об экспрессии в реберную структуру сети. Мы предполагали линейную зависимость между значениями экспрессий каждой вершины и значениями экспрессий соответствующих регулирующих вершин. Описанную линейную статистическую модель обучали с помощью техники ридж-регрессии. Степень регуляции определяли как коэффициент детерминации. Мету силы взаимодействия ребра определяли как абсолютное значение стандартизованного бета коэффициента регрессии. По умолчанию удаляли все вершины, не проходящие порог 0.3 по степени регуляции. Также удаляли ребра с мерой силы взаимодействия меньше 10% перцентили соответствующего распределения.

Транскриптомные тест-системы. Использовали следующие наборы данных транскриптомов злокачественных опухолей молочной железы, полученные с портала Gene Expression Omnibus (GEO):

- Наборы данных Affymetrix Human Genome U133A Array: GSE1456, GSE3494, GSE6532, GSE12093 и GSE17705;
- Наборы данных Affymetrix Human Genome U133 Plus 2.0 Array: GSE6532 и GSE20685.

Для классификации опухолевых образцов с и без рецидива использовали алгоритм, реализованный в пакете ExhauFS (Nersisyan et al., 2022). Для каждой из возможных длин геновых подпоследовательностей (от 1 до 15) обучали модель метода опорных векторов на обучающей выборке, качество оценивали на двух фильтрационных выборках. В случае прохождения пороговой величины 0.65 метриками TPR (чувствительность), TNR (специфичность) и ROC AUC (площадь под ROC кривой) качество классификатора оценивали на независимой валидационной выборке. В случае микрочипов Affymetrix Human Genome U133A объединение выборок GSE3494 и GSE6532 использовали в качестве обучающей выборки, GSE12093 и GSE17705 использовали для фильтрации, валидацию проводили на выборке GSE1456. Для микрочипов Affymetrix Human Genome U133 Plus 2.0 настройку и фильтрацию классификаторов проводили на данных GSE20685 и GSE6532, соответственно. Для валидации на выборке TCGA-BRCA (секвенирование РНК) проводили рандомизированную кросс-валидацию с порогом 0.6 для усредненных TPR, TNR и ROC AUC.

Программирование и статистика. Разработанный метод miRGTF-net реализовывали с помощью языка программирования Python, исходный код программы доступен по адресу <https://github.com/s-a-nersisyan/miRGTF-net>. Реализацию портала isomiRTar (<https://isomirtar.hse.ru>) проводили с использованием модуля Flask версии 2.0 языка программирования Python и базы данных PostgreSQL. Исходный код портала доступен по адресу <https://github.com/s-a-nersisyan/isomiRTar>. Также использовали следующие модули языка программирования Python: SciPy (применение статистических тестов), NetworkX (построение и анализ сетей), scikit-learn (методы машинного обучения), lifelines (анализ выживаемости), Pandas, NumPy (всевозможные вычисления), Seaborn, Matplotlib (построение графиков).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1. Профили экспрессии 5'-изоформ микроРНК

1.1. Общий и специфический профиль экспрессии 5'-изоформ микроРНК в злокачественных опухолях человека

Мы составили исчерпывающий список из 1022 5'-изоформ микроРНК, экспрессированных в 31 типе рака, используя данные секвенирования микроРНК проекта TCGA. Данный список содержал 168 5'-изоформ, отмеченных как высокоэкспрессированные по крайней мере при одном виде рака, включая 31 неканоническую изоформу. Приблизительно 38% высокоэкспрессированных микроРНК (63 молекулы) были экспрессированы в 10 или большем числе типов злокачественных опухолей человека, включая 11 неканонических 5'-изоформ микроРНК. Наконец, 18 микроРНК, включая большинство представителей семейств let-7 и miR-30, были высокоэкспрессированы во всех 31 проекте TCGA. При этом большинство высокоэкспрессированных микроРНК (105 из 168, 62%) были экспрессированы менее чем при 10 типах раков (опухоль-специфическая экспрессия). Из них 39 экспрессировались лишь в одном типе опухолевых клеток. Высокий процент опухоль-специфических микроРНК хорошо согласуется с недавним исследованием Телониса и соавторов. А именно, авторами данного исследования было показано, что различные виды рака могут быть высокоточно классифицированы на основе информации о «высокой» или «низкой» экспрессии каждой изоформ микроРНК в образце (Telonis et al., 2017).

В подавляющем большинстве случаев высокая экспрессия неканонической изоформы микроРНК была сцеплена с высокой экспрессией соответствующей канонической микроРНК, что подтверждает валидность разметки канонических микроРНК на портале miRBase (у данного правила было лишь два исключения). Интересно, что шесть микроРНК имели более двух высокоэкспрессированных 5'-изоформ.

1.2. Точность работы ферментов Drosna и Dicer

Используя информацию об экспрессии изоформ микроРНК в различных опухолевых тканях, для каждой шпильки при-/пре-микроРНК мы оценивали точность разрезания ферментами Drosna и Dicer. Для этого, в первую очередь, для каждой микроРНК определяли наиболее экспрессированную 5'-изоформу. Далее, экспрессию данной мажорной изоформы нормировали на суммарную экспрессию всех изоформ данной микроРНК, получая оценку гетерогенности работы ферментов Drosna (для -5p цепей микроРНК) и Dicer (для -3p цепей микроРНК). Высокие значения (95% и более) соответствовали случаям, когда на мажорную изоформу микроРНК приходится подавляющее большинство ридов секвенирования, то есть разрезание соответствующей шпильки производится точно; в обратном случае наблюдалось неточная работа Drosna или Dicer.

Оказалось, что характер распределений долей наиболее экспрессированных изоформ микроРНК схож во всех видах злокачественных опухолей как для Drosna, так и для Dicer. В случае Drosna подавляющая масса распределения была сосредоточена в районе единицы, в случае Dicer, производящим обработку 5'-концов -3p плеч микроРНК, более половины вероятностной массы распределения стабильно приходилось на гомогенное разрезание.

Объединение 168 микроРНК, обладающих высокой экспрессией при по крайней мере одном виде рака, позволило нам оценить точность работы ферментов Drosna и Dicer в терминах доли шпилек при- и пре-микроРНК, на которых производилось гомогенное

разрезание. Точность Drosha составила 87% (58 из 67 микроРНК -5p), точность Dicer – 58% (37 из 64 микроРНК -3p). Отметим, что Dicer осуществляет свою работу в цитоплазме клетки, что может быть одной из возможных причин более низкой точности работы по сравнению с внутриядерной обработкой при-микроРНК Drosha (большее количество третьих факторов, таких как РНК-связывающие белки, может связываться с пре-микроРНК в цитоплазме). С другой стороны, механизм может полностью скрываться в структурных особенностях данных белков. Для ясности изложения позиции разрезания ферментами Drosha/Dicer, их расположение (ядро/цитоплазма) и значения точности работы приведены на Рисунке 1.

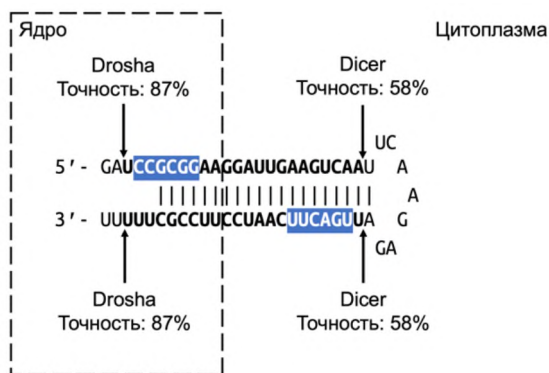


Рисунок 1. Точность работы ферментов Drosha и Dicer. Seed регион двух возможных зрелых целей микроРНК выделен синим.

Для того, чтобы углубить полученные результаты, были проведены сравнения точностей работы указанных ферментов в первичных опухолях и образцах прилежащих здоровых тканей. Для получения статистически достоверных результатов особое внимание уделяли пяти видам рака, для которых доступны данные секвенирования микроРНК по крайней мере 50 образцов нормальных тканей: TCGA-BRCA (рак молочной железы), TCGA-KIRC (светлоклеточный рак почки), TCGA-LIHC (гепатоцеллюлярная карцинома), TCGA-PRAD (рак простаты) и TCGA-TLCA (рак щитовидной железы). Аналогично сравнению точности Drosha/Dicer между образцами различных опухолей, сравнение опухолей и здоровых тканей также позволило заключить, что средняя точность работы ферментов не меняется при злокачественном переходе.

2. Анализ мишеней 5'-изоформ микроРНК, портал isomiRTar, программа miRGTF-net

2.1. Предсказания мишеней 5'-изоформ микроРНК на основе их последовательностей

Две широко распространенные программы miRDB и TargetScan были использованы для предсказания генов-мишеней всех рассмотренных 5'-изоформ микроРНК. Медиана числа предсказанных мишеней была равна 209. Ранее опубликованные исследования сообщали о том, что молекулы микроРНК имеют в среднем от 200 до 606 мишеней в зависимости от стратегии биоинформатического предсказания (Lewis, Burge & Bartel, 2005; Krek et al., 2005; Chen & Wang, 2020), что сопоставимо с полученными нами оценками для 5'-изоформ микроРНК.

Далее, полученные предсказания были использованы для оценки влияния вариации одного нуклеотида с 5'-конца микроРНК на множество ее мишеней. Для каждой пары 5'-изоформ микроРНК, отличающихся друг от друга сдвигом на 1 нуклеотид с 5'-конца (соседние

изоформы микроРНК), была рассчитана мера Жаккара для множеств соответствующих генов-мишеней. Медиана меры Жаккара, рассчитанная по всем парам соседних 5'-изоформ микроРНК, была равна 0.06, максимальное значение – 0.39, минимальное значение – 0 (Рисунок 2). Таким образом, различные 5'-изоформы микроРНК, образуемые из одной молекулы предшественника, могут рассматриваться как независимые функциональные единицы.

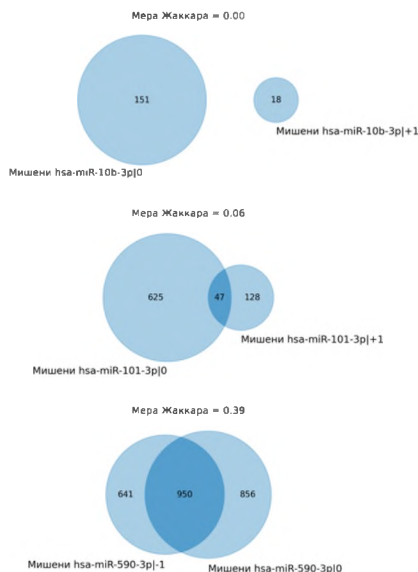


Рисунок 2. Диаграммы Эйлера для множеств предсказанных мишеней (объединение miRDB и TargetScan) соседних изоформ микроРНК.

2.2. Корреляционный анализ уровней экспрессии 5'-изоформ микроРНК и их мишеней

Для того чтобы выяснить, какие из предсказанных мишеней могут действительно подавляться соответствующими 5'-изоформами микроРНК на уровне транскриптома, мы выбрали взаимодействия изоформ микроРНК и их мишеней, подкрепленные значимой отрицательной корреляцией уровней экспрессии в выборке образцов TCGA (анализ проводили отдельно для каждого типа рака). Результаты данного анализа также были представлены в виде регуляторных сетей (графов) для каждого вида рака, где 5'-изоформы были соединены ориентированными ребрами с вершинами, соответствующими антикоррелированным мишеням. Для удобства дальнейшего изложения *активностью* 5'-изоформы микроРНК в некоторой выборке образцов будем обозначать число значимо антикоррелированных мишеней данной молекулы (то есть полустепень исхода вершины в соответствующей сети взаимодействий).

Активность 5'-изоформ микроРНК существенно варьировалась между различными видами рака, однако в большинстве случаев медиана активности не превышала 10, что на порядок ниже соответствующего значения для предсказаний, основанных лишь на нуклеотидных последовательностях микроРНК и потенциальной мишени. Тем не менее, практически в каждом виде рака присутствовало несколько очень активных изоформ микроРНК (активность более 100).

Число предсказанных с помощью miRDB и TargetScan мишеней во многом определяло активность изоформы микроРНК при раке: соответствующие коэффициенты корреляции Спирмена были положительны и статистически значимо отличны от нуля для всех видов рака. Уровни экспрессии 5'-изоформ микроРНК также положительно и статистически значимо коррелировали с их активностью для всех видов рака ($p < 10^{-3}$), однако значения коэффициентов корреляции были невысоки, варьируясь между 0.2 и 0.4. Важно отметить близкую к нулю корреляцию активности и экспрессии во множестве высокоэкспрессированных 5'-изоформ микроРНК при каждом виде рака. Таким образом, высокая экспрессия является необходимым, но не достаточным условием для подавления экспрессии многих генов-мишеней.

2.3. Портал isomiRTar

Для того чтобы сделать профили экспрессии 5'-изоформ микроРНК и результаты предсказаний их мишеней легкодоступными, мы разработали веб-портал isomiRTar (<https://isomiRTar.hse.ru>). Главная страница портала позволяет осуществлять поиск в пяти различных режимах: по виду рака, по 5'-изоформе микроРНК, по гену и по тройке (5'-изоформа микроРНК, ген, вид рака). Наиболее важными частями портала являются интерактивная сеть взаимодействий 5'-изоформ микроРНК и их мишеней при выбранном виде рака, базовая статистика для 5'-изоформ микроРНК в выбранном типе рака, визуализация экспрессии 5'-изоформ выбранной микроРНК, результаты корреляционного анализа между выбранной изоформой микроРНК и ее предсказанными мишенями, или предсказанными регуляторами выбранного гена, а также визуализация совместного распределения соответствующих уровней экспрессии (Рисунок 3).

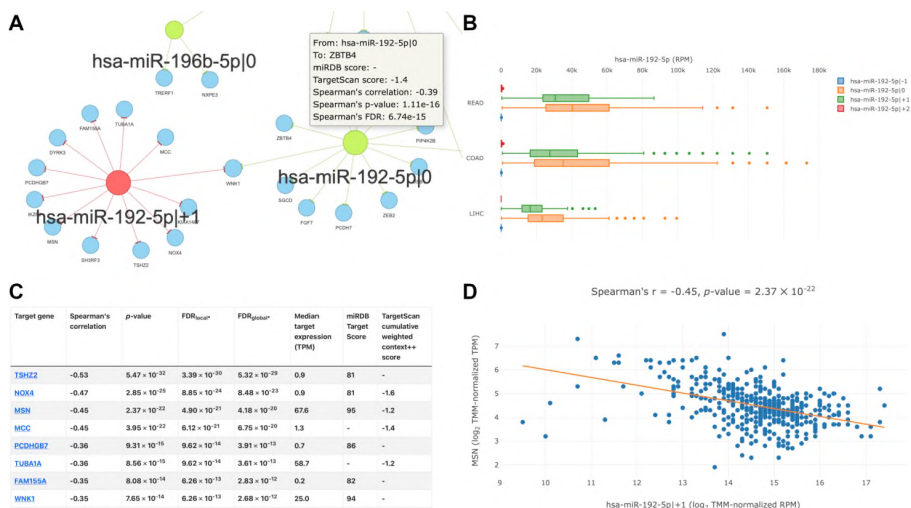


Рисунок 3. Портал isomiRTar. **(А)** Фрагмент сети взаимодействий 5'-изоформ микроРНК и мишеней колоректального рака. **(Б)** Экспрессия 5'-изоформ hsa-miR-192-5p при различных типах рака. **(С)** Предсказанные мишени и результаты корреляционного анализа для hsa-miR-192-5p+1 в тканях колоректального рака. **(Д)** Совместное распределение экспрессии hsa-miR-192-5p и гена мишени *MSN* при колоректальном раке.

В отличие от двух альтернативных порталов об изоформах микроРНК (Tumor IsomiR Encyclopedia, IsomiR Bank), isomiRTag содержит информацию о мишенях 5'-изоформ микроРНК как на уровне предсказаний с помощью актуальных программ miRDB и TargetScan, так и на уровне проведенного анализа ко-экспрессии, в результате чего представленные взаимодействия являются опухоль-специфическими.

2.4. Алгоритм miRGTF-net

На разработанном нами портале isomiRTag приведены результаты исследований регуляторных сетей взаимодействий 5'-изоформ микроРНК и их генов-мишеней. Однако, наличие данных транскриптомного анализа позволяет численно оценивать и другие типы взаимодействий, в частности, регуляцию транскрипции генов и микроРНК транскрипционными факторами. Для построения регуляторных сетей взаимодействий микроРНК-ген-ТФ нами был разработан алгоритм, названный miRGTF-net.

Рассматривался ориентированный граф, вершины которого были ассоциированы с генами и молекулами микроРНК (или изоформами микроРНК), и в котором вершины A , B соединялись ребром в следующих случаях:

- A – ген, кодирующий ТФ, B – ген, являющийся мишенью A ;
- A – ген, кодирующий ТФ, B – микроРНК, являющаяся мишенью A ;
- A – микроРНК, B – ген, являющийся мишенью A ;
- A – ген, B – интронная микроРНК гена-хозяина A ;

Информация о таких ребрах бралась из баз данных экспериментально валидированных взаимодействий или из результатов биоинформатических предсказаний.

После построения литературного графа взаимодействий проводилась оценка сил взаимодействий с использованием парного профиля экспрессии микроРНК и мРНК в выборке образцов. Получаемая в результате работы алгоритма сеть взаимодействий содержала вершины, имеющие значительное влияние на некоторую другую вершину и/или значимо регулируемую другой вершиной.

Основным преимуществом использования данных экспрессии является возможность построения сетей взаимодействий, специфичных для выбранного типа тканей. С другой стороны, использование данных экспрессии порождает одно из главных ограничений метода: так как некоторые вершины сети регулируются десятками и даже сотнями других молекул, для использования метода (а именно, обучения регрессионных моделей) необходимы выборки из десятков и сотен образцов. Отметим также, что разработанный алгоритм не зависит от технологии профилирования экспрессии и может быть использован как с данными секвенирования РНК, так и с данными микрочипового анализа.

3. Роль микроРНК и их изоформ в патогенезе РМЖ и КРР

3.1. Регуляторная сеть микроРНК-ген-ТФ ER⁺ РМЖ

Нами были рассмотрены злокачественные опухоли молочной железы, экспрессирующие рецепторы эстрогена (ER⁺ РМЖ). Лишь одна неканоническая 5'-изоформа микроРНК (hsa-miR-101-3p|+1) имела активность более 10 в данных клетках. Учитывая данное обстоятельство, мы сузили множество 5'-изоформ микроРНК до канонических форм, при этом появилась возможность использования литературной базы данных miRTarBase, включающей в себя результаты экспериментально валидированных мишеней канонических форм

микроРНК. Построенная сеть сдержала 371 вершину, включая 312 генов и 59 микроРНК. Наибольшее число взаимодействий пришлось на регуляцию экспрессии генов (371 ребро) и микроРНК (73 ребра) с помощью ТФ, связыванию микроРНК и мишеней (69 ребер), и лишь 11 ребер соответствовали ко-экспрессии генов-хозяев и интронных микроРНК. Для дальнейшей работы была выбрана только главная 336-вершинная компонента слабой связности (Рисунок 4). Полученные на данном этапе результаты предсказывают незначимую роль взаимодействий между интронными микроРНК и их генами-хозяевами (у большинства человеческих микроРНК есть независимые сайты для связывания факторов транскрипции), тогда как большое количество исследований автоматически рассматривает всевозможные взаимодействия данного вида.

Для каждой вершины графа были рассчитаны полустепени захода и исхода, чтобы найти наиболее регулируемые и регулирующие молекулы. Тринадцать генов имели статистически значимую полустепень исхода, включая ТФ E2F1, SP1 и SP11, демонстрирующих особенно высокие полустепени: 38, 38 и 21, соответственно. Функциональная аннотация генов сети выявила перепредставленность биологических путей, связанных с раком и метастазированием, включая клеточную пролиферацию (86 генов), апоптоз (74 гена), клеточный цикл (51 ген), ответ на повреждения ДНК (45 генов), ответ на гипоксию (33 гена), клеточную адгезию (21 ген).

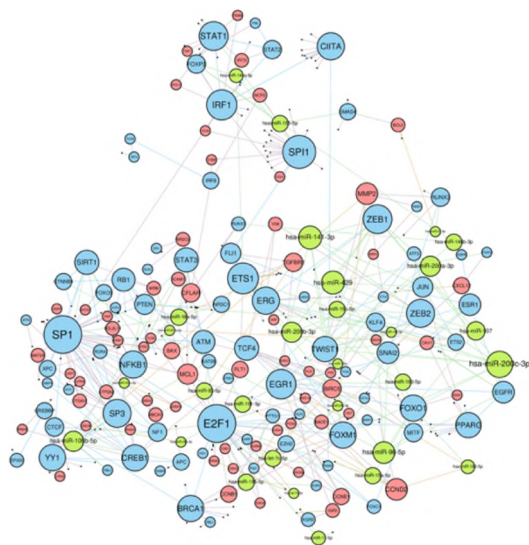


Рисунок 4. Главная компонента сети взаимодействий микроРНК-ген-ТФ при ER⁺ РМЖ. Синие вершины соответствуют генам, кодирующим ТФ, красные – генам, не кодирующим ТФ, зеленые – микроРНК. Цвета ребер соответствуют типам взаимодействий. Размер вершины линейно зависит от ее степени.

3.2. Транскриптомные тест-системы ER⁺ РМЖ

Построенная сеть взаимодействий состояла из вершин со значимой активностью на внутриклеточном уровне. Мы предположили, что такие вершины могут быть включены в большое количество “испорченных” внутриклеточных взаимодействий, а значит, могут являться качественными предикторами рецидива болезни. Для проверки гипотезы были построены классификаторы риска развития рецидива ER⁺ РМЖ в течение пяти лет после операции на основе данных об экспрессии генов в первичной опухоли. В качестве алгоритма интерпретируемой классификации использовался метод опорных векторов (SVM). В

найденных нами выборках (помимо TCGA-BRCA) профилирование транскриптомов проводили с помощью микрочипов Affymetrix Human Genome U133A. Так как данная технология не позволяет измерять экспрессию микроРНК, соответствующие вершины были исключены из последующего анализа.

В результате применения алгоритма на множестве генов из регуляторной сети было найдено 134 прогностические комбинации (генные подписи), прошедшие фильтрационные пороги качества: чувствительность, специфичность и ROC AUC выше 0.65. Для некоторых длин генных подписей качество было высоким и на валидационной выборке. А именно, для длин комбинаций генов $k = 5, 6, 8, 10$ более 85% классификаторов демонстрировали высокое качество с ROC AUC, чувствительностью и специфичностью выше 0.65. Построенные классификаторы далее проверялись на выборках опухолей, транскриптомы которых анализировали с помощью микрочипов Affymetrix Human Genome U133 Plus 2.0 (новая версия Affymetrix Human Genome U133A) и секвенирования РНК.

В результате применения дополнительной фильтрации была отобрана генная подпись, состоящая из восьми генов (*CDK1*, *FOXMI*, *LRIG2*, *MSH2*, *PLK1*, *RACGAP1*, *RRM2* и *TMPO*) и показывающая высокую точность для всех платформ транскриптомного анализа: средние ROC AUC, чувствительность и специфичность классификатора на наборе данных TCGA-BRCA были равны 0.72, 0.61 и 0.73, в то время как те же метрики были равны 0.78, 0.67 и 0.71 на данных микрочипового анализа (Рисунок 5). Такая точность ставит полученную прогностическую подпись в один ряд с такими известными коммерческими тест-системами как OncotypeDX (Paik et al., 2004) и MammaPrint (van 't Veer et al., 2002) согласно отчету рабочей группы EGAPP (Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group, 2009). В соответствии с высокой точностью классификации, проведенный анализ выживаемости для всех пациентов из микрочиповой когорты также выявил значимое различие в безрецидивной выживаемости в группах предсказанного высокого и низкого рисков ($p = 0.0015$).

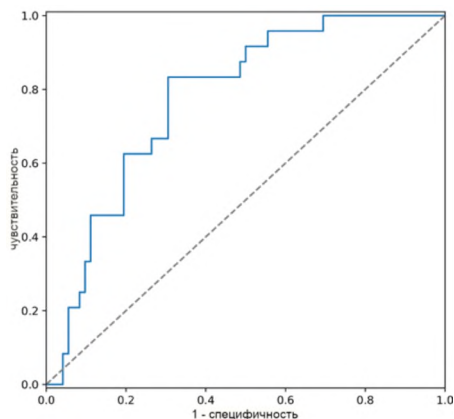


Рисунок 5. ROC кривая (AUC = 0.78) прогностической модели SVM, основанной на экспрессии восьми генов: *CDK1*, *FOXMI*, *LRIG2*, *MSH2*, *PLK1*, *RACGAP1*, *RRM2* и *TMPO*, рассчитанная по валидационной выборке GSE1456.

3.3. ТФ E2F1 – предположительный регулятор прогностических генов ER⁺ РМЖ

Совокупность прогностических комбинаций, прошедших пороги фильтрации, содержала 49 уникальных генов. Четырнадцать из них были значимо перепредставлены в комбинациях (биномиальный тест $p < 0.05$), при этом большинство из них было вовлечены в клеточные

процессы, связанные с раком и метастазированием. Отметим, что семь из рассмотренных генов (*ATAD2*, *CCNB1*, *FOXM1*, *KIF2C*, *PLK1*, *RACGAP1* и *RRM2*) были напрямую активированы фактором транскрипции E2F1. В частности, данное пересечение было статистически значимым по отношению к полустепени исхода вершины E2F1 (гипергеометрический тест $p = 0.00073$).

Пациенты из проанализированных выборок принимали гормональную терапию (тамоксифен), направленную на блокировку активности эстрогенового рецептора альфа, кодируемого геном ESR1. Как транскрипционный фактор ESR1 напрямую активирует транскрипцию E2F1 (Wang et al., 1999; Ngwenya & Safe, 2003). В то время как *ESR1* и *E2F1* не были коррелированы в безрецидивной группе ($r = 0.086$, $p = 0.34$), статистически значимая корреляция была отмечена в опухолях пациентов с рецидивом ($r = 0.22$, $p = 0.013$). Интересно, что ESR1 и E2F1 (как факторы транскрипции) имеют множество общих мишеней: 15 общих генов из 76- и 134-элементных множеств мишеней, соответственно. В числе общих мишеней были ключевые для метастазирования белки p53, Мус, регулятор апоптоза Bcl-2. Таким образом, взаимодействие ESR1-E2F1 может быть ключевым при развитии рецидива ER-положительного рака молочной железы.

3.4. Регуляторная сеть 5'-изоформ микроРНК и их мишеней при КРР

В отличие от РМЖ клетки КРР содержали несколько неканонических 5'-изоформ микроРНК с большим числом предсказанных и антикоррелированных мишеней (Рисунок 6). Четыре из них имели наиболее высокие полустепени исхода: hsa-miR-142-3p|+1 (34), hsa-miR-22-3p|+1 (31), hsa-miR-203a-3p|+1 (17) и hsa-miR-192-5p|+1 (12).

Обе формы hsa-miR-203a-3p (каноническая и неканоническая) подавляли экспрессию генов, связанных с ЭМП. Экспрессия hsa-miR-203a-3p|+1 отрицательно коррелировала с экспрессией предсказанной мишени *ZEB1*, являющейся ключевым ТФ для ЭМП (Dongre & Weinberg, 2019). Каноническая форма hsa-miR-203a-3p|0 потенциально подавляла экспрессию *ADAM12*, известного индуктора ЭМП (Ruff et al., 2015), и ЭМП-ТФ *SNAI2* (Ma et al., 2017). Схожие результаты были получены для hsa-miR-192-5p. Экспрессия неканонической изоформы hsa-miR-192-5p|+1 отрицательно коррелирована с экспрессией предсказанных онкогенов-мишеней *NOX4* (Lin et al., 2017) и *MSN* (Kim et al., 2012), а канонической формы с ЭМП-ТФ *ZEB2* (Dongre & Weinberg, 2019) и стимулятором ангиогенеза *WINK1* (Sie et al., 2020). Оба взаимодействия были ранее валидированы с использованием репортерных конструкций (Elvira-Matelot et al., 2010; Geng et al., 2014). Таким образом, предсказанные потенциальные мишени неканонических изоформ hsa-miR-203a-3p и hsa-miR-192-5p, отличающихся от канонических форм отсутствием первого нуклеотида с 5'-конца, согласуются с известной противоопухолевой ролью соответствующих канонических микроРНК.

Анализ обогащения мишеней 5'-изоформ микроРНК по функциональной принадлежности выявил статистически значимое обогащение мишеней hsa-miR-148a-3p|0 про-опухолевыми биологическими путями, включая адгезию (KEGG hsa04510), взаимодействия клеточных рецепторов и внеклеточного матрикса (hsa04512), сигнальные пути TGF- β (hsa04350) и PI3K-Akt (hsa04151). Указанные пути разделяли множество мишеней hsa-miR-148a-3p|0: *COL4A1*, *ITGA5*, *ITGA11* и *LAMB2* присутствовали сразу в трех путях, *FLT1* и *ROCK1* – в двух, а *CSF1*, *FBN1*, *LTBP1* и *TGFB2* – в одном. Взаимодействия с половиной из данных мишеней (*CSF1*, *ITGA5*, *ITGA11*, *ROCK1*, *TGFB2*) были ранее валидированы с использованием репортерных конструкций.

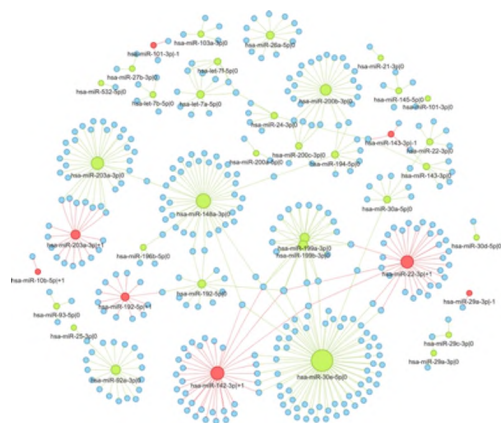


Рисунок 6. Сеть регуляторных взаимодействий 5'-изоформ и их мишеней при колоректальном раке. Зеленые вершины соответствуют каноническим 5'-изоформам микроРНК, красные – неканоническим, синие – генам мишеням. Цвета ребер соответствуют цветам инцидентных 5'-изоформ микроРНК. Размер вершины линейно зависит от ее полустепени исхода.

3.5. Гипоксия как регулятор экспрессии микроРНК при КРР

Нами был проведен анализ данных мРНК и микроРНК секвенирования клеточных линий колоректального рака Сасо-2 и НТ-29 под воздействием двух химических агентов (CoCl₂ и оксихинолин), эффективно моделирующих гипоксию. Список изменивших экспрессию микроРНК клеток Сасо-2 включал ранее упомянутую микроРНК hsa-miR-148a-3р, составляющую 6.9% от всех ридов секвенирования микроРНОма Сасо-2. Экспрессия hsa-miR-148a-3р была снижена в полтора раза при обоих воздействиях, индуцирующих гипоксию в клетках Сасо-2. Также была отмечено значимое повышение экспрессии двух мишеней hsa-miR-148a-3р: *ITGA5* и *PRNP*.

Далее, пациенты с КРР были разделены на две группы согласно экспрессии гена *ITGA5* и на две группы согласно экспрессии *PRNP*. В полном соответствии с предположительно протективной ролью hsa-miR-148a-3р, высокая экспрессия *ITGA5* и *PRNP* была ассоциирована с пониженной общей выживаемостью пациентов с КРР (логранк тест $p = 0.0133$ для *ITGA5* и $p = 0.0119$ для *PRNP*, Рисунок 7). Таким образом, экспрессию генов-мишеней hsa-miR-148a-3р можно использовать в качестве прогностических биомаркеров КРР.

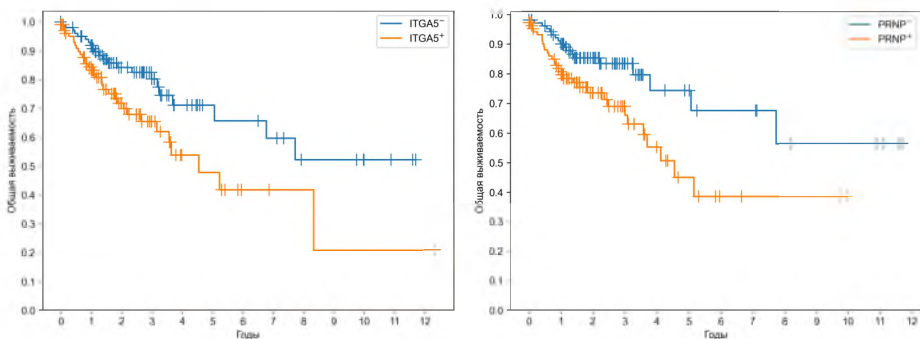


Рисунок 7. Высокая экспрессия *ITGA5* и *PRNP* в первичной опухоли связана с пониженной выживаемостью пациентов с КРР.

ЗАКЛЮЧЕНИЕ

Используя данные секвенирования микроРНК и мРНК первичных опухолей, соответствующих 31 виду рака, мы провели оценку точности разрезания молекул при- и пре-микроРНК ферментами Droscha и Dicer, соответственно. Оказалось, что рассчитанные точности фактически одинаковы в различных типах опухолевых клеток: Droscha производит гомогенное разрезание шпильки в 86% случаев, Dicer – в 58%.

Далее, нуклеотидные последовательности 1022 5'-изоформ микроРНК использовали для предсказания генов-мишеней с помощью программ miRDB и TargetScan. Полученные предсказания были использованы для анализа ко-экспрессии, в результате чего были получены суженные множества опухоль-специфических мишеней. Одной изоформе микроРНК в среднем соответствовало 209 генов-мишеней при предсказании на основе последовательности и на порядок меньше при дополнительном опухоль-специфическом анализе. При этом пара изоформ одной микроРНК с отличием в один нуклеотид с 5'-конца в среднем имела лишь 6% общих мишеней, что подтверждает важность рассмотрения изоформ микроРНК в качестве отдельных функциональных единиц.

Полученные данные экспрессии и активности изоформ микроРНК при различных видах рака были опубликованы на разработанном нами веб-портале isomiRTag. Также мы разработали и программно реализовали алгоритм miRGTF-net, позволяющий интегрировать данные мишеней микроРНК с активностью транскрипционных факторов. Научное сообщество может свободно использовать портал isomiRTag и программу miRGTF-net при решении различных задач, связанных с микроРНК и их изоформами в контексте рака.

В результате применения алгоритма miRGTF-net к данным ER⁺ ПМЖ была выявлена генная подпись, состоящая из восьми генов (*CDK1*, *FOXM1*, *LRIG2*, *MSH2*, *PLK1*, *RACGAP1*, *RRM2* и *TMPO*) и позволяющая предсказывать 5-летний рецидив заболевания с высокой точностью (ROC AUC, чувствительность и специфичность классификатора близки к 70%). Важно, что значимая часть полученных прогностических генов регулировалась ТФ E2F1, что позволило сформулировать гипотезу о важности взаимодействия ESR1-E2F1 при развитии рецидива заболевания.

Анализ экспрессии и активности изоформ микроРНК в клетках КРП позволил предсказать противоопухолевую роль изоформ микроРНК miR-148a, miR-192 и miR-203a. Дальнейшие эксперименты показали значимое уменьшение уровней экспрессии miR-148a и увеличение экспрессии ее генов-мишеней *ITGA5* и *PRNP5* в клетках КРП Caco-2 при гипоксическом ответе. Значимая связь повышенной экспрессии *ITGA5* и *PRNP5* с пониженной общей выживаемостью пациентов с КРП дополнительно подтверждает противоопухолевую роль miR-148a.

Таким образом, все поставленные задачи были выполнены в полном объеме, цель исследования достигнута.

ВЫВОДЫ

1. Точность работы ферментов Drosha и Dicer на 5'-концах микроРНК одинакова в различных видах злокачественных клеток и составляет в среднем 86% и 58%, соответственно.
2. Среднее число предсказанных по нуклеотидным последовательностям мишеней микроРНК равно 209, при дополнительном использовании данных секвенирования микроРНК и мРНК – на порядок ниже. Изоформы микроРНК, различающиеся одним нуклеотидом с 5'-конца, в среднем имеют 6% общих мишеней.
3. Создан веб-портал isomiRTag, содержащий информацию об экспрессии и активности 5'-изоформ микроРНК в образцах 31 вида злокачественных опухолей человека.
4. Разработан и программно реализован алгоритм miRGTF-net, позволяющий строить регуляторные сети взаимодействий микроРНК, генов и ТФ по входным данным транскриптомного анализа в некоторой выборке образцов и базам данных взаимодействий.
5. В результате анализа регуляторной сети взаимодействий микроРНК, генов и ТФ при ER+ РМЖ, построенной с помощью программы miRGTF-net, выявлены восемь генов (*CDK1*, *FOXM1*, *LRIG2*, *MSH2*, *PLK1*, *RACGAP1*, *RRM2* и *TMPO*), образующих точный и кроссплатформенный классификатор прогноза 5-летнего рецидива ER+ РМЖ (ROC AUC более 0.7). Показано, что значимую часть данных генов регулирует ТФ E2F1.
6. Предсказана противоопухолевая роль изоформ микроРНК miR-148a, miR-192 и miR-203a при КРР. Показано, что при гипоксии клеток КРР значимо снижается экспрессия miR-148a и значимо повышается экспрессия ее потенциальных генов-мишеней *ITGA5* и *PRNP*.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

СТАТЬИ

1. Нерсиян С. Изоформы микроРНК miR-148a и miR-203a предположительно играют роль супрессоров колоректального рака. **Вестник РГМУ.** 2022; 3:34-40. doi: 10.24075/vrgmu.2022.028
2. Zhiyanov A, Nersisyan S, Tonevitsky A. Hairpin sequence and structure is associated with features of isomiR biogenesis. **RNA Biol.** 2021; 18(sup1):430-438. doi: 10.1080/15476286.2021.1952759
3. Nersisyan S, Galatenko A, Chekova M, Tonevitsky A. Hypoxia-Induced miR-148a Downregulation Contributes to Poor Survival in Colorectal Cancer. **Front Genet.** 2021; 12:662468. doi: 10.3389/fgene.2021.662468
4. Nersisyan S, Galatenko A, Galatenko V, Shkurnikov M, Tonevitsky A. miRGTF-net: Integrative miRNA-gene-TF network analysis reveals key drivers of breast cancer recurrence. **PLoS One.** 2021; 16(4):e0249424. doi: 10.1371/journal.pone.0249424
5. Пшкурников М, Нерсиян С, Оселин А, Мальцева Д, Князев Е. Различия в работе ферментов Drosha и Dicer в образцах колоректального рака и нормальной ткани толстой кишки. **Докл РАН.** 2020; 493(1):400-403. doi: 10.31857/S2686738920040216
6. Nersisyan S, Shkurnikov M, Poloznikov A, Turchinovich A, Burwinkel B, Anisimov N, Tonevitsky A. A Post-Processing Algorithm for miRNA Microarray Data. **Int J Mol Sci.** 2020; 21(4):1228. doi: 10.3390/ijms21041228
7. Shkurnikov M, Nikulin S, Nersisyan S, Poloznikov A, Zaidi S, Baranova A, Schumacher U, Wicklein D, Tonevitsky A. LAMA4-Regulating miR-4274 and Its Host Gene SORCS2 Play a Role in IGFBP6-Dependent Effects on Phenotype of Basal-Like Breast Cancer. **Front Mol Biosci.** 2019; 6:122. doi: 10.3389/fmolb.2019.00122

ՆԵՐՄԻՍՅԱՆ ԱՏԵՓԱՆ ԱՇՈՏԻ

ՄԻԿՐՈՌՆԹ-ՆԵՐԻ ԵՎ ԴՐԱՆՑ ԻՉՈՉԵՎԵՐԻ ՖՈՒՆԿՑԻՈՆԱԼ ԱԿՏԻՎՈՒԹՅԱՆ ԴԵՐԸ ՉԱՐՈՐԱԿ ՈՒՌՈՒՑՔՆԵՐԻ ՊԱԹՈԳԵՆԵԶՈՒՄ

ԱՄՓՈՓԱԳԻՐ

Հանգուցային բառեր. միկրոՌՆԹ, միկրոՌՆԹ-ի իզոմներ, քաղցկեղ, կարգավորում, Դրոշա, Դայսեր, ցանցերի վերլուծություն

ՄիկրոՌՆԹ-ները կարճ չկոդավորող ՌՆԹ-ներ են, որոնք մասնակցում են գեների էքսպրեսիայի բացասական կարգավորմանը: ՄիկրոՌՆԹ-ների աբերանտ էքսպրեսիան և թիրախավորումը վճռորոշ դեր են խաղում քաղցկեղի տարբեր տեսակների զարգացման և մետաստազների առաջացման մեջ: ՄիկրոՌՆԹ-ների վերջերս հայտնաբերված իզոմները՝ իզոմիՌ-ները, տարբերվում են կանոնական միկրոՌՆԹ-ներից 1-3 նուկլեոտիդներով համապատասխան հաջորդականության ծայրերում: Այդ չնչին տարբերությունը կարող է նշանակալիորեն ազդել միկրոՌՆԹ-ների թիրախ գեների վրա՝ նոր պատկերացում ավելացնելով գեների էքսպրեսիայի կարգավորման ուսումնասիրություններին:

Սույն հետազոտության հիմնական նպատակն է որոշել միկրոՌՆԹ-ների և դրանց իզոմների դերը մարդու չարորակ ուռուցքների պաթոգենեզում՝ կենսաինֆորմատիկական վերլուծության մեթոդների միջոցով: Նշված նպատակին հասնելու համար առաջադրվել են խնդիրներ, որոնք ներառում են իզոմիՌ-ների կենսագենետիկ վերլուծությունը, իզոմիՌ-ների թիրախների կանխատեսումը և կարգավորիչ ցանցերի կառուցումը մարդու չարորակ ուռուցքների համատեքստում: Բացի այդ, նպատակադրվել է մշակել օգտագործման համար հարմարավետ վեբ հարթակ, որը կպարունակի իզոմիՌ-ների էքսպրեսիայի և ակտիվության մասին տվյալներ:

Աշխատանքում օգտագործվել են քաղցկեղի 31 տեսակների (The Cancer Genome Atlas՝ TCGA, նախագծի տվյալների հավաքածու) պատկանող առաջնային ուռուցքների միկրոՌՆԹ-ների և դրանց իզոմների սեքվենավորման տվյալներ՝ գնահատելու պրի- և պրե-միկրոՌՆԹ-ների մոլեկուլների, համապատասխանաբար՝ Դրոշա և Դայսեր ֆերմենտներով կտրատման ճշգրտությունը: Պարզվել է, որ տարատեսակ ուռուցքային բջիջներում ճշգրտության ցուցանիշները գրեթե նույնն են. պրի-միկրոՌՆԹ-ի մշակման (Դրոշայով) արդյունքում առաջացել են 86% միատարր կտրտված «մազկալներ», մինչդեռ պրե-միկրոՌՆԹ-ների (Դայսերով) մշակումը ցուցաբերել է ավելի քիչ՝ 58% ճշգրտություն:

Այնուհետև, իզոմիՌ-ների նուկլեոտիդային հաջորդականություններն օգտագործվել են թիրախ գեների կանխատեսման համար՝ miRDB և TargetScan ծրագրերի օգնությամբ: Ստացված կանխատեսումները ենթարկվել են կո-էքսպրեսիայի վերլուծության TCGA նախագծի սեքվենավորման տվյալների հետ, ինչի արդյունքում նեղացվել է քաղցկեղ-բնորոշ թիրախների շրջանակը: Յուրաքանչյուր իզոմիՌ-ին միջինում համապատասխանում էին 209 թիրախ գեներ՝ կանխատեսված հաջորդականության

հիման վրա, և մոտ 10 անգամ ավելի պակաս գեներ՝ քաղցկեղ-բնորոշ վերլուծության դեպքում: Ընդ որում՝ նույն միկրոՌՆԹ-ի երկու իզոմերներ՝ 5՝-ծայրում մեկ նուկլեոտիդի տարբերությամբ, ունենին 6% ընդհանուր կանխատեսված թիրախներ, ինչը հաստատում է իզոմիՌ-ների՝ որպես առանձին ֆունկցիոնալ միավորներ դիտարկելու կարևորությունը:

Աշխատանքի շրջանակներում մշակվել է *isomiRTar* վեբ հարթակը, որտեղ տեղադրվել են քաղցկեղի տարբեր տեսակների հետ կապված իզոմիՌ-ների էքսպրեսիայի և թիրախավորման տվյալները: Բացի այդ, մշակվել և գործարկվել է *miRGTF-net* ալգորիթմը, որը թույլ է տալիս միկրոՌՆԹ-ների թիրախների տվյալները միավորել տրանսկրիպցիոն գործոնների ակտիվության տվյալների հետ: *ER⁺* կրծքագեղձի քաղցկեղի տրանսկրիպտոմային տվյալների նկատմամբ *miRGTF-net* ալգորիթմի կիրառման արդյունքում բացահայտվել է ութ գեներից կազմված հուսալի համախումբ (*CDK1, FOXM1, LRIG2, MSH2, PLK1, RACGAP1, RRM2* և *TMPO*), որը թույլ է տալիս բարձր ճշգրտությամբ (*ROC AUC > 0.7*) կանխատեսել հիվանդության կրկնության հավանականությունն առնվազն 5 տարվա կտրվածքով: Կարևոր է նշել, որ ստացված կանխագուշակիչ գեների զգալի մասը կարգավորվում է *E2F1* տրանսկրիպցիոն գործոնով:

Կոլոռեկտալ քաղցկեղի բջիջներում իզոմիՌ-ների էքսպրեսիայի և ակտիվության վերլուծությունը բացահայտել է *miR-148a*, *miR-192* և *miR-203a* իզոմերների հնարավոր հակառուռոցքային դերը: Հետագա հետազոտությունները ցույց են տվել *miR-148a*-ի էքսպրեսիայի նշանակալի նվազում և դրա թիրախ գեներ *ITGA5*-ի և *PRNP*-ի էքսպրեսիայի աճ հիպօքսիայի ենթարկված կոլոռեկտալ քաղցկեղի բջիջներում:

Ամփոփելով, ստացված արդյունքները զգալի նորամուծություն են ապահովում միկրոՌՆԹ-ների և դրանց իզոմերների ուսումնասիրություններում քաղցկեղային հիվանդությունների ոլորտում: Աշխատանքի շրջանակներում մշակված *isomiRTar* հարթակը և *miRGTF-net* ծրագիրը դյուրամատչելի են գիտական հանրության համար և կարող են օգտագործվել տարբեր խնդիրների լուծման նպատակով: Գործնական տեսանկյունից, ստացված կանխագուշակիչ տվյալները և գեների էքսպրեսիայի կարգավորման մեխանիզմները կրծքագեղձի և կոլոռեկտալ քաղցկեղների դեպքում նոր հեռանկարներ են բացում անհատականացված բուժման ընտրության և թերապևտիկ նոր եղանակների մշակման համար:

**THE ROLE OF MIRNA AND ISOMIR FUNCTIONAL ACTIVITY IN THE
PATHOGENESIS OF HUMAN CANCER**

SUMMARY

Key words: miRNA, isomiR, cancer, regulation, Drosha, Dicer, network analysis.

MiRNAs are short non-coding RNAs performing negative regulation of gene expression. Aberrant expression and targeting of miRNAs play a crucial role in various cancer types' development, progression, and metastasis. Recently discovered isomiRs differ from the canonical miRNAs by 1-3 nucleotides at the corresponding sequence ends. Such a difference could dramatically affect the targetome of miRNA, adding a new dimension to gene expression regulation.

The primary research purpose was to study the role of miRNAs and isomiRs in the pathogenesis of human cancer using bioinformatics methods. The research included objectives related to the analysis of isomiR biogenesis, isomiR target prediction, and the construction of regulatory networks in the context of human cancer. In addition, we aimed to develop a convenient web portal interface to the isomiR expression and targeting data.

First, we used miRNA and mRNA sequencing data of primary tumors corresponding to 31 cancer types (TCGA dataset) to infer the cleavage accuracy of pri- and pre-miRNA molecules by Drosha and Dicer enzymes, respectively. It turned out that these accuracies were practically the same in different tumor cells: pri-miRNA processing by Drosha yielded 86% of homogeneously cleaved hairpins. In contrast, pre-miRNA processing by Dicer was less accurate: 58%.

Then, isomiR nucleotide sequences were used to predict target genes using miRDB and TargetScan tools. The obtained predictions were then subjected to the co-expression analysis based on TCGA sequencing data, which resulted in the narrowed sets of cancer-specific targets. Each isomiR had 209 sequence-based predicted targets on average, while cancer-specific sets contained an order of magnitude fewer genes. In addition, two isomiRs originating from the same miRNAs and having one nucleotide difference on the 5'-end had a median of 6% common predicted targets, which confirms the importance of considering isomiRs as distinct functional units.

Next, we developed the isomiRTar web portal to conveniently present cancer-related isomiR expression and targeting data. We also developed and implemented the miRGTF-net algorithm, which integrates miRNA targeting data with transcription factors activity. As a result of applying the miRGTF-net to ER⁺ breast cancer transcriptomic data, we obtained a reliable eight-gene signature (*CDK1*, *FOXMI*, *LRIG2*, *MSH2*, *PLK1*, *RACGAP1*, *RRM2*, and *TMPO*) allowing to predict 5-year disease recurrence with ROC AUC > 0.7. Importantly, transcription factor E2F1 regulated a significant part of the predictive genes.

Analysis of isomiR expression and activity in colorectal cancer revealed possible tumor-suppressive roles of miR-148a, miR-192, and miR-203a isomiRs. Further experiments showed a significant decrease in miR-148a expression and upregulation of its target genes *ITGA5* and *PRNP* in colorectal cell lines exposed to hypoxia.

In summary, the obtained results bring much novelty to miRNA and isomiR research in the context of cancer. The scientific community could easily use the developed isomiRTar portal and the miRGTF-net program for various purposes. From a practical point of view, the inferred prognostic data and gene expression regulatory mechanisms in breast and colorectal cancers could open up new possibilities in the personalized treatment selection and the development of new treatments.

