

На правах рукописи

Арзуманян Виктория Арменовна

Молекулярный профиль опухолевой клеточной линии HepG2

1.5.8. – «Математическая биология, биоинформатика»

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2024

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича» (ИБМХ), в лаборатории интерактомики протеоформ.

Научный руководитель: кандидат биологических наук,
Поверенная Екатерина Владимировна

Официальные оппоненты: **Колпаков Федор Анатольевич**,
доктор биологических наук, Автономная
некоммерческая образовательная организация
высшего образования «Научно-технологический
университет «Сириус», Научный центр генетики и
наук о жизни, научный руководитель направления
«Вычислительная биология»

Храмеева Екатерина Евгеньевна,
доктор биологических наук, Сколковский институт
науки и технологий, Центр молекулярной и
клеточной биологии, доцент

Ведущая организация: Федеральное государственное бюджетное
учреждение науки Институт общей генетики им.
Н.И. Вавилова Российской академии наук.

Защита состоится «13» февраля 2025 г. в 13 часов на заседании диссертационного совета 24.1.172.01 (Д 001.010.01) при Федеральном государственном бюджетном научном учреждении «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича» по адресу: 119121, Москва, ул. Погодинская, д. 10, стр. 8.

С диссертацией можно ознакомиться в библиотеке ИБМХ и на сайте www.ibmc.msk.ru.

Автореферат разослан «___» _____ 2024 г.

Учёный секретарь
диссертационного совета,
кандидат химических наук

Карпова Е.А.

1. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

1.1. Актуальность работы, цель и задачи

Современные исследования в области наук о жизни включают в себя анализ генома, эпигенома, экзома, транскриптома, протеома и метаболома на уровне отдельных клеток и целых тканей. Эти методы входят в рамки омикс-анализа и, в отличие от традиционных узконаправленных подходов, позволяют в совокупности получить целостное представление об объекте. Несмотря на растущий объем данных, полученные результаты не позволяют полностью заполнить пробелы в понимании молекулярных путей. Существует высокая потребность в унифицированных подходах для обработки данных, поскольку универсального метода для комбинированного анализа различных омикс-уровней или предсказания данных одного уровня на основе другого пока не разработано.

Удобным объектом для отработки методов системного анализа и расшифровки биологических процессов являются клеточные линии (культуры), выведенные из клеток животных или человека. На настоящий момент существует 1 970 клеточных линий человека (в соответствии с биобанком ATCC), полученных как из опухолевых, так и из нормальных клеток. Клеточные культуры используются в широком спектре исследований, включая моделирование клеточных процессов, поиск лекарственных мишеней и доклинических испытаний. Популярность клеточных культур обусловлена их единообразием, относительной схожестью с исходными клетками и сохранностью молекулярных профилей в процессе культивирования. Клетки могут поддаваться многократному, а иногда и бесконечному культивированию.

Недавние работы по анализу массива образцов одной клеточной линии опровергли постулат об их единообразии. Выполненное в 2018 году масштабное тестирование противоопухолевых препаратов на 27 образцах клеточной линии MCF-7 показало наличие как минимум двух подтипов клеток, по-разному реагирующих на воздействие лекарств (Ben-David, U. et al, 2018). Годом позже было проведено исследование 14 образцов клеточной линии HeLa из 13 лабораторий, в котором показано, что образцы одной клеточной линии могут существенно отличаться между собой, формируя три подтипа клеток (Liu, Y. et al, 2019).

Для эффективного применения клеточных линий и корректного использования ранее полученных для этих объектов омикс-данных необходимо детальное понимание их молекулярной специфики. Предпочтительно начать анализ с наиболее популярных клеточных линий. Востребованность печеночной клеточной линии HepG2 в

фармакологических и токсикологических исследованиях обосновывает ее приоритетность для оценки молекулярной гетерогенности образцов внутри культуры.

Целью данной работы является построение и сравнительный анализ молекулярного портрета опухолевой клеточной линии HepG2 на основе данных, охватывающих разные уровни передачи молекулярно-генетической информации.

Для достижения указанной цели были поставлены следующие **задачи**:

1. Оценить гетерогенность клеточной линии HepG2 на основе геномных, эпигеномных, транскриптомных и протеомных данных.
2. Изучить взаимосвязь между копийностью генов и соответствующими продуктами реализации молекулярно-генетической информации.
3. Определить элементы регуляции, обеспечивающие взаимодействие между различными уровнями омикс-данных.
4. Оценить общие и частные характеристики опухолевой клеточной линии HepG2 и других печеночных культур.

1.2. Научная новизна работы

В работе впервые сформирован целостный молекулярный портрет опухолевой клеточной линии HepG2 на основе результатов широкомасштабных анализов генома, метилома, транскриптома и протеома. Результаты работы позволят корректно использовать данную линию в экспериментах и минимизировать ошибки при интерпретации результатов.

1.3. Теоретическая и практическая значимость работы

На январь 2024 года по запросу «HepG2» в базе данных PubMed находится 40 059 статей, что является четвертым результатом среди всех клеточных культур и первым среди печеночных клеточных линий. Клеточная линия HepG2 имеет опухолевую природу и является моделью гепатобластомы, однако ее часто используют и как модель нормальных гепатоцитов, несмотря на доказанные различия в ряде молекулярных механизмов (Wilkening et al., 2003). Детальное описание и сопоставление молекулярных профилей, полученных для различных образцов HepG2, позволит использовать ее более корректно при дизайне экспериментов и будет способствовать улучшению качества и точности научных выводов.

Полученные сведения об общих и частных отличиях печеночных клеточных культур дают основу для более детального изучения регуляторных механизмов, определяющих тканеспецифические свойства. Выполненный анализ взаимосвязей молекулярных уровней в рамках исследования одного образца открывает возможности

для направленного изучения скрытых процессов регуляции экспрессии и трансляции при масштабировании такого мультиомного профилирования.

1.4. Личный вклад автора

Автором были сформулированы концепции анализа и обработки данных молекулярного профилирования, выполненных в рамках реализации проектов в ИБМХ и опубликованных в открытом доступе. Самостоятельно проведен поиск и биоинформатический анализ необработанных данных с применением новых интеграционных протоколов и актуальных алгоритмов; выполнены статистические тесты и визуализация результатов. Автор провел поиск и анализ опубликованной информации в научных статьях при написании литературного обзора и изложении результатов.

1.5. Основные положения, выносимые на защиту

1. Для клеток HepG2 характерными стабильными абберациями генома являются t(1;21)(p36;p11), +2, +6(p11.1-p13.3, q21-q24.3), +17(q12qter), +16(p12.2-p13.2), +16(p12.2-p13.2), del(9)(q12), +20, +20, del(Y)(p11.32, q11.23qter), del(SATp).
2. Образцы клеточной линии HepG2 характеризуются высокой степенью сходства транскриптомных профилей. Локальные абберации касаются генов, ответственных за метаболизм ксенобиотиков, и ассоциируются с практиками ведения клеточных линий в отдельных лабораториях.
3. Печеночные клеточные линии (HepG2, Huh7, Hep3B) имеют схожие профили экспрессии генов, которые значительно отличаются от профилей клеточных линий другой природы.

1.6. Степень достоверности и апробация результатов

Полученные в ходе выполнения работы результаты верифицированы должным образом согласно современным научным представлениям. В работе были использованы вычислительные алгоритмы, рекомендованные научным сообществом, в частности GATK и C-HPP, а также актуальная информация из курируемых баз данных, включая Ensembl, NCBI, UniProt и др. Обсуждение результатов проводилось с учетом новейших достижений в области биологических и медицинских наук. Научные положения и выводы, изложенные в диссертации, обоснованы и подтверждены фактическим материалом.

Основные положения диссертационной работы были представлены в виде докладов и обсуждены на конференциях: симпозиуме «XXVIII Биоинформатика и компьютерное конструирование лекарств» (Москва, 2021), «Высокопроизводительное секвенирование в геномике» (Новосибирск, 2022), «Вычислительная биология и искусственный интеллект для персонализированной медицины» (Москва, 2022),

симпозиуме «Белки и пептиды» (Сочи, 2022), «Вычислительная молекулярная биология» (Москва, 2023), Конгрессе международной организации «Протеом человека» (HUPO 2023, Пусан, Южная Корея), на конференции «Мультиомные технологии для прецизионной медицины» (Москва, 2023).

1.7. Публикации

По теме диссертации опубликовано 15 работ, из которых 8 статей в рецензируемых научных журналах и 7 публикаций в трудах конференций.

1.8. Объем и структура диссертации

Диссертация изложена на 184 страницах машинописного текста, содержит 11 таблиц и 34 рисунка. Состоит из следующих глав: введение, обзор литературы, материалы и методы, результаты и их обсуждение, заключение, выводы, а также включает разделы список сокращений, словарь терминов, список литературы, благодарности и 9 приложений.

2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Данные

Для анализа клеточной линии HepG2 использовались как данные, полученные в ИБМХ для одного образца в 2020 и 2022 годах (далее именуемые собственными данными, т.к. большая часть была получена в лаборатории интерактомики протеоформ), так и общедоступные данные молекулярного профилирования (рисунок 1).

Кариотип	Геном	Эпигеном	Транскриптом	Транслятом	Протеом
Собственные данные (проекты 2020 и 2022 годов)					
Кариотипирование проводилось на 20 метафазных пластинках	 WGS, Illumina (2 повтора)	 WGBS, Illumina (1 повтор)	  direct RNA, ONT (5 повторов) RNAseq, Illumina (2 повтора)	 Polysome-seq, ONT (1 повтор)	 ВЭЖХ-МС/МС (2 повтора)
Опубликованные данные					
 4 статьи	 >100 млн прочтений, длина чтений 100 п.н. WGS, Illumina (6 наборов данных)	 WGBS, Illumina (2 набора данных)	 >20 млн прочтений, длина чтений >60 п.н. RNAseq, Illumina (40 наборов данных)	 Нет данных	 ВЭЖХ-МС/МС (14 наборов данных)

Рисунок 1. Сводная таблица анализируемых данных для клеточной линии HepG2.

Собственные данные включали результаты полногеномного (WGS), бисульфитного (WGBS), транскриптомного (RNA-seq и direct RNA-seq), транслятомного (Polysome-Seq) и протеомного (ВЭЖХ-МС/МС) профилирований (таблица 1). Секвенирование нуклеиновых кислот выполнялось с помощью технологии Illumina (NovaSeq 6000). Также транскриптомные и транслятомное профилирования выполнялись с использованием технологии Oxford Nanopore Technology (ONT), где в режиме direct RNA-seq был

выполнен анализ пяти повторов, а Polysome-Seq (в режиме cDNA) – один повтор.

Таблица 1. Перечень собственных данных омикс-профилирования клеток HepG2.

Тип профилирования	Название образца	ID набора данных
2020 год		
WGS	G-IBMC.1	PRJNA765908
WGBS	M-IBMC.1	
RNA-seq	T-IBMC.1.1, T-IBMC.1.2, T-IBMC.1.3	
direct RNA-seq	T-IBMC.1.O1, T-IBMC.1.O2, T-IBMC.1.O3, T-IBMC.1.O4, T-IBMC.1.O5	
ВЭЖХ-МС/МС	P-IBMC.1	Mendeley Data: yrmd8ygncl/1
2022 год		
WGS	G-IBMC.2	PRJNA1151259
RNA-seq	T-IBMC.2.1, T-IBMC.2.2, T-IBMC.2.3	PRJNA956723
Polysome-seq	Tr-IBMC.1	PRJNA972889
ВЭЖХ-МС/МС	P-IBMC.2	PXD056076

*WGS – полногеномное секвенирование, WGBS – полногеномное бисульфитное секвенирование, RNA-seq – транскриптомное секвенирование, direct RNA-seq – секвенирование РНК, ВЭЖХ-МС/МС – панорамное протеомное профилирование.

Поиск опубликованных данных проводился по базам данных NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra>) и PRIDE (<https://www.ebi.ac.uk/pride/archive>). После контроля качества были определены критерии: данные должны быть представлены «сырыми» файлами и получены для необработанных клеток.

Для иных печеночных клеточных линий по аналогичным вышеуказанным параметрам (рисунок 1) соответствовали опубликованные данные в виде двух наборов данных WGS и 21 набор данных RNA-seq для клеток Huh7, а также 12 наборов данных RNA-seq для клеток Hep3B (данные WGS отсутствуют).

Был также проведен поиск и аналогичный анализ транскриптомных данных первичных гепатоцитов (ПГ) и непеченочных клеточных линий, включая MCF-7 и MDA-MB-231 (рак молочной железы), HeLa (рак шейки матки), HCT116 (рак толстой кишки), LNCaP (рак простаты), HEK293T (почки), HaCaT (эпидермис) и лейкемические линии K562 и THP-1. Поисковые критерии были аналогичными. Для ПГ был найден 31 набор данных, для непеченочных линий в среднем доступными были 13 наборов данных.

2.2. Обработка омикс-данных

2.2.1. Полногеномное секвенирование

Оценка качества данных полногеномного секвенирования осуществлялась с помощью программы FastQC (версия 0.12.1). Удаление чтений низкого качества и адаптеров проводилось с использованием TrimGalore (версия 0.6.7). Отфильтрованные fastq файлы выравнивались на референсный геном человека hg38 с использованием алгоритма Bowtie2 (версия 2.4.4). В результате были получены bam-файлы для дальнейшего анализа.

Оценка копийности хромосом была выполнена с помощью программы mosdepth (версия 0.3.9), которая генерировала bed-файл с усредненными данными по покрытию для окон размером 10 000 нуклеотидов. Данные о покрытии накладывались на хромосомы с помощью библиотеки R – karyoploteR.

Для детекции несинонимичных однонуклеотидных замен (нсОНП) и инделов использовались рекомендации GATK best practices от Broad Institute (<https://gatk.broadinstitute.org/hc/en-us>). Однонуклеотидные замены были обнаружены с помощью GATK Haplotypecaller (версия 3.7), результаты были перекалиброваны с использованием GATK VQSР с наборами обучающих данных. Аннотация нсОНП и инделов проводилась с помощью программы SnpEff (версия 4.3).

2.2.2. Эпигеном

Оценка качества прочтений бисульфитного секвенирования была проведена с помощью программы FastQC (версия 0.12.1), обрезка чтений низкого качества выполнялась с помощью TrimGalore (версия 0.6.7). Для идентификации CpG участков был использован протокол анализа программы Bismark (версия 0.24.2), данные были выровнены на референсный геном человека hg38.

2.2.3 Транскриптомное и транслятомное профилирование

Качество данных секвенирования, полученных с помощью технологии Illumina (формат fastq), проверяли с использованием программы FastQC. Корректировка чтений низкого качества осуществлялась с помощью программы TrimGalore (версия 0.6.7). Анализ экспрессии генов и транскриптов выполнялся с использованием Salmon (версия 1.6.0), экспрессия генов и транскриптов рассчитывалась в TPM (транскрипт на миллион).

Данные прямого секвенирования РНК (direct RNA-seq), полученные с использованием технологии Oxford Nanopore (формат файлов fast5), конвертировали в формат fastq с помощью программы распознавания азотистых оснований Guppy (версия 6.2.1). Полученные fastq файлы выравнивались с помощью Minimap2 (версия 2.24). Экспрессия генов и транскриптов также рассчитывалась с использованием программы

Salmon (версия 1.6.0). Данные транскриптомного профилирования анализировались по тому же протоколу, что и транскриптомные данные прямого секвенирования РНК.

2.2.4 Эпитранскриптом

Анализ данных direct RNA-seq для поиска сайтов модификации мбА проводился с использованием Nanopolish (версия 0.14.0) и алгоритма mбAnet (версия 2.0.2). Участки с коэффициентом модификации менее 0,9 исключались из анализа. Использовался термин «возможные участки модификаций мбА» в соответствии с терминологией mбAnet.

2.2.5. Протеомное профилирование

Протеомные данные преобразовывались в формат mgf с помощью MSConvert (версия 3) и анализировались с помощью программного пакета SearchGUI (версия 3.2.20) с использованием поискового алгоритма X!Tandem. Карбамидометилирование цистеина указывалось как фиксированная модификация, окисление метионина – как переменная в случае данных ИБМХ. Для опубликованных данных определяли параметры, исходя из описания получения наборов данных. Были использованы результаты идентификации, удовлетворяющие уровню показателя достоверности $FDR \leq 1\%$. Для полуквантитативного анализа был использован индекс нормированных спектральных коэффициентов содержания (NSAF). Референсная база данных включала канонические и альтернативно сплайсированные последовательности белков из SwissProt. Кастомизированная база данных для HerG2 была создана на основе результатов полногеномного и транскриптомного анализов.

2.3. Статистическая обработка

Обработка данных проводилась в среде Linux (Ubuntu 22). Для статистического анализа и визуализации использовались R (версия 4.1) и Python (версия 3.11). Анализ дифференциальной экспрессии генов выполнялся с использованием пакета DESeq2, графики вулкана – EnhancedVolcano. Для анализа обогащения наборов генов применялся ClusterProfiler. Источниками биологических данных были WikiPathways и Gene Ontology (GO) с акцентом на молекулярные функции. «Эффект партии» корректировался с помощью pyComBat.

Аннотация регионов метилирования выполнялась с пакетом annotatr с использованием аннотации «hg38_basicgenes». Файл GTF для аннотации генома был загружен из базы данных Ensembl (GRCh38, версия 103), анализ ограничивался белок-кодирующими генами (БКГ) и стандартными хромосомами (1-22, X и Y).

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

3.1. Гетерогенность клеточной линии HepG2

3.1.1. Анализ кариотипа по цитогенетическим данным

Для оценки стабильности кариотипа клеточной линии HepG2 были проанализированы результаты кариотипирования, выполненного цитогенетическими методами как собственного образца (20 метафазных пластинок), так и опубликованных данных. Собственный образец характеризуется следующей общей формулой кариотипа: 49~55, XY, +Y[6], add(1)(q44)[6], t(1;21)(p36;p11)[20], +2[19], +6[20], der(6;?)(q13;?)[20], +16[20], add(16)(p13)[17], +20[20], +20[17], add(21)(p11)[5], +1~2mar[cp20].

В большинстве опубликованных данных (Zhou et al., 2019; Simon et al., 1982; Chen et al., 1993; Wong et al., 2000), как и в собственных результатах, отмечены транслокация между короткими плечами хромосом 1 и 21, трисомия хромосом 2 и 6 и тетрасомия хромосомы 20. Трисомии хромосом 16 и 17 были выявлены в двух результатах кариотипирования, однако в собственных данных трисомия хромосомы 17 не наблюдалась. В каждом кариотипировании были обнаружены одна-две маркерные хромосомы, состоящие из частей других хромосом.

3.1.2. Оценка ploидности генома

На основе данных полногеномного секвенирования в виде покрытия и соответствующего процентного содержания GC-состава соответствующих областей (Zhou et al., 2019) была получена анеуплоидная карта с высоким разрешением для детализации кариотипа клеток HepG2.

По собственным результатам секвенирования клеточной линии HepG2 (G-IBMC.1 и G-IBMC.2) были выявлены пять кластеров (рисунок 2), соответствующих различной копийности генов (CN). Кластерам присвоены значения на основе соотношения их среднего охвата к охвату нижнего кластера (CN = 1).

Кластер CN2,5, вызванный дублированием участка генома на одной из хроматид при стандартной копийности или делецией в случае трисомии, значительно отличается между двумя наборами данных одного образца, полученными в 2020 году (G-IBMC.1) и 2022 году (G-IBMC.2). В G-IBMC.1 было выявлено 2 966 уникальных генов в кластере CN2,5, что в 2,4 раза больше, чем в G-IBMC.2 (1 222 гена). Гены с уникальной копийностью для образца G-IBMC.1 участвуют в регуляции ответа на повреждения ДНК, тогда как в G-IBMC.2 связаны с метаболизмом ретиноидов. Практически все отличные гены между образцами «восстанавливают» свою копийность до нормальной в образце G-IBMC.2.

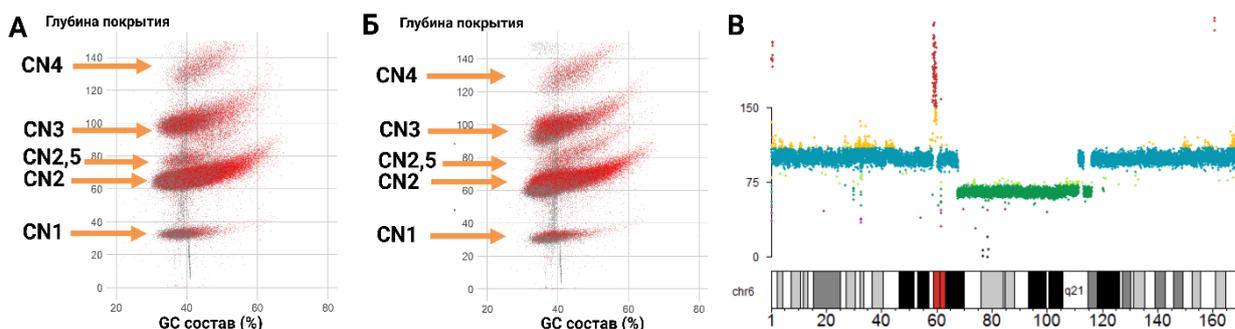


Рисунок 2. Анеуплоидные карты для двух образцов клеточной линии HepG2. (А, Б) анеуплоидная карта, составленная на основании GC-состава и покрытия ридами. Красным цветом обозначены участки БКГ; (В) распределение покрытия ридами на примере хромосомы 6.

3.1.3. Кариотипирование на основе полногеномных данных

Для подтверждения и уточнения результатов кариотипирования цитогенетическими методами оценивали копийность хромосом согласно распределению глубины покрытия. Для этого полученные данные о покрытии на 10 000 нуклеотидов и копийности этих участков были наложены на координаты хромосом с помощью пакета karyoploteR (Gel et al., 2017). Для собственных образцов покрытие на хромосомы оказалось идентичным, несмотря на разницу по данным анеуплоидных карт (5 107 БКГ отличались по копийности).

Иллюстративным примером преимущества интеграции двух типов анализа является хромосома 6 (рисунок 2В). По данным цитогенетического анализа хромосома 6 вся находится в триплоидном состоянии, а при анализе геномных данных выявляется ее неполная трисомия – часть q-плеча (q12-q21) остается в диплоидном состоянии.

Суммарно во всех образцах также выявлены: потеря региона q12 хромосомы 1, трисомия хромосомы 2, частичная трисомия хромосомы 6 (p11.1-p13.3, q21-q24.3), трисомия части длинного плеча 17 (после региона q12) и частичная тетрасомия хромосомы 16 (p12.2-p13.2), делеция части хромосомы 9 (q12), коротких плеч у SAT-хромосом – 13, 14, 15, 21 и 22, тетрасомия хромосомы 20 и делеция части короткого (p11.32) и длинного плеч Y-хромосомы. (после региона q11.23).

Ни в одном из опубликованных образцов не обнаружено высокой копийности в начале короткого плеча хромосомы 1, специфичной для образцов G-IBMC.1 и G-IBMC.2, хотя во всех опубликованных образцах также наблюдалась потеря части короткого плеча хромосомы 21. Еще одним отличием собственных данных от опубликованных является отсутствие повышенной копийности короткого плеча хромосомы 10 в последних. Во всех случаях нормальное диплоидное состояние без видимых aberrаций сохранено у хромосом 3, 4, 5, 7, 8, 11, 12 и 18.

Данные копийности хромосом позволили расширить информацию о кариотипе клеточной линии HepG2 и выявить делецию 25 БКГ: POTED (хромосома 21); OR4M2B и OR4N4C (хромосома 15); USP17L2, USP17L3, USP17L8, USP17L7 (хромосома 8) и USP17L23 (хромосома 4); CT47A12, CT47A11, CT47A10, CT47A9, CT47A8, CT47A7, CT47A6, CT47A5, CT47A4, CT47A3, CT47A2, CT47A1, CXorf51A, GAGE12B, OPN1MW, OPN1MW2 и OPN1MW3 (хромосома X).

Согласно результатам анализа покрытия хромосом ридами, 86% генома остаются стабильными во всех восьми наборах данных: в CN1 находится 5% генома, в CN2 – 64%, CN3 – 13%, CN4 – 3% и CN5 – 1%. Вариативная по копийности часть генома включала хромосомы 1, 10, 19 и 22 для собственного образца и хромосомы 2, 6 и 16 в опубликованных данных.

3.1.4. Разнообразие генетических вариантов на уровне генома

Сравнительный анализ двух повторов полногеномного секвенирования собственного образца клеточной линии HepG2 показал совпадение только 52,5% нсОМП, что составляет 8 524 общих варианта между двумя образцами, имеющих разный период культивирования. Кроме того, общими оказались только 25,7% инделов – 255 вариантов. Несмотря на то что образцы происходят из одной культуры, обнаруженные различия в генетических вариантах подчеркивают важность генетического мониторинга и контроля клеточных линий.

В результате анализа полногеномного секвенирования клеточной линии HepG2 во всех восьми наборах данных было обнаружено 29 560 генетических вариантов, из которых 27 618 классифицировались как нсОМП и 1 942 как инделы. Большинство вариантов (54% нсОМП и 73% инделов) были выявлены в одном или двух образцах. Генетические варианты, обнаруженные в клеточной линии HepG2, демонстрируют значительное разнообразие, и многие из них связаны с онкогенезом и нарушением клеточных процессов. Большинство генетических вариантов (11 159 нсОМП и инделов), выявленных более чем в половине образцов, являются постоянными для клеток HepG2. При этом 20% уникальных генетических вариантов относятся к китайским образцам, включая два нсОМП — rs429358 (APOE) и rs1346044 (WRN). К этим вариантам проявляется повышенный интерес в контексте изучения старения, поскольку ассоциируются с когнитивными способностями и рисками развития возрастных заболеваний (Zins et al, 2015; Kuningas et al, 2006).

3.1.5. Метилом

Метилирование ДНК – важная эпигенетическая модификация, влияющая на активность генов, однако его воздействие на экспрессию зависит от локализации в гене. Метилирование в промоторах обычно подавляет экспрессию, предотвращая связывание факторов транскрипции (Lande-Diner et al, 2005). В других случаях метилирование на расстоянии 1-5 кб от промотора, в 5'UTR, 3'UTR, интронах и экзонах может как повышать, так и понижать активность генов.

В результате анализа метиломных данных для собственного образца клеточной линии HepG2 было обнаружено 23 637 476 локусов метилирования, из которых 2 796 515 относятся к 19 017 БКГ. Таким образом, метилирование не было обнаружено в 895 БКГ. Для дальнейшего анализа рассчитывался средний процент метилирования для каждой из областей гена. Если уровень метилирования ДНК на участке выше 80%, то он классифицируется как участок с высоким уровнем метилирования, менее 20% – как неметилированный, а остальные – как частично метилированные (Luo et al., 2020).

Так как для собственных данных наблюдалось низкое покрытие на референсный геном (22X) при рекомендуемых 30X, дополнительно были проанализированы опубликованные данные, чтобы по усредненным результатам провести дальнейший анализ взаимодействий. В базе данных NCBI SRA было найдено два проекта – SRP346999 и SRP012412 со средним, не дублирующим покрытием данных 29X. В опубликованных данных было обнаружено от 15 до 25 млн сайтов метилирования.

При объединении сайтов метилирования собственных и опубликованных данных были получены схожие цифры – определено 2 937 854 локусов, относящихся к 19 037 генам. По совокупности для клеток HepG2 выявлено 718 генов со стабильно высоким уровнем метилирования во всех обнаруженных для них регионах. Эти гены связаны с воспалением, иммунным ответом, метаболическими нарушениями и дифференцировкой клеток, особенно в развитии тканей и органов, согласно анализу с помощью WikiPathway. Для дальнейшего анализа с целью устранения «шума» и снижения вероятности ошибок при обнаружении сайтов метилирования были использованы совпадающие данные об уровне метилирования в областях 18 091 БКГ между всеми анализируемыми наборами данных.

3.1.6. Транскриптом

Транскриптомный анализ 42 наборов данных клеточной линии HepG2, среди которых два набора данных получены в ИБМХ, показал, что в среднем в каждом образце экспрессировались 14 609 БКГ и 37 844 транскриптов (TPM > 0). Между образцами выявлена положительная корреляция по критерию Спирмена по данным экспрессии генов

($\rho = 0,88$) и транскриптов ($\rho = 0,73$) (рисунок 3А, Arzumanyan et al., 2023).

Иерархическая кластеризация данных транскриптомного профилирования клеточной линии HepG2 выявила два кластера. В первом (желтый) преобладали данные из Китая и Израиля, во втором (зеленый) — образцы из США, Японии, Южной Кореи, Турции, Колумбии, Италии и России (ИБМХ).

Согласно анализу 2 824 дифференциально экспрессирующихся генов, между кластерами обнаружены различия в путях метаболизма, таких как «Окисление через цитохром P450» и «Метапуть биотрансформации фазы I и II». Эти пути важны для клеточных линий, особенно для HepG2, используемых в исследованиях метаболизма и токсичности лекарств. Во второй группе образцов отмечается повышенная экспрессия генов, ответственных за метаболизм лекарств и детоксикацию, таких как CYP3A5, CYP3A7 и CYP20A1, а также увеличение экспрессии генов SULT, участвующих в сульфатировании субстратов. Эти изменения указывают на активацию процессов детоксикации и метаболизма ксенобиотиков.

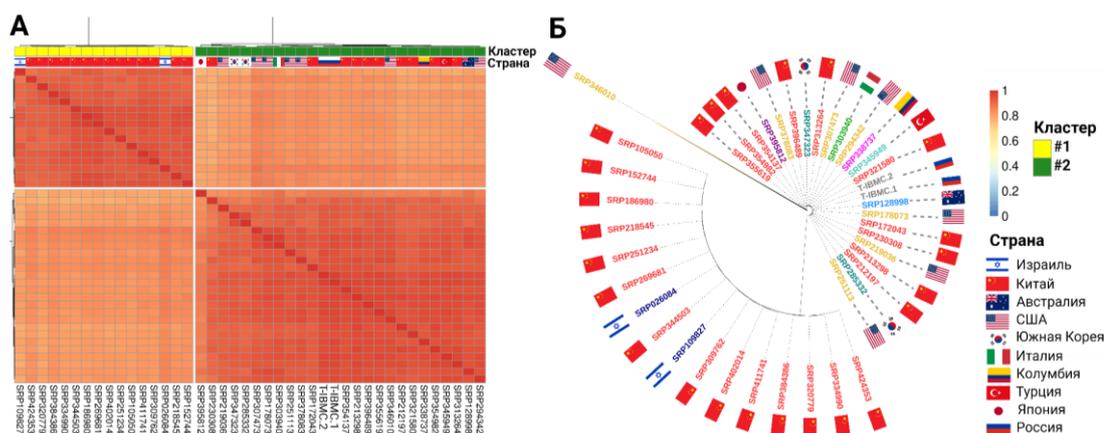


Рисунок 3. (А) Матрица корреляции между образцами клеточной линии HepG2 на уровне экспрессии генов. (Б) Филогенетическое дерево, построенное на основе модели эволюции последовательностей.

В кластере #1 наблюдается повышение активности пути «Гликолиз в состоянии старения клеток» ($FDR = 0,03$). Старение клеток представляет собой процесс, при котором старые или поврежденные клетки перестают делиться в ответ на различные виды стресса. Дополнительно в кластере #2 наблюдается активация сигнальных путей клеточного ответа, таких как MAPK, Wnt/ β -catenin и p53. Эти изменения свидетельствуют о влиянии на онкогенез и метастазирование клеток печени.

Для оценки эволюции наборов данных клеточной линии HepG2 были проанализированы данные RNA-seq на наличие ОНП и исключены вставки и делеции. Филогенетическое дерево было построено на данных о 3 604 396 ОНП, которые суммарно обнаружены во всех 42 наборах данных. «Предком» оказался образец из проекта

SRP346010 (Медицинский университет Иллинойса, 2021). Образцы кластера #2, включая данные из ИБМХ и ряда стран, были ближе к «предку». Кластер #1 с образцами из Китая и Израиля, оказался дальше и происходил из кластера #2 (рисунок 3Б).

3.1.7. Эпитранскриптом

На основе данных прямого секвенирования РНК (direct RNAseq), выполненного в пяти повторах, был проведен поиск сайтов модификации N6-метиладенозина (мбА) (Arzumaniyan et al., 2023). Суммарно было идентифицировано 3 968 возможных сайтов мбА модификаций для 2 851 транскрипта, соответствующего 1 396 генам, где только 250 (6,3%) сайтов, соответствующих 132 генам (9,46%), оказались общими для всех повторов. Гены с обнаруженными мбА сайтами ассоциированы с процессами убиквитинирования. Согласно предыдущим исследованиям, метилирование РНК (мбА) играет значимую роль в убиквитинировании гистонов при злокачественных новообразованиях.

3.1.8. Транслятом

В течение последнего десятилетия резко возросло понимание роли трансляции как ключевого регуляторного узла реализации генов. Транслятом является промежуточным слоем между транскриптомом и протеомом. Анализ результатов транслятомного профилирования, выполненного с использованием технологии ONT методом полисомного профилирования (Polysome-seq), выявил экспрессию 10 430 генов и 22 513 транскриптов при $TPM > 0,1$, в среднем было выявлено два транслята на ген. В сравнении с результатами транскриптомного профилирования не было детектировано экспрессии 2 172 генов.

3.1.9. Протеом

В результате протеомного профилирования 16 наборов данных клеточной линии НерG2, включая два повтора из ИБМХ, было обнаружено в среднем 7 602 белка и 558 сплайс-вариантов. Во всех наборах данных идентифицировано 204 белка, 351 обнаружен в 50% образцов, а 2 734 — один-два раза. Белки, присутствующие во всех образцах, участвуют во взаимодействии с нуклеиновыми кислотами и малыми органическими соединениями (GO:0003723, FDR < 0,001; GO:0003676, FDR < 0,001).

Для собственного образца клеточной линии НерG2 на протеомном уровне идентифицировано 5 950 белков и 273 сплайс-варианта. Коэффициент корреляции Спирмена между пятью образцами ИБМХ по данным полуколичественной оценки содержания белка NSAF составил 0,52 (рисунок 4).

Как и ожидалось, анализ корреляции данных показал, что сходство протеомного профиля образцов ниже, чем на уровне транскриптома. В среднем значение корреляции

Спирмена между данными составляло 0,51 (рисунок 4). В результате были выделены два кластера: в первый вошли образцы из России, Швеции, Дании, Испании, Германии и Китая, во второй — образцы из России, Израиля, Германии, США и Китая.

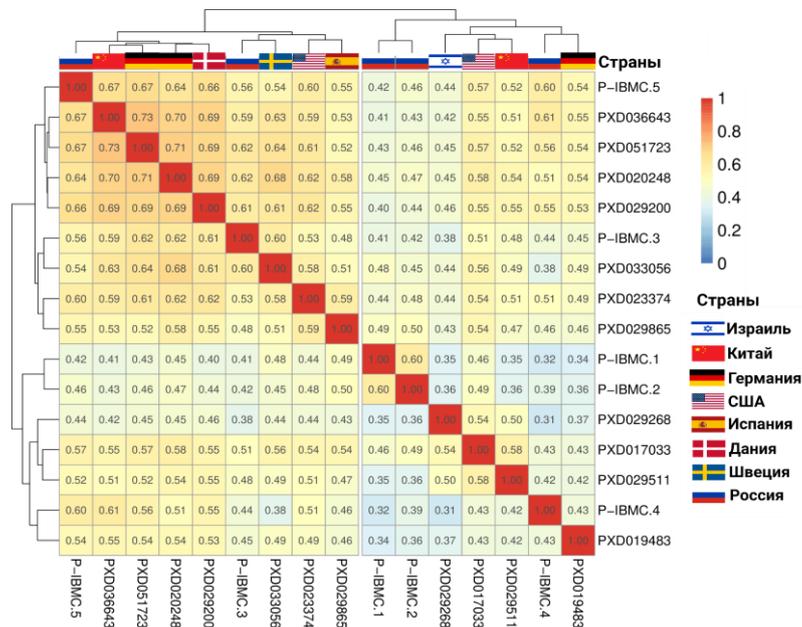


Рисунок 4. Матрица корреляции между образцами клеточной линии HerG2 на уровне экспрессии белков.

Анализ рассмотренных нами уровней передачи генетической информации для клеточной линии HerG2 показывает ожидаемое увеличение гетерогенности от геномного уровня к протеому. На транскриптомном уровне наблюдалось формирование кластеров по географическому происхождению — отдельно для образцов из Китая и Европы. Внутри этих кластеров дополнительных различий не обнаружено, что свидетельствует о стабильности клеточной линии в рамках каждого региона.

3.2. Взаимосвязь между копийностью гена и продуктами молекулярно-генетической информации

Копийность генов является одним из структурных вариантов генома. Такие геномные изменения играют роль в регуляции генетической активности и могут влиять на фенотип. Зависимость между дозой гена и молекулярными событиями, такими как уровень метилирования генов или их экспрессия, позволяет понять, как изменения в количестве копий генов влияют на клеточные процессы. При сопоставлении копийности генов и их метилирования не выявлено значимых изменений. Для генов с участками копийности CN2,5 наблюдается эпигенетическая компенсация (рисунок 5А). Не было выявлено взаимосвязей между копийностью генов и экспрессией ($\rho = 0,16$). При повышенной копийности генов вдвое увеличивается доля средне- и высокоэкспрессирующих генов ($TPM > 20$), которая практически неизменна при любом варианте увеличенной ploidy (рисунок 5Б).

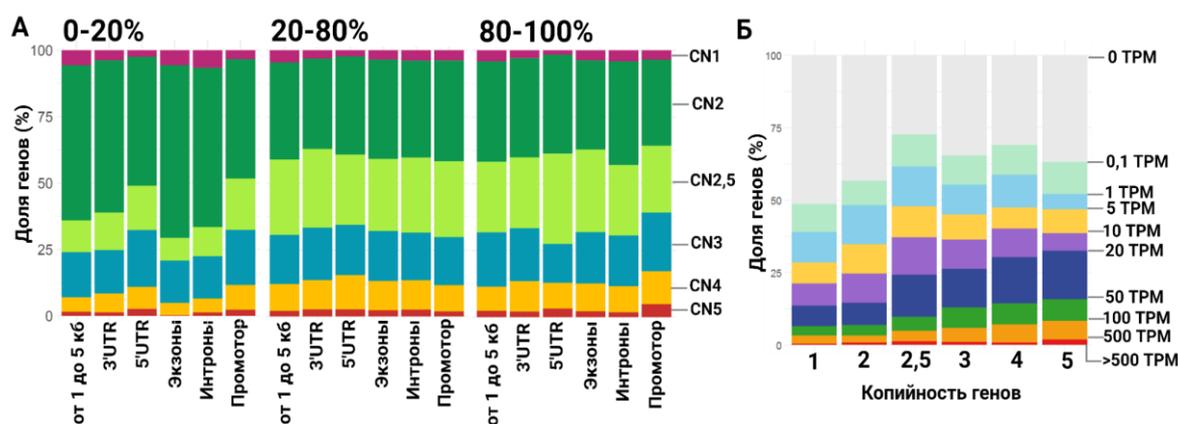


Рисунок 5. А) Распределение генов с разным уровнем метилирования при разной копияности. По оси X представлены регионы генов. Б) Взаимосвязь копияности гена и его экспрессией.

3.3. Взаимосвязи между уровнями реализации генетической информации

3.3.1. Зависимость экспрессии генов от уровня метилирования

Метилирование цитозина в положении CpG является одним из механизмов эпигенетического контроля экспрессии генов. Считается, что метилирование играет ген- и контекст-зависимую роль в регулировании экспрессии. В случае собственных данных высокий процент метилирования в регионах 3'UTR и от 1 до 5 к.б. выше промотора связаны с увеличением экспрессии соответствующих генов. В то же время при увеличении процента метилирования в промоторах и 5'UTR уровень экспрессии соответствующих генов снижается (рисунок 6).

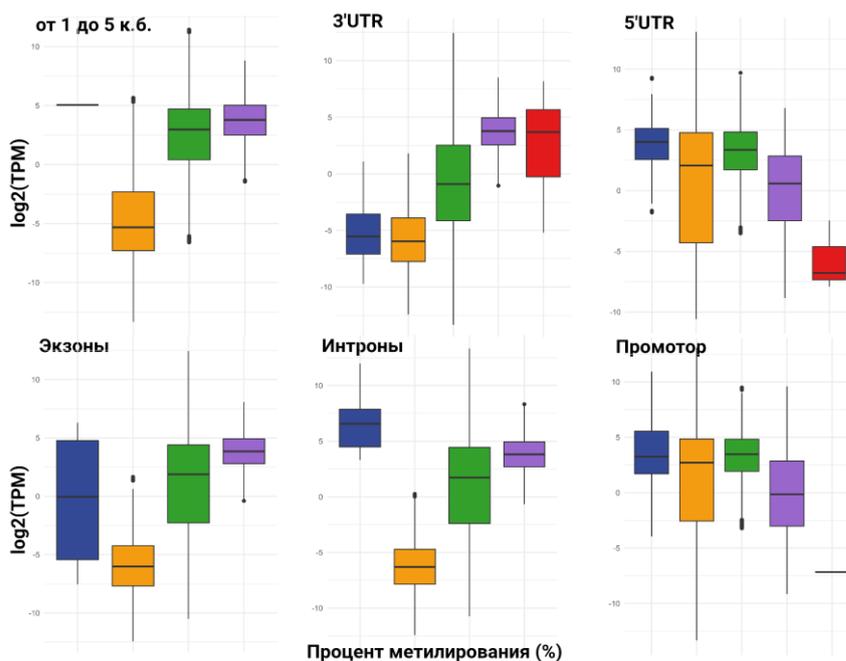


Рисунок 6. Зависимость экспрессии генов от уровня метилирования их отдельных регионов: от 1 до 5 к.б. – регион 1-5 к.б. выше промотора гена; 3'UTR и 5'UTR – 5' и 3'-нетранслируемые области. Слева направо: синим цветом показан процент метилирования от 0 до 1%, желтым – 1-20%, зеленым – 20-80%, фиолетовым – 80-99% и красным – 99-100%.

3.3.2. Поведение экспрессии генов с возможными сайтами модификации мбА

Для анализа влияния мбА модификаций на трансляцию генов были сопоставлены данные транскриптома (экспрессия 12 602 генов, выраженная в TPM > 0,1), транслятома (экспрессия 10 430 генов, выраженная в TPM > 0,1) и протеома (уровень белков, кодируемых 5 901 геном, выраженных в NSAF > 0). Гены с возможными мбА сайтами, на трансляционном уровне составили 1 355 из 1 396, а на протеомном уровне только 835 генов (Arzumanian et al., 2023). В результате сопоставления данных с учетом и без генов с выявленными модификациями была обнаружена высокая корреляция между транскриптомным и трансляционным уровнями ($\rho = 0,85$). При этом корреляция между протеомным и другими уровнями была несколько ниже, что может свидетельствовать о сложности перевода уровня mRNA в функциональный белок или других регуляторных процессах (рисунок 7).

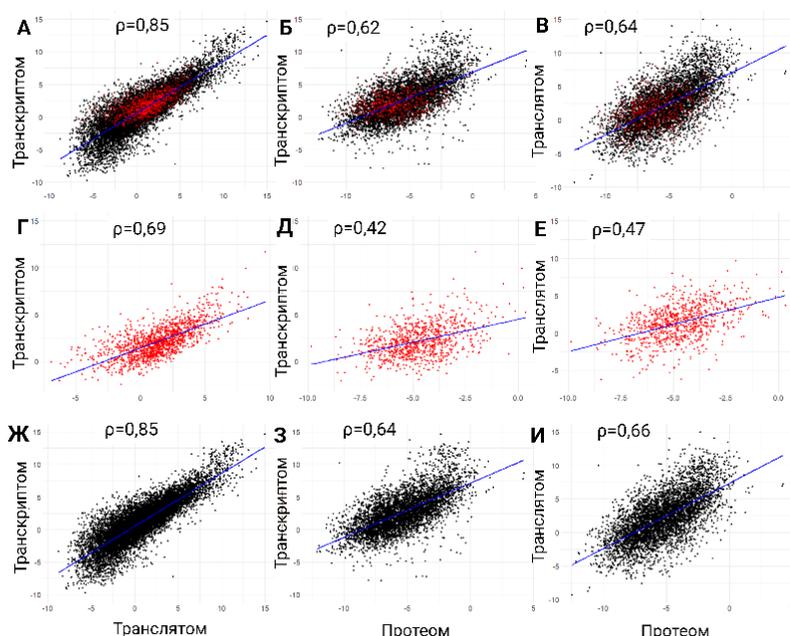


Рисунок 7. Графики корреляции между транскриптомным, трансляционным и протеомным уровнями. А, Б, В) всех генов между тремя уровнями реализации генетической информации. Красным цветом отмечены гены с мбА модификациями; Г, Д, Е) всех генов с возможными мбА модификациями; Ж, З, И) всех генов за исключением генов с возможными мбА модификациями. Уровни экспрессии приведены в логарифмических значениях: $\log_2(\text{TPM})$ для транскриптома и транслятома и $\log_2(\text{NSAF})$ для протеома.

Большинство идентифицированных потенциальных сайтов модификации мбА уменьшают трансляцию связанных с ними генов. Это подтверждается анализом корреляции генов, исключая 1 396 модифицированных генов, где значение корреляции Спирмена между транскриптомом и транслятомом оказалось самым высоким из всех проведенных сравнений (рисунок 7Ж). Корреляции между транскриптомом и протеомом, а также транслятомом и протеомом остаются неизменными (рисунок 7З и 7И) по сравнению со всеми генами.

3.4. Общее и частное в молекулярных профилях печеночных клеточных линий (HepG2, Huh7 и Hep3B)

Для изучения биологии и патологии печени широко используются *in vitro* модели клеток, такие как HepG2, Huh7 и Hep3B. Тем не менее информация о схожести этих клеточных линий на различных уровнях передачи генетической информации остается ограниченной.

В результате построения молекулярных портретов для Huh7 и Hep3B были выявлены следующие общие черты. По данным цитогенетических исследований, общими с клеточной линией HepG2 хромосомными аномалиями являются трисомии хромосом 2 и 6. В одной из статей была также отмечена тетрасомия хромосомы 20 для клеточной линии Huh7 (Zhou et al., 2020), аналогичная той, что наблюдалась в клетках HepG2. Суммарно наблюдается большее количество aberrаций в клетках Huh7 и Hep3B по сравнению с клетками HepG2.

Полногеномные данные опубликованы только для клеточной линии Huh7, анализ которых показал, что 49,4% генома было с базовой копийностью. В клетках Huh7, как и в HepG2, были зафиксированы следующие изменения кариотипа: потеря участка q12 на хромосоме 1, потеря участка q11.2 на хромосоме 16, повышенная копийность хромосомы 20, а также потеря коротких плеч у SAT-хромосом – 13, 14, 15, 21 и 22. Кроме того, в Huh7 наблюдалась неполная трисомия хромосомы 6, но, в отличие от HepG2, здесь увеличена копийность по всему длинному плечу этой хромосомы. Для хромосомы 2 увеличение копийности наблюдалось в серединной части, в то время как в HepG2 высокая копийность присутствовала вдоль всей хромосомы.

Согласно результатам анализа транскриптомного профилирования печеночных клеточных линий, среднее количество экспрессируемых БКГ для клеток Huh7 составило 14 879, что соответствует 38 662 транскриптам, а для Hep3B – 13 940, что соответствует 35 864 транскриптам (TPM > 0). Как и в случае HepG2, между образцами одной культуры Huh7 и Hep3B наблюдается высокая корреляция – 0,87 и 0,91 соответственно.

Так, средний коэффициент корреляции Спирмена между образцами печеночных клеточных линий (HepG2 и Huh7, HepG2 и Hep3B, а также Huh7 и Hep3B) по данным экспрессии генов составил 0,86. По данным экспрессии транскриптов показатели ожидаемо были ниже – 0,71 для HepG2 и Huh7, а также 0,72 для HepG2 и Hep3B, в случае клеток Huh7 и Hep3B коэффициент корреляции Спирмена был 0,72.

Для оценки природы схожести печеночных клеточных линий был проведен PCA анализ по данным транскриптомного профилирования с включением в качестве контроля клеток иного генеза – MCF-7 и MDA-MB-231 (рак молочной железы), HeLa (рак шейки

матки), HCT116 (рак толстой кишки), LNCaP (рак простаты), HEK293T (почки), HaCaT (эпидермис), а также лейкемические линии K562 и THP-1. По данным экспрессии генов (рисунок 8А) и транскриптов (рисунок 8Б) наблюдается схожесть печеночных линий, но с большей вариабельностью, чем образцов иной природы, сгруппировавшихся в одну область.

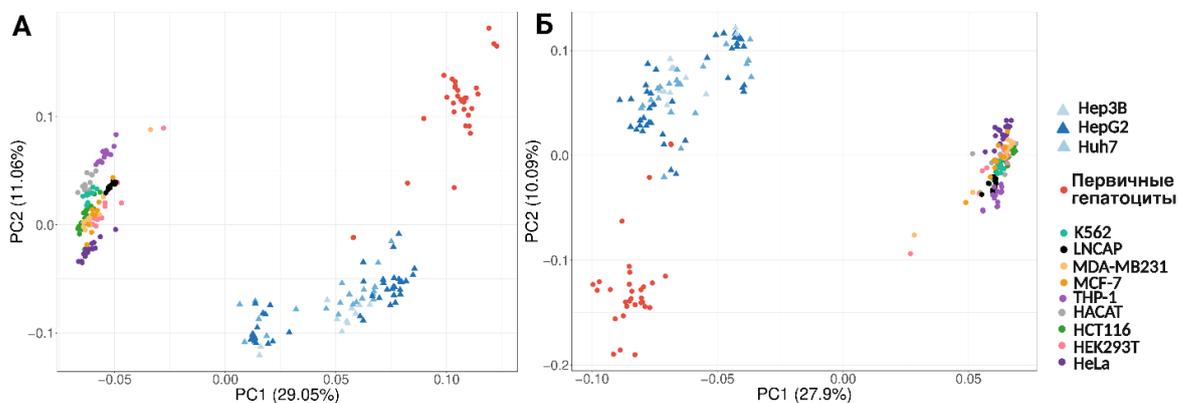


Рисунок 8. PCA для 13 клеточных линий. А) на уровне экспрессии генов; Б) на уровне экспрессии транскриптов.

В результате анализа данных протеомного профилирования клеточной линии Huh7 (два набора данных) детектировано 5 235 белков и 195 сплайс-вариантов, при этом коэффициент корреляции Спирмена между полученными данными составил 0,42. Для клеточной линии Hep3B (также два набора данных) было детектировано 5 626 белков и 224 сплайс-вариантов, а коэффициент корреляции Спирмена составил 0,28.

В результате анализа протеомного профилирования 20 наборов данных печеночных клеточных линий HepG2, Huh7 и Hep3B был установлен более высокий уровень схожести между клеточными линиями Huh7 и Hep3B, в отличие от HepG2 по отношению к образцам одного типа ($\rho = 0,55$), HepG2 и Hep3B (0,50), HepG2 и Huh7 (0,41). Во всех трех клеточных культурах было обнаружено 4 485 белков, которые участвуют в процессах регуляции связывания РНК, синтеза белков, клеточной адгезии, а также энергетического и сигнального обмена в клетке.

3.5. Различия между печеночными клеточными линиями и первичными гепатоцитами

Важной характеристикой молекулярного профиля опухолевых клеточных культур является их отличие от максимально нативных клеток-предшественников. Первичные гепатоциты выделены непосредственно из ткани печени и считаются золотым стандартом для изучения функции и заболеваний печени.

Ввиду ограниченности доступных для ПГ данных, построить полный молекулярный профиль не удалось, тем не менее имеется достаточно количество

опубликованных результатов транскриптомного профилирования для сопоставления профилей ПГ с опухолевыми клеточными линиями – HepG2, Huh7, Hep3B. Несмотря на высокую корреляцию между клеточными линиями и ПГ ($\rho = 0,84$, а транскриптов – 0,82), при анализе методом главных компонент (PCA) ПГ группируются в отдельный кластер (рисунок 8).

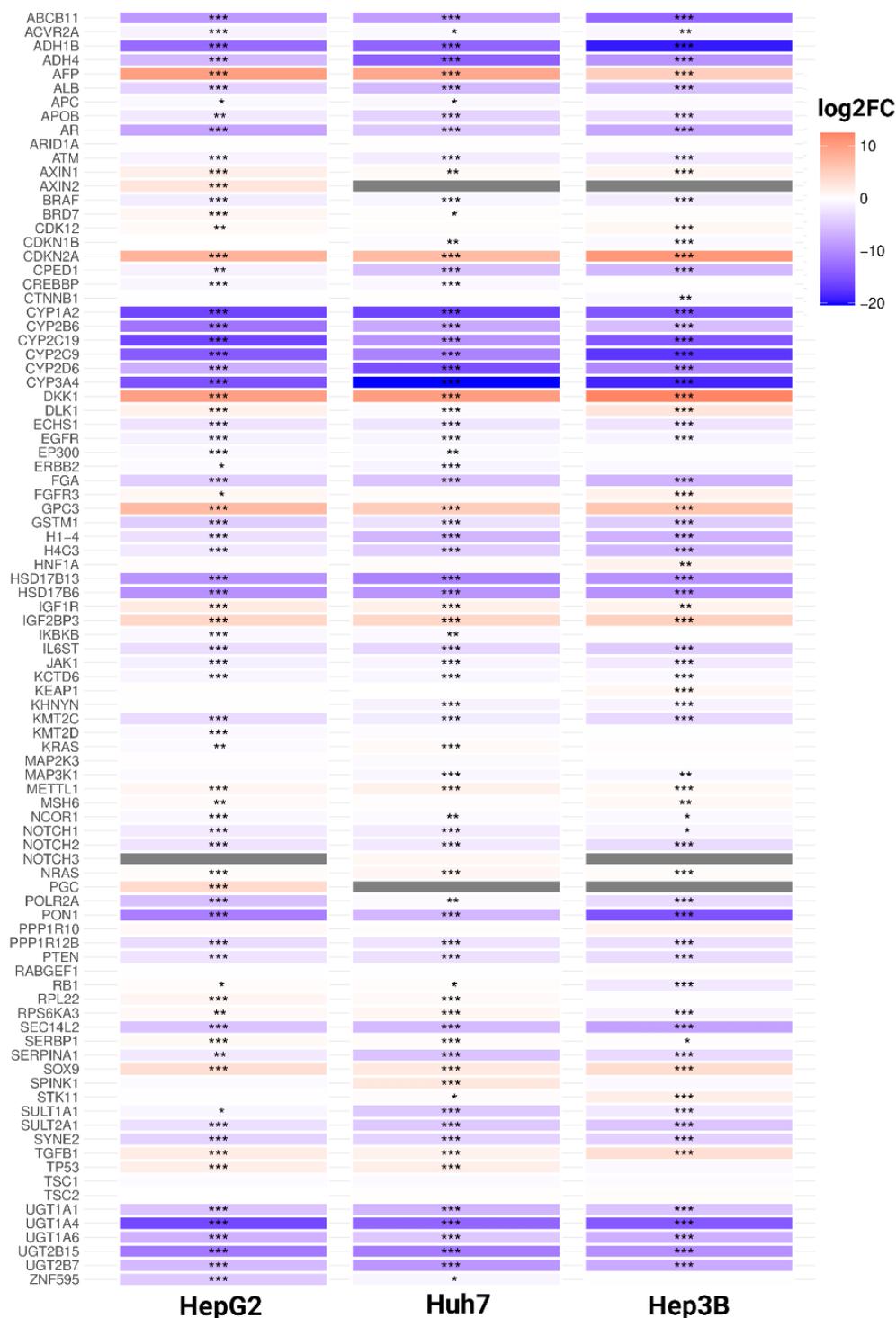


Рисунок 9. Изменение 91 из 102 потенциальных генов-маркеров рака печени в HepG2, Huh7 и Hep3B относительно ПГ. FDR меньше 0,05 считались статистически значимыми. Звездочки обозначают уровень значимости: *, ** и *** означают FDR < 0,05, FDR < 0,01 и FDR < 0,001 соответственно. Серый цвет означает отсутствие изменений.

Для всех трех рассматриваемых опухолевых клеточных культур показаны схожие изменения в дифференциальной экспрессии генов по отношению к ПГ. Гены, участвующие в метаболизме лекарств и токсинов, такие как гены из семейства цитохрома P450 и UGT, значимо понижались в клетках HepG2, Huh7 и Hep3B по сравнению с ПГ. Аналогично гены семейства гидроксистероиддегидрогеназ, такие как HSD17B13 и HSD17B6, которые играют ключевую роль в метаболизме стероидов, андрогенов и эстрогенов, также демонстрировали пониженную экспрессию. Значимо сниженная экспрессия также выявлена и в случае генов, кодирующих ферменты метаболизма алкоголя – ADH4 и ADH1B.

Для ряда генов, которые ассоциированы с развитием и прогрессированием рака печени – AFP, DKK1, IGF2BP3, CDKN2A и SOX9, – наоборот, наблюдается повышенная экспрессия в печеночных клеточных линиях. Ген DLK1 значимо повышается исключительно в клетках линии HepG2 и Hep3B, а ген TP53 – только в клетках Huh7 и HepG2. В то же время, экспрессия генов SPINK1 значимо повышалась только в линии Huh7, главный фактор, связанный с фиброзом печени, который возникает при хронических заболеваниях, таких как гепатит или алкогольная болезнь печени.

По результатам анализа построены профили изменения экспрессии 91 из 102 потенциальных генов-маркеров рака печени (Arzumanyan V.A. et al., 2021), по отношению к ПГ на основании результатов дифференциального анализа экспрессии генов (рисунок 9). Полученные сведения открывают новые возможности для теоретических исследований механизмов генной регуляции, а также для практического применения при разработке лекарственных средств и изучении токсинов, что, в свою очередь, может способствовать созданию более эффективных методов терапии и лечению рака печени.

4. ЗАКЛЮЧЕНИЕ

В результате детального анализа доступных данных генетического и молекулярного профилирования клеток НерG2 была построена подробная карта данной клеточной культуры. На рисунке 10 приведена усредненная по двум годам профилирования карта для образца, используемого в ИБМХ, который является 5-7 пассажем сертифицированных клеток, полученных от компании Sigma-Aldrich. Построение таких карт позволяет визуализировать и сопоставлять особенности реализации генетической информации как в целом, так и предметно для каждого участка генома.

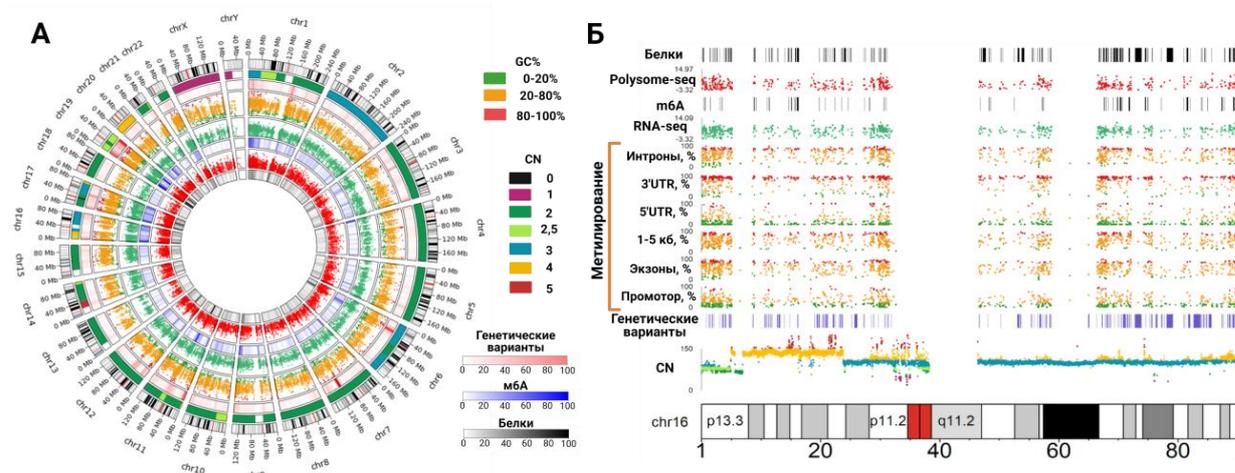


Рисунок 10. А) Молекулярная карта клеточной линии НерG2. От внешнего круга к внутреннему: координаты хромосом, CN, генетические варианты (нсОНП и инделы), метилирование участков, экспрессия генов, сайты м6А, экспрессия транслятов и встречаемость белков. Б) Молекулярная карта хромосомы 16.

В качестве примера на рисунке 10Б приведена молекулярная карта хромосомы 16, которая находится в трисомии (CN=3) с частичной тетрасомией (CN=4) в середине короткого плеча (p12.2-p13.2), а также с миниделцией начала длинного плеча (q11.2). В кодирующих регионах обнаружено 559 несинонимичных однонуклеотидных замен для 257 генов. Наибольшее количество генов (459 генов) было обнаружено с высокометилированными сайтами в области 3'UTR, в то время как наименьшее число генов (33 гена) – с высокометилированными сайтами в промоторных регионах. Высокое метилирование в области 3'UTR повышает экспрессию генов в регионе p13.3.

Анализ, проведенный в рамках данного исследования, углубил понимание молекулярных особенностей каждой хромосомы клеточной линии НерG2, включая генетические изменения, уровни экспрессии и эпигенетические модификации.

5. ВЫВОДЫ

1. Генотип клеточной линии HepG2 содержит устойчивые хромосомные aberrации: транслокацию между 1p и 21p, трисомию 2-й хромосомы, частичную трисомию 6 (p11.1-p13.3, q21-q24.3), 17 (после региона q12) хромосом, частичную тетрасомию хромосомы 16 (p12.2-p13.2), делецию части короткого плеча 9-й хромосомы (q12), тетрасомию 20-й хромосомы, потерю части короткого (p11.32) и длинного (q11.23qter) плеч Y-хромосомы и короткого плеча всех SAT-хромосом.
2. Клеточная линия HepG2 имеет высокий уровень гомогенности между образцами: коэффициент корреляции Спирмена (ρ) на уровне экспрессии генов составляет 0,88, на уровне экспрессии транскриптов – 0,73, по данным полуколичественной оценки содержания белка NSAF – 0,51. Более 80% генома имеют стабильную копийность.
3. Наблюдается эпигенетическая компенсация для генов с нестабильной копийностью. При повышенной копийности генов вдвое увеличивается доля средне и высокоэкспрессирующих генов (TPM > 20), которая практически неизменна при любом варианте увеличенной ploидности.
4. Модификации м6А в генах, связанных с фолдингом белков, влияют на снижение уровня их трансляции.
5. Печеночные клеточные линии HepG2, Huh7 и Hep3B демонстрируют общие устойчивые изменения генотипа, такие как трисомия 2-й и 6-й хромосомы, потеря р-плеча у всех SAT-хромосом и повышенная копийность 20-й хромосомы. На транскриптомном уровне между этими клеточными линиями не наблюдается существенных различий по уровню экспрессии генов (средний коэффициент корреляции $\rho = 0,86$).

6. СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых научных журналах:

1. **Arzumanian V.A.**, Kiseleva O.I., Poverennaya E.V. The Curious Case of the HepG2 Cell Line: 40 Years of Expertise // International Journal of Molecular Sciences. – 2021. – Т. 22 – № 23. – С. 13135.
2. Pyatnitskiy M.A., **Arzumanian V.A.**, Radko S.P., Ptitsyn K.G., Vakhrushev I.V., Poverennaya E.V., Ponomarenko E.A. Oxford Nanopore MinION Direct RNA-Seq for Systems Biology // Biology (Basel). – 2021. – Т. 10. – № 11. – С. 1131.
3. **Arzumanian V.A.**, Dolgalev G.V., Kurbatov I.Y., Kiseleva O.I., Poverennaya E.V. Epitranscriptome: Review of Top 25 Most-Studied RNA Modifications // International Journal of Molecular Sciences. – 2022. – Т. 23 – № 22. – С. 13851.
4. **Arzumanian V.**, Pyatnitskiy M., Poverennaya E. Comparative Transcriptomic Analysis of Three Common Liver Cell Lines // International Journal of Molecular Sciences. – 2023. – Т. 24. – № 10. – С. 8791.
5. **Arzumanian V.A.**, Kurbatov I.Y., Ptitsyn K.G., Khmeleva S.A., Kurbatov L.K., Radko S.P., Poverennaya E.V. Identifying N6-Methyladenosine Sites in HepG2 Cell Lines Using Oxford Nanopore Technology // International Journal of Molecular Sciences. – 2023. – Т. 24. – № 22. – С. 16477.
6. Ponomarenko E.A., Krasnov G.S., Kiseleva O.I., Kryukova P.A., **Arzumanian V.A.**, Dolgalev G.V., Ilgisonis E.V., Lisitsa A.V., Poverennaya E.V. Workability of mRNA Sequencing for Predicting Protein Abundance // Genes (Basel). – 2023. – Т. 14. – № 11. – С. 2065.
7. Poverennaya E.V., Pyatnitskiy M.A., Dolgalev G.V., **Arzumanian V.A.**, Kiseleva O.I., Kurbatov I.Y., Kurbatov L.K., Vakhrushev I.V., Romashin D.D., Kim Y.S., Ponomarenko E.A. Exploiting Multi-Omics Profiling and Systems Biology to Investigate Functions of TOMM34 // Biology (Basel). – 2023. – Т. 12. – № 2. – С. 198.
8. Поверенная Е.В., Киселева О.И., **Арзуманян В.А.**, Пятницкий М.А., Вахрушев И.В., Пономаренко Е.А. Несинонимичные однонуклеотидные замены и инделы: вклад в молекулярный постгеномный портрет клеточной линии HepG2 // Успехи современной биологии. – 2023. – Т. 143 – № 2. – С. 107-113.

Работы, опубликованные в сборниках материалов научных конференций:

9. **Арзуманян В.А.**, Поверенная Е.В. Оценка гетерогенности транскриптомов опухолевой клеточной линии HepG2 // XXVII Симпозиум «Биоинформатика и компьютерное конструирование лекарств». Сборник научных трудов XXVII симпозиума. Москва, Россия, 2021. С. 52.

10. **Арзуманян В.А.**, Пятницкий М.А., Поверенная Е.В. Оценка гетерогенности опухолевой клеточной линии HepG2 на основе транскриптомных данных // III Объединенный научный форум физиологов, биохимиков и молекулярных биологов. Материалы: VII Съезд Биохимиков России. X Российский Симпозиум «Белки и Пептиды». VII Съезд Физиологов СНГ. Том 2. Москва, Россия, 2021. С. 169.
11. **Arzumanian V.A.**, Poverennaya E.V. Biological or technical nature of transcriptomes heterogeneity // XXVIII Symposium of Bioinformatics and Computer-Aided Drug Discovery. Proceedings book. Moscow, Russia, 2022. P. 63.
12. **Арзуманян В.А.**, Поверенная Е.В. Оценка гетерогенности транскриптомов печеночных клеточных линий // Высокопроизводительное секвенирование в геномике (HSG-2022). Сборник тезисов III Всероссийской конференции. Новосибирск, Россия, 2022. С. 90.
13. **Arzumanian V.A.**, Poverennaya E.V. Identification of m6A modifications in the HepG2 cell line using direct RNA-seq Oxford Nanopore technologies // XXIX International Symposium "Bioinformatics and Computer-Aided Drug Discovery". Proceedings book. Moscow, Russia, 2023. С. 90.
14. **Арзуманян В.А.**, Пятницкий М.А., Поверенная Е.В. Сравнительный транскриптомный анализ трех популярных клеточных линий печени // 11-ая Московская конференция по вычислительной молекулярной биологии МССМВ'23. Москва, Россия, 2023. (<https://mccmb.belozersky.msu.ru/2023/thesis/pages/authors.utf8.html>).
15. **Arzumanian V.A.**, Radko S.P., Ptitsyn K.G., Poverennaya E.V. Analysis of m6A Modifications in HepG2 Cells // Scientific Program with Abstracts of The HUPO 2023 Congress. Busan, South Korea, 2023. С. 96. (<https://2023.hupo.org/scientific-program-program-at-a-glance/>).

7. ФИНАНСИРОВАНИЕ

Программа фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021 - 2030 годы) (№122030100168-2). Соглашение о предоставлении из федерального бюджета гранта в форме субсидии от 12 июля 2024 г. № 075-15-2024-643.

8. БЛАГОДАРНОСТИ

Автор выражает благодарность научному руководителю за поддержку и наставления на всех этапах подготовки диссертационной работы. Особая признательность коллегам за помощь в получении данных, их интерпретации и обсуждении полученных результатов.