

Всероссийская конференция «Биомедицинская химия: наука и практика»

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

МОЛЕКУЛЯРНЫЙ ПРОФИЛЬ ОПУХОЛЕВОЙ КЛЕТОЧНОЙ ЛИНИИ HepG2

*В.А. Арзуманян**, *М.А. Пятницкий*, *И.В. Вахрушев*, *К.Г. Птицин*,
С.П. Радько, *В.Г. Згода*, *О.И. Киселева*, *Е.В. Поверенная*

Научно-исследовательский институт биомедицинской химии им. В.Н. Ореховича,
119121, Москва, ул. Погодинская, 10; *e-mail: arzumanian.victoria@gmail.com

Клеточные линии широко используются в научных исследованиях благодаря своей доступности, стабильности и функциональной схожести с оригинальными клетками. Линия HepG2, будучи четвёртой по популярности клеточной культурой, часто применяется в токсикологических и метаболических исследованиях благодаря частичному сохранению свойств гепатоцитов. В нашей работе впервые построен молекулярный портрет клеточной культуры HepG2. Для построения портрета мы использовали ранее полученные нами данные для одного образца, включающие результаты полногеномного (WGS), метиломного (WGBS), транскриптомного (RNA-seq), трансляционного (Polysome-seq) и протеомного (LC-MS/MS) профилирования. Для оценки гетерогенности клеточной линии HepG2 анализировали полногеномные и транскриптомные данные, опубликованные в базе данных NCBI SRA, а также результаты протеомных исследований, доступных в ресурсе PRIDE. Наше исследование показало, что клеточная линия HepG2 демонстрирует в целом высокую степень стабильности на геномном и транскриптомном уровнях, однако образцы из Китая требуют более пристального внимания при переносе результатов транскриптомных и протеомных экспериментов. Генотип HepG2 характеризуется устойчивыми хромосомными перестройками, такими как транслокация между короткими плечами хромосом 1p и 21p, тетрасомия хромосомы 20, потеря короткого плеча у всех SAT-хромосом и длинного плеча Y хромосомы. При отсутствии на геномном уровне 1216 белок-кодирующих генов, на транскриптомном уровне экспрессируется 12602 гена, из которых только 10461 детектируются на уровне транслятов, а на протеомном уровне идентифицировано лишь 1027 генов, что связано с ограничением чувствительности масс-спектрометрических методов. В результате проведённого анализа омикс данных мы представили детальный молекулярный портрет клеточной культуры HepG2, иллюстрирующий мультиомный профиль каждого гена.

Ключевые слова: HepG2; клеточная линия; омикс данные; регуляция генов; молекулярный портрет

DOI: 10.18097/BMCRM00239

ВВЕДЕНИЕ

За последние десятилетия молекулярные биологи значительно расширили знания об “омах”: геноме, метиломе, транскриптоме, протеоме и метаболоме для различных организмов и клеточных линий. Однако, несмотря на накопленные данные, взаимосвязи между уровнями реализации генетической информации остаются фрагментарными [1].

Пробелы в понимании взаимосвязей ограничивают возможности интерпретации наблюдаемых эффектов и могут приводить к некорректной постановке экспериментов, что усложняет поиск и разработку новых биомаркеров и методов лечения заболеваний. Клеточные линии часто используют как унифицированный и доступный объект исследований, поскольку они обеспечивают воспроизводимый молекулярный профиль и уменьшение количества повторностей в биомедицинских исследованиях [2, 3].

Убеждённости в стабильном молекулярном портрете клеточных линий до 15–20 пассажа (в зависимости от типа клеток) ослабило исследование одной из популярных клеточных линий HeLa [4].

Группа под руководством Liu проиллюстрировала наличие трёх подтипов клеток с различными геномными и транскриптомными профилями. Это подчёркивает важность изучения гетерогенности клеточных линий на разных молекулярных уровнях [4].

Особое внимание заслуживает опухолевая линия печени HepG2, частично сохранившая свойства гепатоцитов. Данную культуру клеток часто используют в качестве модели для токсикологических и фармакологических исследований, а также для изучения рака. На 2024 год в базе данных PubMed по запросу “HepG2” находится 40059 статей, что делает её четвертой по популярности клеточной линией после HeLa (рак шейки матки), MCF-7 (рак молочной железы) и A549 (рак лёгкого).

Мы собрали молекулярный портрет опухолевой клеточной линии HepG2 на основе молекулярно-генетических данных на геномном, метиломном, транскриптомном, эпигенетическом и протеомном уровнях, что в итоге позволит более корректно использовать данную линию в экспериментах и минимизировать ошибки при интерпретации результатов.



МАТЕРИАЛЫ И МЕТОДЫ

Данные

Для анализа клеточной линии HepG2 были использованы как собственные, так и общедоступные данные. Собственные данные включали результаты цитогенетического анализа, а также полногеномного (WGS), метиломного (WGBS), транскриптомного (RNA-seq), трансляционного (Polysome-seq) и протеомного (LC-MS/MS) профилирований, ранее полученные для одного образца [5].

Необработанные собственные данные WGS, WGBS, RNA-seq и Polysome-seq были загружены в NCBI SRA: PRJNA765908, PRJNA956723 и PRJNA972889, а протеомные данные доступны в Mendeley Data [6].

Для геномного и транскриптомного секвенирования использовали данные из NCBI SRA, полученные с помощью технологии Illumina. Критерии для поиска полногеномных и эпигеномных данных: >100 млн прочтений, длина чтений 100 пар нуклеотидов (п.н.), метод WGS/WGBS, для транскриптомных данных: >20 млн прочтений, длина чтений >60 п.н. Суммарно найдено восемь образцов полногеномного секвенирования HepG2, доступные эпигеномные данные не соответствовали критериям. В соответствии с критериями найдено 40 транскриптомных образцов HepG2. В PRIDE найдено 11 протеомных данных HepG2, согласно следующим критериям: данные должны быть представлены “сырыми” файлами и получены из необработанных клеток.

Дополнительно был проведён поиск статей с опубликованными данными цитогенетического анализа для клеточной линии HepG2. В результате было найдено шесть результатов кариотипирования.

Цитогенетический анализ

Цитогенетическое исследование проводили с помощью G-дифференциального окрашивания 20 метафазных пластинок для определения хромосомного состава клеток.

Анализ данных секвенирования нового поколения

Для анализа собственных и опубликованных данных использовали одинаковые протоколы. Данные WGS, WGBS и RNA-seq (Illumina) были представлены в виде необработанных файлов fastq. Первый этап анализа включал оценку качества с использованием программы FastQ. Вторым этапом чтения низкого качества и адаптеры удаляли с помощью программы TrimGalore.

Данные транскриптомного и трансляционного профилирований (Oxford Nanopore) были представлены в формате fast5. Первый этап обработки включал распознавание азотистых оснований с помощью программы Guppy, результатом которого были файлы fastq. Затем файлы выравнивались с помощью алгоритма Minimap2.

Полногеномное секвенирование. Файлы fastq после фильтрации выравнивали на референсный геном человека (Ensembl GRCh38) с использованием

алгоритма Bowtie2. Полученные bam-файлы использовали для оценки копийности хромосом с помощью программы mosdepth. Для детекции несинонимичных однонуклеотидных замен (SNP) и инделов использовали рекомендации GATK.

Эпигеном. Идентификацию CpG участков выполняли с помощью алгоритма Bismark с выравниванием на геном человека (Ensembl GRCh38).

Транскриптомное профилирование. Анализ экспрессии генов и изоформ, полученных с помощью Illumina и ONT, был выполнен программой Salmon. Экспрессию транскриптов рассчитывали в TPM (транскрипты на миллион), а экспрессию генов — суммированием TPM соответствующих транскриптов.

Эпитранскриптом. Анализ данных прямого секвенирования РНК (direct RNA-seq) клеточной линии HepG2 для поиска сайтов модификации N6-метиладенозина (m6A) проводили с использованием программы Nanopolish и алгоритма m6Anet. Участки с коэффициентом модификации менее 0.9 исключали из анализа.

Трансляционное профилирование. Данные трансляционного профилирования в двух повторах были получены с помощью Oxford Nanopore и анализировались по тому же протоколу, что и транскриптомные данные прямого секвенирования РНК.

Протеомное профилирование. Протеомные данные для клеточной линии HepG2 были проанализированы по единому протоколу с учётом рекомендации протеомного сообщества [7]. Данные преобразовывали в формат mgf с помощью программы MSConvert и анализировали в программе SearchGUI с использованием алгоритма X!Tandem. Для полуквантитативного анализа использовался индекс нормированных спектральных коэффициентов обилия (NSAF). Референсная база данных включала канонические и образующиеся в результате альтернативного сплайсинга последовательности белков для человека из SwissProt. Опубликованные протеомные данные также анализировали в SearchGUI с уникальными параметрами для каждого образца.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ кариотипа

Клеточная линия HepG2 характеризуется гипердиплоидным кариотипом, содержащим в среднем от 50 до 60 хромосом [8–11]. Анализ собственного и опубликованных кариотипирований клеток HepG2 выявил устойчивые хромосомные аномалии, включая транслокацию между короткими плечами хромосом 1p и 21p, тетрасомию хромосомы 20 и трисомию хромосомы 6 с потерей её короткого плеча. Также наблюдались и отличия для хромосом 16 и 17 между собственными и опубликованными данными [10].

Мы провели анализ глубины покрытия и GC состава для всех 10 образцов полногеномного секвенирования клеточной линии HepG2. На основе этих данных была получена анеуплоидная карта

с высоким разрешением, выявившая значительные изменения числа копий в хромосомных областях клеточной линии HepG2. В 70% образцов обнаружено пять кластеров копийности. В среднем, геном клеточной линии HepG2 имеет 65% базового числа копий (CN=2).

Результаты кариотипирования клеток HepG2 подтверждены анализом покрытия хромосом на геномном уровне. Сравнение с опубликованными данными показало стабильные хромосомные аномалии: транслокация между короткими плечами хромосом 1p и 21p, трисомия хромосомы 2, частичная трисомия хромосомы 6 (q1–q21), делеция короткого плеча хромосомы 9, тетрасомия хромосомы 20, потеря длинного плеча Y хромосомы и делеция коротких плеч у хромосом, имеющих спутники (SAT-хромосомы) — 13, 14, 15, 21 и 22.

Таким образом, кариотип клеточной линии HepG2 сохраняет стабильные ключевые изменения, подтверждённые как собственными исследованиями, так и опубликованными данными, несмотря на различия межлабораторной воспроизводимости.

Анализ генетических вариантов

Сопоставление собственных и опубликованных данных на уровне несинонимичных однонуклеотидных замен показало, что только 7% вариантов встречались во всех образцах, 5,3% — в половине образцов, и 53% — только в одном или двух образцах. Наибольшее количество мутаций обнаружено в генах семейства MUC (муцины) и HLA (антиген человеческого лейкоцита). Муцины участвуют в формировании защитных слизистых барьеров, и их мутации могут способствовать прогрессированию опухоли и метастазированию. Гены HLA играют ключевую роль в иммунном ответе, и их мутации связаны с уклонением опухоли от иммунного надзора [12]. Такие значительные геномные aberrации подчеркивают важность генетического мониторинга для обеспечения надёжности и воспроизводимости результатов исследований.

В результате анализ полногеномных данных также было выявлено отсутствие 1216 белок-кодирующих генов (БГК).

Анализ метиломных данных

Анализ метиломных данных для клеточной линии HepG2 показал, что 44% генома подверглось метилированию, что соответствует общему уровню метилирования в клетках человека [13]. Было определено 23637476 сайтов метилирования в 17369 БГК.

Анализ транскриптомных данных

В результате анализа транскриптомных данных клеточной линии HepG2, полученных с помощью технологии Illumina, в среднем была обнаружена экспрессия 12602 генов и 32583 транскриптов при TPM >0.1. Коэффициент корреляции Спирмена между образцами составил 0.82 на уровне транскриптов и 0.93 на уровне генов.

Для оценки гетерогенности клеток HepG2 были сопоставлены опубликованные данные транскриптомного профилирования 42 образцов. Между образцами выявлена сильная положительная корреляция: по данным экспрессии генов коэффициент корреляции Спирмена составил 0.88, а по данным экспрессии транскриптов — 0.77.

Примечательно, что, несмотря на высокие значения корреляции, образцы разделились на два кластера. В первом кластере преобладали данные из китайских институтов, тогда как во втором — из институтов США, Японии, Южной Кореи, Турции, Колумбии, Италии и России. Анализ дифференциальной экспрессии выявил 2824 гена с различной экспрессией между этими двумя кластерами. Между кластерами была отмечена дисрегуляция генов, участвующих в метаболизме лекарств и детоксикации. Также были выявлены различия в путях “Окисление с помощью цитохрома P450” и “Метаболизм эстрогенов”. Снижение экспрессии этих генов указывает на потенциальное ухудшение способности клеток к метаболизму и детоксикации лекарств, что может влиять на точность и значимость анализов метаболизма лекарств, проводимых с использованием клеток HepG2 [14].

Анализ эпитранскриптомных данных

В клеточной линии HepG2 в пяти технических повторах суммарно обнаружено 3968 возможных сайтов m6A модификаций в 1396 генах. В результате сравнения была показана низкая воспроизводимость: только 6.3% (250 из 3968) возможных сайтов m6A были обнаружены во всех повторностях. Однако все гены с выявленными сайтами m6A модификаций демонстрировали ассоциированность с процессами убиквитилирования, несмотря на низкую воспроизводимость [15]. Таким образом, при наличии множественных технических повторов возможно использовать совокупную информацию для повышения надёжности данных.

Анализ транслятомных данных

За последние десять лет значительно увеличилось понимание роли трансляции как ключевого регулятора экспрессии генов. Транслятом выступает промежуточным звеном между транскриптомом и протеомом. Анализ транслятомного профилирования собственного образца клеточной линии HepG2 показал экспрессию 10461 гена и 22513 транскриптов при TPM >0.1.

Анализ протеомных данных

По результатам анализа нашего образца в трёх повторах было выявлено 1027 белков (NSAF >0). В результате анализа протеомного профилирования опубликованных данных для 14 образцов клеточной линии HepG2 были получены низкие значения корреляции Спирмена между данными полуколичественной оценки NSAF — 0.46.

ЗАКЛЮЧЕНИЕ

В ходе исследования установлено, что клеточная линия HepG2 демонстрирует в целом высокую степень стабильности на геномном и транскриптом уровнях, однако образцы из Китая требуют более пристального внимания при переносе результатов транскриптомных и протеомных экспериментов. В частности для клеток HepG2, выявлены устойчивые хромосомные перестройки, такие как транслокация между короткими плечами хромосом 1p и 21p, тетрасомия хромосомы 20, потеря короткого плеча у всех SAT-хромосом и длинного плеча Y хромосомы. При отсутствии на геномном уровне 1216 БКГ, на транскриптомном уровне экспрессируется 12602 гена, из которых только 10461 детектируются на уровне транслятов, а на протеомном уровне идентифицировано лишь 1027 генов (рис. 1).

Впервые была построена подробная молекулярная карта клеточной линии HepG2 для каждой хромосомы (рис. 2), позволяющая оценить все молекулярные события на каждом участке генома. В качестве примера такой интеграции на рисунке 3 приведена хромосома 16 — благодаря полногеномному анализу уточнено, что не вся хромосома находится в трисомии, как следует из цитогенетического анализа. В середине короткого плеча (p13.2-12.2), находящегося в тетрасомном состоянии, уровень метилирования, экспрессия транскриптов и транслятов ниже, чем в целом для остальных участков хромосом, хотя количество сайтов m6A чуть выше. Данный регион содержит 122 БКГ, включая 10 из 24 генов, относящихся к семейству белков, взаимодействующих с ядерным поровым комплексом (Nuclear pore complex-interacting protein family member). Среди этих 122 генов 82 были детектированы на транскриптомном уровне, 78 на транслятомном и 6 на протеомном.

Таким образом, полученные результаты способствуют как более качественной интерпретации получаемых данных для клеточной линии HepG2, так и дальнейшей расшивке взаимосвязей регуляции клеточных процессов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Работа не содержит каких-либо исследований с участием людей или использованием животных в качестве объектов.

БЛАГОДАРНОСТИ

Эксперименты выполнены с использованием оборудования ЦКП “Протеом человека” Научно-исследовательского института биомедицинской химии им. В.Н. Ореховича. Кариотипирование выполнялось в ЦКП “Группа геномных технологий” в Институте биологии развития им. Н.К. Кольцова РАН.

ФИНАНСИРОВАНИЕ

Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021–2030 годы) (№ 122030100168-2).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

ЛИТЕРАТУРА

1. Yu, Y., Zhang, N., Mai, Y., Ren, L., Chen, Q., Cao, Z., Chen, Q., Liu, Y., Hou, W., Yang, J., Hong, H., Xu, J., Tong, W., Dong, L., Shi, L., Fang, X., Zheng, Y. (2023) Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method. *Genome Biol.*, **24**(1), 201. DOI: 10.1186/s13059-023-03047-z
2. Goodspeed, A., Heiser, L.M., Gray, J.W., Costello, J.C. (2016) Tumor-derived cell lines as molecular models of cancer pharmacogenomics. *Mol. Cancer Res.*, **14**(1), 3–13. DOI: 10.1158/1541-7786.MCR-15-0189
3. Gillet, J.-P., Varma, S., Gottesman, M.M. (2013) The clinical relevance of cancer cell lines. *Journal of the Natl. Cancer Inst.*, **105**(7), 452–458. DOI: 10.1093/jnci/djt007
4. Liu, Y., Mi, Y., Mueller, T., Kreibich, S., Williams, E.G., van Droegen, A., Borel, C., Frank, M., Germain, P.-L., Bludau, I., Mehnert, M., Seifert, M., Emmenlauer, M., Sorg, I., Bezrukov, F., Bena, F.S., Zhou, H., Dehio, C., Testa, G., Saez-Rodriguez, J., Antonarakis S.E., Hardt W.D., Aebersold R. (2019) Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat. Biotechnol.*, **37**(3), 314–322. DOI: 10.1038/s41587-019-0037-y
5. 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. (2023) Exploiting multi-omics profiling and systems biology to investigate functions of TOMM34. *Biology*, **12**(2), 198. DOI: 10.3390/biology12020198
6. Mendeley Data. Exploiting multi-omics profiling and systems biology to investigate functions of TOMM34. Retrieved December 6, 2022, from: <https://data.mendeley.com/datasets/yrm8ygnncn/1>
7. Deutsch, E.W., Lane, L., Overall, C.M., Bandeira, N., Baker, M.S., Pineau, C., Moritz, R.L., Corrales, F., Orchard, S., van Eyk, J.E., Paik, Y.-K., Weintraub, S.T., Vandenbrouck, Y., Omenn, G.S. (2019) Human proteome project mass spectrometry data interpretation guidelines 3.0. *J. Proteome Res.*, **18**(12), 4108–4116. DOI: 10.1021/acs.jproteome.9b00542
8. Zhou, B., Ho, S.S., Greer, S.U., Spies, N., Bell, J.M., Zhang, X., Zhu, X., Arthur, J.G., Byeon, S., Pattni, R., Saha, I., Huang, Y., Song, G., Perrin, D., Wong, W.H., Ji, H.P., Abyzov, A., Urban, A.E. (2019) Haplotype-resolved and integrated genome analysis of the cancer cell line HepG2. *Nucleic Acids Res.*, **47**(8), 3846–3861. DOI: 10.1093/nar/gkz169
9. Wong, N., Lai, P., Pang, E., Leung, T.W., Lau, J.W., Johnson, P.J. (2000) A comprehensive karyotypic study on human hepatocellular carcinoma by spectral karyotyping. *Hepatology (Baltimore, Md)*, **32**(5), 1060–1068. DOI: 10.1053/jhep.2000.19349



Рисунок 1. Детектируемый молекулярный профиль одного образца клеточной линии НерG2.

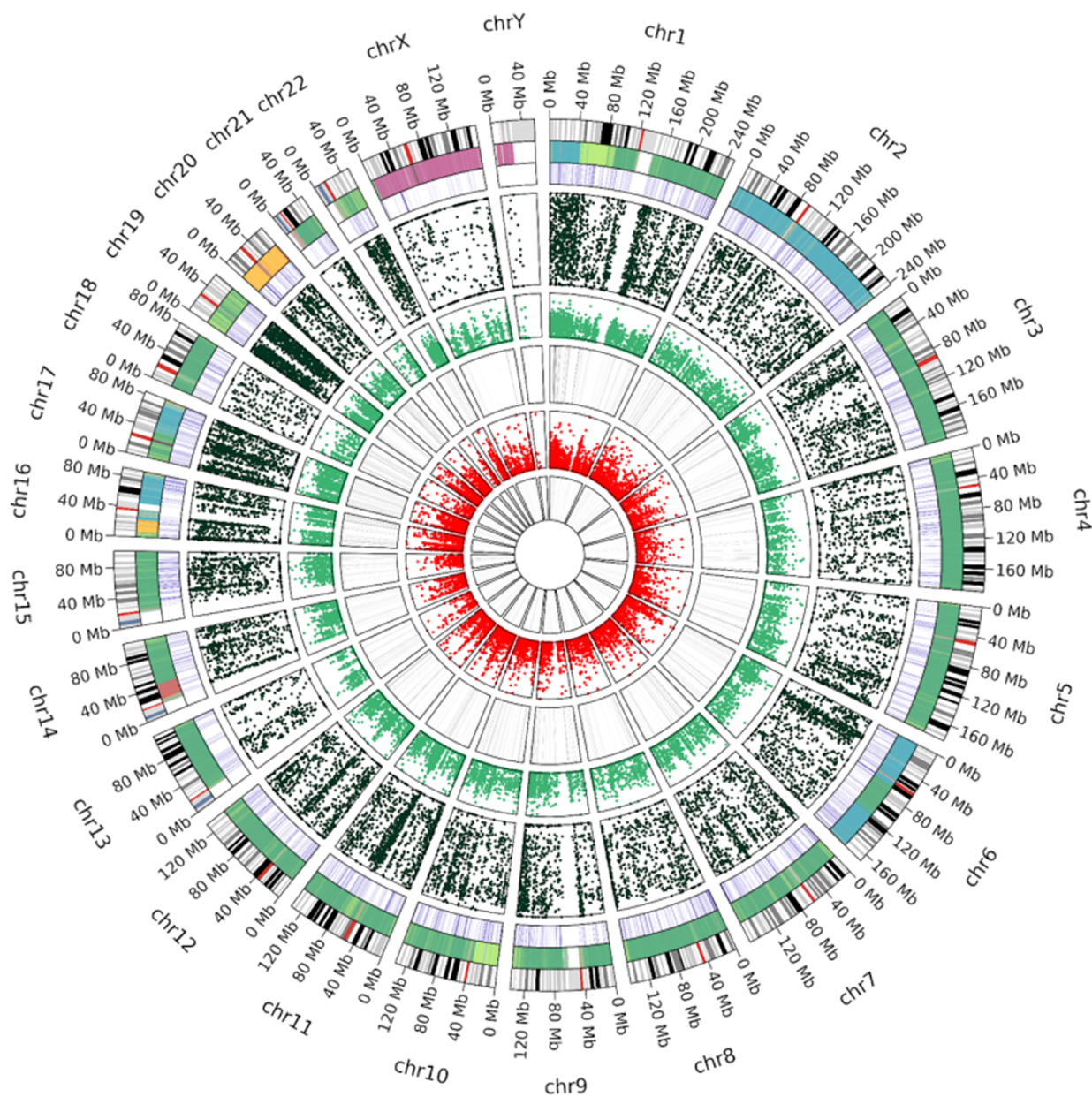


Рисунок 2. Молекулярная карта для клеточной линии НерG2. От внешнего круга: координаты хромосомы, копияность генов, генетические варианты, метилирование участков, экспрессия генов, m6A сайты, экспрессия транслятов и встречаемость белков.

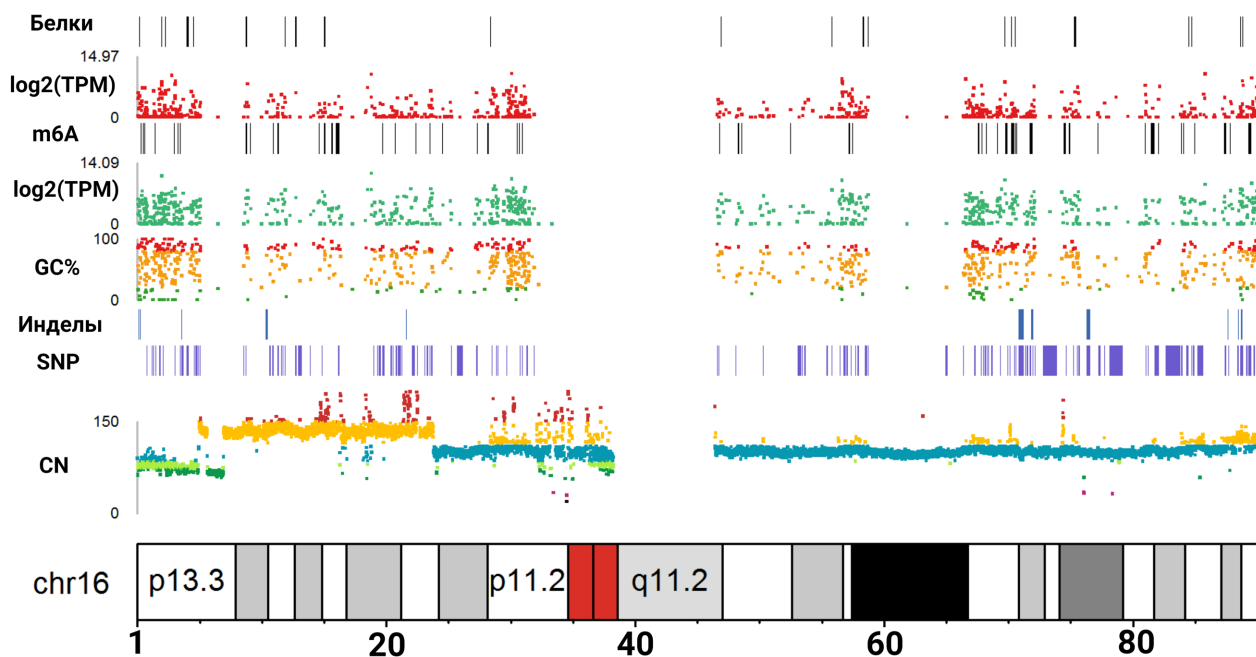


Рисунок 3. Молекулярная карта для хромосомы 16. Снизу вверх: координаты хромосомы, CN – копияность генов, SNP – несинонимичные однонуклеотидные замены, Инделлы – вставки и делеции, GC% – метилирование участков, log2(TPM) – экспрессия генов, m6A сайты, log2(TPM) – экспрессия транскриптов и встречаемость белков. Цвета на уровне CN: чёрный означает копияность 0, розовый – копияность 1, зелёный – копияность 2, салатовый – копияность 2,5, голубой – копияность 3, жёлтый – копияность 4 и красный – копияность 5. Цвета на уровне GC%: красный – гены с высоким уровнем метилирования, жёлтым – средним уровнем, и зелёным – низким уровнем.

10. Simon, D., Aden, D.P., Knowles, B.B. (1982) Chromosomes of human hepatoma cell lines. *Int. J. Cancer*, **30**(1), 27–33. DOI: 10.1002/ijc.2910300106

11. Chen, H.L., Chiu, T.S., Chen, P.J., Chen, D.S. (1993) Cytogenetic studies on human liver cancer cell lines. *Cancer Genet. Cytogen.*, **65**(2), 161–166. DOI: 10.1016/0165-4608(93)90227-d

12. Guo, J.-J., Ye, Y.-Q., Liu, Y.-D., Wu, W.-F., Mei, Q.-Q., Zhang, X.-Y., Lao, J., Wang, B., Wang, J.-Y. (2022) Interaction between human leukocyte antigen (HLA-C) and killer cell Ig-like receptors (KIR2DL) inhibits the cytotoxicity of natural killer cells in patients with hepatoblastoma. *Front. Med.*, **9**, 947729. DOI: 10.3389/fmed.2022.947729

13. Strichman-Almashanu, L.Z., Lee, R.S., Onyango, P.O., Perlman, E., Flam, F., Frieman, M.B., Feinberg, A.P. (2002) A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res.*, **12**(4), 543–554. DOI: 10.1101/gr.224102

14. Arzumanyan, V.A., Kiseleva, O.I., Poverennaya, E.V. (2021) The curious case of the HepG2 cell line: 40 years of expertise. *Int. J. Mol. Sci.*, **22**(23), 13135. DOI: 10.3390/ijms222313135

15. Arzumanyan, V.A., Kurbatov, I.Y., Ptitsyn, K.G., Khmeleva, S.A., Kurbatov, L.K., Radko, S.P., Poverennaya, E.V. (2023) Identifying N6-methyladenosine sites in HepG2 cell lines using oxford nanopore technology. *Int. J. Mol. Sci.*, **24**(22), 16477. DOI: 10.3390/ijms242216477

Поступила: 09. 07. 2024.

После доработки 25. 07. 2024.

Принята к публикации: 29. 07. 2024.

MOLECULAR PROFILE OF THE HepG2 TUMOR CELL LINE

*V.A. Arzumanian**, *M.A. Pyatnitsky*, *I.V. Vakhrushev*, *K.G. Ptitsyn*,
S.P. Radko, *V.G. Zgoda*, *O.I. Kiseleva*, *E.V. Poverennaya*

Institute of Biomedical Chemistry,
10 bld. 8, Pogodinskaya str., Moscow, 119121 Russia; *e-mail: arzumanian.victoria@gmail.com

Cell lines are widely used in scientific research due to their accessibility, stability, and functional similarity to the original cells. The HepG2 line, being the fourth most popular cell culture, is often used in toxicological and metabolic studies due to its partial retention of hepatocyte properties. In our study, the molecular portrait of the HepG2 cell culture was constructed for the first time. To build this portrait, we used previously obtained data for a single sample, including results of whole-genome sequencing (WGS), whole-genome bisulfite sequencing (WGBS), transcriptome (RNA-seq), translome (Polysome-seq), and proteome (LC-MS/MS) profiling. For the assessment of HepG2 cell line heterogeneity, we analyzed whole-genome and transcriptome data published in the NCBI SRA database, as well as proteome research results available in the PRIDE resource. Our study showed that the HepG2 cell line generally demonstrates a high degree of stability at the genomic and transcriptomic levels; however, samples from China require closer attention when transferring the results of transcriptomic and proteomic experiments. The HepG2 genotype is characterized by stable chromosomal rearrangements, such as translocation between the short arms of chromosomes 1p and 21p, tetrasomy of chromosome 20, loss of the short arm of all SAT chromosomes, and the long arm of the Y chromosome. Despite the absence of 1216 protein-coding genes at the genomic level, 12,602 genes are expressed at the transcriptomic level, of which only 10,461 are detected at the translome level, and only 1027 genes are identified at the proteome level, which is related to the limitations of mass spectrometry sensitivity. As a result of the omics data analysis, we presented a detailed molecular portrait of the HepG2 cell culture, illustrating the omics profile at various levels for each gene.

Key words: HepG2; cell line; omics data; gene regulation; molecular portrait

FUNDING

The work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030) (no. 122030100168-2).

Received: 09.07.2024; revised: 25.07.2024; accepted: 29.07.2024.