

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

©Коллектив авторов

ДЕТЕКЦИЯ НИЗКОКОПИЙНЫХ БЕЛКОВ В ПРОТЕОМНЫХ ИССЛЕДОВАНИЯХ: ПРОБЛЕМЫ И ПУТИ РЕШЕНИЯ

А.И. Арчаков, Н.Э. Вавилов, В.Г. Згода*

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича,
119121, Москва, ул. Погодинская, 10; *эл. почта: n.vavilov95@gmail.com

Одна из наиболее актуальных проблем современной протеомики — детекция низкокопийных белков в комплексных биологических образцах. Главная причина малоэффективной детекции низких концентраций белков заключается в недостаточной чувствительности масс-спектрометрического детектора и высоком динамическом диапазоне концентраций белков. В данной работе были исследованы возможности и ограничения метода таргетного масс-спектрометрического анализа на примере реконструированной системы стандартных белков UPS1 (Универсальный Протеомный Стандарт 1, Universal Proteomic Standard 1). Показано, что на чувствительность метода влияет концентрация целевых белков системы UPS1, а также высокий уровень биологического шума в виде белков цельного лизата *E. coli*. Ограничения метода удалось преодолеть с помощью концентрирования и предварительного фракционирования пептидов образца в хроматографической системе на обращённой фазе в щелочных условиях элюции. Протеомный анализ биологического образца — белков клеточной линии гепатоцеллюлярной карциномы человека HepG2, кодируемых генами 18 хромосомы человека, — показал повышение чувствительности метода по сравнению со стандартным таргетным масс-спектрометрическим анализом. Это позволило зарегистрировать 94 белка, кодируемых генами 18 хромосомы человека.

Ключевые слова: протеомика; масс-спектрометрия; HepG2; продукты генов 18 хромосомы человека

DOI: 10.18097/PBMC20247005342

ВВЕДЕНИЕ

Проект “Протеом человека” был инициирован в 2010 году на международном конгрессе в Сиднее [1]. Основной целью проекта стал поиск и описание белков и кодирующих их мРНК, которые содержатся в геноме человека. Также предложен хромосомоцентричный проект “Протеом человека” — C-HPP (Chromosome-centric Human Proteome Project), в котором странам-участникам предлагалось исследовать продукты генов конкретной хромосомы человека [2]. Основная цель проекта — детекция всех белков, кодируемых генами человека. Выполнение проекта позволит расширить знания о биологии человека на клеточном уровне, что в свою очередь послужит основой будущих прогностических, диагностических, терапевтических и превентивных медицинских приложений, основанных на данных о белках, полученных в ходе выполнения проекта. Группа исследователей из России под руководством А.И. Арчакова выбрала 18 хромосому человека для поиска и описания продуктов белок-кодирующих генов. В ходе проекта разработан масс-спектрометрический метод регистрации и измерения абсолютной концентрации белков, который обладает высокими показателями чувствительности и селективности. Разработанный метод был апробирован на различных типах биоматериала человека: клеточной линии гепатоцеллюлярной карциномы человека HepG2, клетках печени человека и сыворотке крови человека [3]. Зарегистрированные белки, кодируемые генами 18 хромосомы человека, покрывают

12 порядков концентраций от 10^{-6} М до 10^{-18} М. Детекции белков в ультранизкой концентрации удалось достичь методом необратимого связывания аналитов в растворе [4].

Однако в ходе работы выяснилось, что при исследовании биологических образцов человека современные протеомные методы регистрации белков не обладают достаточной чувствительностью [5]. Например, методы панорамной масс-спектрометрии при анализе комплексного образца клеточной линии гепатоцеллюлярной карциномы человека позволяют зарегистрировать около 20000 уникальных пептидов. Однако программное обеспечение для обработки протеомных данных также обнаруживает и считает количество не зарегистрированных прекурсорных ионов в исследуемом образце. Обычно количество незарегистрированных пептидов значительно выше и составляет около 120000 прекурсоров. Таким образом, панорамная методика обеспечивает высокое количество идентификаций, но минимальная концентрация белков обычно не превышает 10^{-9} М [6]. Также существуют методы направленной масс-спектрометрии, которые характеризуются повышенной чувствительностью. Однако для сохранения чувствительности на высоком уровне необходимо ограничивать количество регистрируемых белков до 200–300 на метод. Для получения глубокого протеома 264 белков, кодируемых генами 18 хромосомы человека, в клетках гепатоцеллюлярной карциномы человека HepG2 был использован метод таргетной масс-спектрометрии с использованием внутренних стандартов

в комбинации с предварительным фракционированием. Для определения чувствительности методики при различном соотношении уровня сигнал/шум была использована упрощенная белковая система UPS1.

МЕТОДИКА

Реактивы

В работе были использованы следующие реактивы: муравьиная кислота, ("Sigma-Aldrich", Швейцария), трифторуксусная кислота, 2-хлорацетамид, трис-2-карбоксиэтилфосфин, дезоксихолат натрия, ДМЕМ, Universal Proteomic Standard 1 (UPS1), фосфатный буфер ("Sigma-Aldrich", США), мочевины, хлорид натрия ("Acros organics", США), ацетонитрил ("Fisher Chemical", Китай), триэтиламмония бикарбонат (ТАБ) ("Fluca", Швейцария), трипсин свиной модифицированный ("Promega", США), фетальная бычья сыворотка (ФБС) ("Biowest", Франция), пенициллин/стрептомицин ("Gibco", США) трипсин/ЭДТА стерильный раствор ("Панэко", Россия).

Культивирование клеточной линии гепатоцеллюлярной карциномы человека HepG2

Культура клеток гепатоцеллюлярной карциномы HepG2 была получена из клеточного банка Института Биомедицинской химии им. В.Н. Ореховича (ИБМХ, Москва, Россия). Клеточная культура была выращена в среде с добавлением ДМЕМ, 10% ФБС, антибиотиков (пенициллин/стрептомицин 100 единиц/мл). Клетки культивировали в инкубаторе при температуре 37°C и в атмосфере 5% CO₂. При культивировании среду обновляли каждые 2–3 дня. Клетки были собраны при конfluence более 80% с помощью раствора 0,25% трипсин/ЭДТА для получения суспензии клеток.

Лизис и гидролиз образцов клеток линии HepG2

Клеточный осадок промывали от сыворотки в 100 мМ фосфатно-солевом буфере pH 7,4. После центрифугирования (10 мин, 500 g) отбирали супернатант и в клеточный осадок добавляли 10 мкл лизирующего буфера (4 М мочевины, 1% дезоксихолат натрия, 15% ацетонитрил, 5 мМ трискарбоксиэтилфосфин, 100 мМ фосфатный буфер pH 7,4, 100 мМ хлорид натрия). Раствор подвергали обработке ультразвуком для снижения вязкости (5 циклов работы прибора Q125 Sonicator ("Qsonica Sonicators", США): 10 с — озвучивание, 30 с — пауза с мощностью 30%). В дальнейшем гомогенат нагревали до 60°C в течение 30 мин. После охлаждения до комнатной температуры добавляли 2-хлорацетамид в конечной концентрации 50 мМ. Инкубацию проводили при комнатной температуре в темноте. После этого добавляли 90 мкл 100 мМ ТАБ и 1 мкг трипсина и проводили инкубацию при температуре 38°C; через 3 ч добавляли ещё одну аликвоту трипсина (1 мкг). Суммарное время гидролиза составило 18 ч. Реакцию останавливали добавлением муравьиной кислоты до конечной концентрации 2%, после этого проводили центрифугирование (10 мин, 10000 g) и отбор

супернатанта. Супернатант переносили в стеклянный сосуд ("Agilent", Германия), упаривали в вакуумном концентраторе (Concentrator plus, "Eppendorf", Германия) и перерастворяли в 0,1% растворе муравьиной кислоты.

Подготовку белков системы UPS1 к масс-спектрометрическому анализу проводили, как описано ранее [7].

Масс-спектрометрический анализ

Анализ образцов проводили на масс-спектрометре, объединённом с хроматографом в единую систему (6495 Triple Quad LC/MS, "Agilent", США). Настройки прибора и метод детекции применяли, как было описано ранее [8].

Визуализацию и обработку результатов таргетного анализа проводили в программном обеспечении Skyline (версия 23.1), как было описано ранее [8].

Транскриптомный анализ

Тотальная РНК была выделена из клеток гепатоцеллюлярной карциномы человека HepG2 с помощью коммерческого набора Extract RNA ("Евроген", Россия). Качество выделения контролировали спектрофотометрически с использованием Bioanalyzer 2100 System ("Agilent Technologies", США). Транскриптомный анализ выполнен на приборе Illumina HiSeq 2500 system ("Illumina", США) согласно протоколу производителя. Для каждого повтора было получено от 32 до 59 миллионов прочтений [9].

Культивирование бактериального штамма E. coli K-12

Культура была выращена по стандартному протоколу. Выделение и подготовку белка к масс-спектрометрическому анализу выполняли, как описано ранее [7].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Определение чувствительности и ограничений методики направленной масс-спектрометрии (SRM) при различной концентрации белков упрощенной белковой системы UPS1

Низкая глубина покрытия белков, кодируемых генами 18 хромосомы человека, объясняется недостаточной чувствительностью современных масс-спектрометрических методов. Причина низкой чувствительности может быть обусловлена одновременной элюцией и ионизацией огромного количества пептидных молекул, из которых лишь малая часть может быть зарегистрирована [10, 11]. Для определения рабочего диапазона метода и способов преодоления недостаточной чувствительности была сконструирована модельная система, которая имитирует биологический образец.

Упрощенная белковая система UPS1 представляет собой эквимольную смесь из 47 рекомбинантных, высокоочищенных белков человека, которую в том числе используют для разработки и валидации новых методик анализа белков [12, 13]. Данный объект

используют в качестве модельного для имитации целевого сигнала пептидов в биологическом образце. Для того чтобы исключить интерференцию сигнала между белками человека и UPS1 в качестве шума был использован другой биологический объект — *E. coli* (штамм K-12).

Таким образом, из этих компонентов было собрано 3 разных варианта модельной системы с различной концентрацией целевых белков UPS1: 1) раствор UPS1 без добавления *E. coli*; 2) раствор UPS1 с добавлением *E. coli*, в котором по мере снижения концентрации UPS1 пропорционально снижается концентрация *E. coli*; 3) раствор UPS1 с добавлением *E. coli* в одинаковой предельно допустимой высокой концентрации. Серия разведений с низким показателем значения сигнал/шум наиболее близко имитирует биологической эффект матрицы, который реализуется при анализе комплексного биологического образца [14].

Результаты направленного масс-спектрометрического анализа демонстрируют, что методика обладает 100% чувствительностью в области высоких концентраций 10^{-9} М, независимо от наличия биологического шума в виде *E. coli* в образце (рис. 1). Наиболее выраженное снижение количества идентификаций целевых белков системы UPS1 при уменьшении их концентрации до 10^{-10} М наблюдается в образце с наиболее высоким уровнем шума — 26 идентификаций, в чистом растворе UPS1 и в образце с низким показателем шума — 45 и 44 соответственно. Данная тенденция сохраняется и далее по мере снижения концентрации целевых белков. Таким образом, при наличии высокого уровня биологического шума уже в концентрации 10^{-12} М не удаётся зарегистрировать ни одного белка системы UPS1, но вместе с тем при отсутствии и низком уровне шума всё ещё удастся зарегистрировать 10 и 8 белков соответственно. Однако уже в концентрации 10^{-13} М даже в отсутствии биологического шума или при низких его показателях не было детектировано ни одного целевого белка. Следовательно, результаты,

полученные на модельном объекте UPS1, можно экстраполировать на белки, кодируемые генами 18 хромосомы, и спрогнозировать количество белков, которые можно зарегистрировать при использовании методики направленного масс-спектрометрического анализа. Таким образом, можно заключить, что методика позволяет зарегистрировать 100% белков в концентрации 10^{-9} М и выше, 56% белков в концентрации 10^{-10} М и лишь 19% белков в концентрации 10^{-11} М. Методика не рассчитана на детекцию белков в более низких концентрациях в комплексных биологических образцах. Таким образом, было продемонстрировано влияние двух факторов, которые влияют на регистрацию белков: концентрация целевых белков, наличие высокого уровня шума.

Для демонстрации влияния концентрации целевых белков на количество регистрируемых белков был также поставлен эксперимент с обратным концентрированием. Образцы UPS1 с добавлением и без добавления *E. coli* в концентрации 10^{-12} М были сконцентрированы в 100 раз до 10^{-10} М и проанализированы методом направленной масс-спектрометрии. Такой эксперимент позволил восстановить количество регистрируемых белков в случае чистого раствора UPS1 и раствора с низким уровнем шума до 45 и 44 соответственно. Однако концентрирование образца с высоким уровнем шума не приводит к улучшению результата, потому что вместе с целевыми белками происходит концентрирование белков матрицы. Чтобы снизить влияние матрицы необходимо применить методику фракционирования образца. Применение предварительного фракционирования методом обращённо-фазовой высокоэффективной жидкостной хроматографии в щелочных условиях элюции позволило восстановить количество идентификаций белков системы UPS1 до 26 (табл. 1).

Таким образом, комбинация методов концентрирования образца и фракционирования позволяет решить проблему низкой концентрации целевых белков, а также снижает уровень шума,

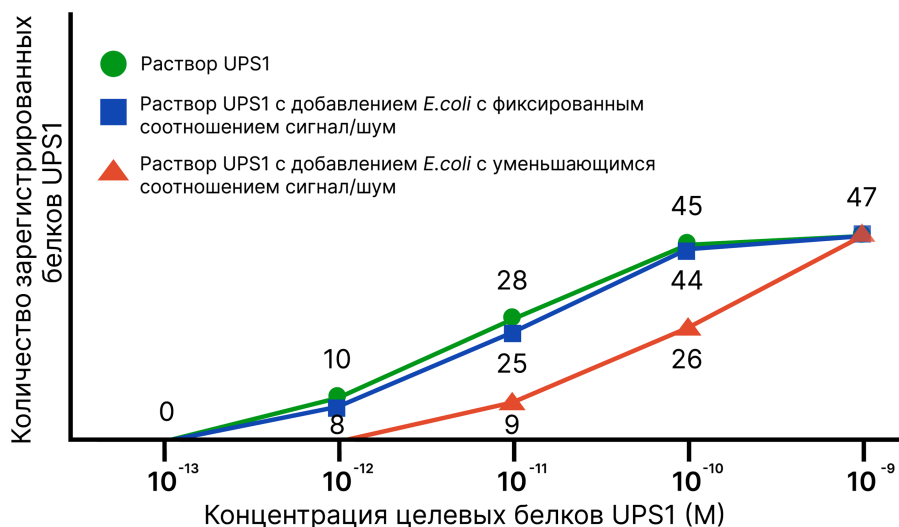


Рисунок 1. Количество зарегистрированных белков системы UPS1 в зависимости от концентрации целевых белков в модельных образцах.

Таблица 1. Результаты эксперимента с концентрированием образцов UPS1

Образец	Количество регистраций белков до проведения концентрирования образца	Количество регистраций белков после проведения концентрирования образца в 100 раз
Раствор UPS1	10	45
Раствор UPS1 с добавлением <i>E. coli</i> с фиксированным соотношением сигнал/шум	8	44
Раствор UPS1 с добавлением <i>E. coli</i> с уменьшающимся соотношением сигнал/шум	0	26*

Примечание: * – было проведено биохимическое фракционирование образца на 12 фракций.

что позволяет увеличить число идентификаций в реконструированной системе белков UPS1/*E. coli*. Следовательно, применение такого метода на биологическом образце — клеточной линии гепатоцеллюлярной карциномы человека HepG2 также может увеличить покрытие белков, кодируемых генами 18 хромосомы человека.

Направленный масс-спектрометрический анализ (SRM SIS) белков, кодируемых генами 18 хромосомы человека, клеточной линии гепатоцеллюлярной карциномы человека — HepG2

Для регистрации белков, кодируемых генами 18 хромосомы человека, был использован метод таргетного масс-спектрометрического анализа в комбинации с предварительным фракционированием в щелочных условиях элюции [7]. Было показано, что применение дополнительного этапа фракционирования позволяет зарегистрировать 26 уникальных белков, которые не были зарегистрированы в не фракционированном образце (рис. 2). По итогам применения обоих методов всего было зарегистрировано 94 белка, кодируемых генами 18 хромосомы человека. Большая часть белков (66%) была детектирована и в одномерном, и в двумерном анализе, однако небольшая часть белков (6,4%) была идентифицирована исключительно методом одномерного SRM анализа. Данное явление, как и значимая разность в измеренных концентрациях некоторых белков, зарегистрированных и в одномерном и в двумерном анализе, объясняется схемой сбора фракций при предварительном фракционировании.

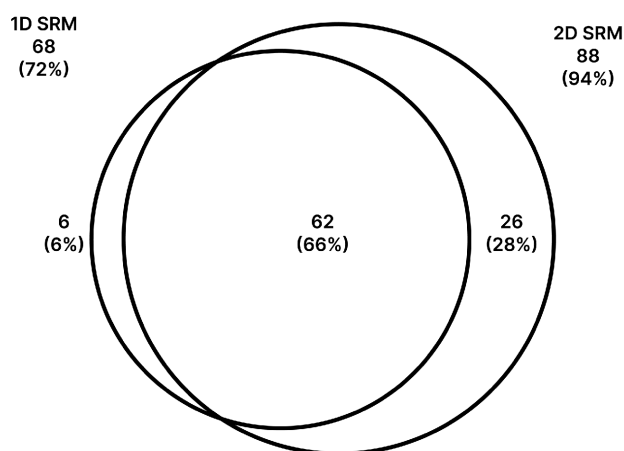


Рисунок 2. Диаграмма Эйлера-Венна, сравнение множеств уникальных белков, обнаруженных при одномерном сканировании не фракционированного образца (1D SRM) и фракционированного образца (2D SRM).

Фракции собираются непрерывно и во время переключения на следующую фракцию часть хроматографического пика пептида может быть собрана в одну фракцию, а другая часть уже в следующую. Таким образом, 6 пептидов могли оказаться в двух разных фракциях, а их концентрация быть ниже предела чувствительности прибора. Данный эффект мы не могли наблюдать при анализе белков системы UPS1. Возможно, этому помешало ограниченная выборка из 47 белков, в данном случае мы наблюдали 94 белка.

Диапазон измеренных концентраций составляет от 10^{-8} М до 10^{-12} М или от 300000 до 20 копий белка на клетку. Среди белков, зарегистрированных только с помощью таргетного анализа с предварительным фракционированием, присутствуют, в том числе, белки, представляющие транскрипционные факторы (Q92908-транскрипционный фактор GATA-6) и протеинкиназы (Q13464-Rho ассоциированная протеинкиназа 1, P31152-митоген-активируемая протеинкиназа 4), которые отвечают за передачу сигнала в клетке и регуляцию экспрессии генов [15–17]. Следовательно, с помощью данной методики можно получить дополнительную биологически значимую информацию об исследуемом объекте, которая недоступна при использовании стандартных методов анализа.

Транскриптопротеомный анализ клеточной линии HepG2

Данные, используемые в данном разделе, представляют собой результаты высокопроизводительного РНК секвенирования, выполненного на платформе Illumina. Для качественной оценки сходимости протеома и транскриптома клеточной линии HepG2 были выбраны разные отсечения по величине RPKM (Reads Per Kilobase Million, количество прочтений нормализованное на глубину прочтений и длину генов) транскриптомных данных, которые сравнивали с соответствующими белками, зарегистрированными методами направленной масс-спектрометрии [18]. Наибольшую сходимость протеомных и транскриптомных данных показывает отсечение на уровне $RPKM \geq 1$ (48%). При данном отсечении количество зарегистрированных транскриптов составляет 145, а соответствующих белков — 94 (табл. 2). При этом корреляционный анализ количественных протеомных и транскриптомных данных при соответствующих уровнях отсечения по RPKM показывает тенденцию к снижению получаемого коэффициента корреляции по мере возрастания значений порога отсечения по RPKM

ДЕТЕКЦИЯ НИЗКОКОПИЙНЫХ БЕЛКОВ В ПРОТЕОМНЫХ ИССЛЕДОВАНИЯХ

Таблица 2. Процентное соотношение обнаруженных уникальных белков, транскриптов и совместно зарегистрированных продуктов генов 18 хромосомы человека при различных порогах отсека по RPKM, также показан коэффициент корреляции количественных протеомных и транскриптомных данных при соответствующем отсеке

Порог отсека по RPKM (\geq)	Одновременно зарегистрированные белки и транскрипты [%]	Зарегистрированные уникальные транскрипты [%]	Зарегистрированные уникальные белки [%]	Коэффициент корреляции Спирмена (rs) количественных протеомных и транскриптомных данных	p-уровень значимости
0,01	42	54	3	0,6	$6,3 \times 10^{-9}$
0,1	43	51	6	0,6	$7,0 \times 10^{-8}$
0,5	47	44	9	0,6	$6,0 \times 10^{-8}$
1	48	42	10	0,6	$2,0 \times 10^{-7}$
5	40	28	31	0,5	$0,1 \times 10^{-3}$
10	31	20	50	0,4	$0,2 \times 10^{-1}$
50	8	2	90	0,4	0,4

от 0,6 до 0,4. Данные по измерению концентрации белка и RPKM соответствующих транскриптов представлены в Дополнительных материалах.

В этой работе данные высокопроизводительного РНК секвенирования были использованы в качестве ориентира экспрессии соответствующего белка в клеточной линии гепатоцеллюлярной карциномы человека HepG2. По аналогии с упрощённой белковой системой UPS1, где производитель гарантировал наличие 47 белков человека, в данном случае детекцию транскрипта рассматривали в качестве гарантированного признака (“золотого стандарта”) наличия соответствующего белка. Для 76 белков был зарегистрирован как транскрипт, так и соответствующий белок. Однако для 67 транскриптов не было зарегистрировано соответствующего белкового продукта (рис. 3). В отличие от набора UPS1 мы не можем знать концентрации белков в клетках. Даже если они и присутствуют в клетке, их концентрация может быть ниже, чем рабочий диапазон методики SRM анализа. Однако было зарегистрировано 16 белков, для которых не обнаружены соответствующие транскрипты. Наиболее вероятное объяснение данному феномену — большое время полураспада, стабильность молекулы белка относительно мРНК [19, 20]. При выделении мРНК, несмотря на все предосторожности, часть молекул может деградировать, в то время как более стабильные по своей химической структуре белки сохраняются. Также в живой клетке период полураспада белков многократно превосходит период полураспада соответствующих молекул мРНК [21, 22]. Измерения периода полураспада белков показывают, что в среднем белки до своей деградации выполняют свои функции десятки часов, в то время молекулы мРНК — считанные секунды и минуты [23, 24]. Таким образом, детекция белка соответствующей мРНК определяется как чувствительностью протеомного анализа, так и периодом полураспада молекулы мРНК в клетке.

Результаты данного исследования показывают, что применение таргетного анализа совместно с методикой предварительного фракционирования образца увеличивает глубину покрытия продуктов генов 18 хромосомы и позволяет зарегистрировать 94 белка, которые составляют 48% зарегистрированного транскриптома и 34% всех белков кодирующих генов, локализованных на 18 хромосоме человека.

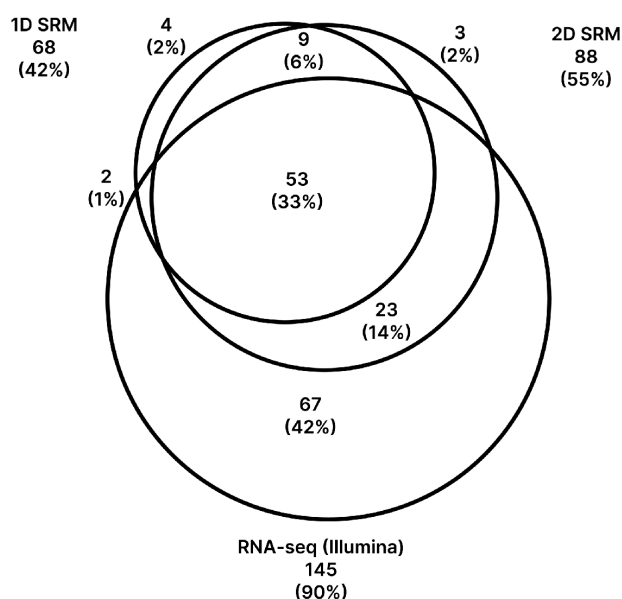


Рисунок 3. Диаграмма Эйлера-Венна, представляет собой сравнение трёх множеств, зарегистрированные в образце HepG2 транскрипты – RNA-seq, белки, кодируемые генами 18 хромосомы человека, зарегистрированные методом 1D SRM, белки, зарегистрированные методом 2D SRM.

ЗАКЛЮЧЕНИЕ

Основной проблемой таргетного анализа комплексных биологических образцов является наличие высокого динамического диапазона концентраций белков исследуемой системы. Эксперименты с реконструированной системой UPS1 показали, что чувствительность метода SRM анализа зависит от концентрации целевых белков, а также от наличия высокого уровня биологического шума. Применение 2D-фракционирования позволило частично решить проблему чувствительности и снизить динамический диапазон белков.

При анализе белков, кодируемых генами 18 хромосомы человека, в клеточной линии гепатоцеллюлярной карциномы человека HepG2 целевыми являются 264 белка, в то время как остальные белки, количество которых может превышать 6 миллионов являются матрицей или другими словами “биологическим шумом” [25].

Таким образом, применение предварительного фракционирования образца позволяет обогатить каждую фракцию целевыми белками и увеличить показатель соотношения сигнал/шум в каждой фракции. Белки в низкой концентрации (10^{-10} М– 10^{-12} М) составляют большую часть (76 из 94) протеома 18 хромосомы человека в клеточной линии HepG2. Соответственно, при переходе к анализу белков в концентрации 10^{-10} М и ниже мы не можем быть уверены, что были зарегистрированы все белки, которые экспрессируются в клетке, так как потери при переходе к аналогичной концентрации в реконструированной белковой системе UPS1 составляют 50%. При этом установлено, что в клетках обнаружены транскрипты 67 продуктов генов, для которых не зарегистрированы соответствующие белки. Полученные данные позволяют предположить, что префракционирование позволяет обнаружить дополнительные белки, однако всё ещё не позволяет выйти на уровень чувствительности транскриптомных методов.

БЛАГОДАРНОСТИ

Исследование выполнено с использованием оборудования Центра коллективного пользования “Протеом человека” в ИБМХ.

ФИНАНСИРОВАНИЕ

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

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Данная работа не включает исследования, в которых в качестве объекта выступают люди или животные.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

Дополнительные материалы доступны в электронной версии статьи на сайте журнала (pbmc.ibmc.msk.ru).

ЛИТЕРАТУРА

1. Anonymous (2010) The call of the human proteome. *Nat. Methods*, **7**(9), 661. DOI: 10.1038/nmeth0910-661
2. Hancock W., Omenn G., Legrain P., Paik Y.K. (2011) Proteomics, human proteome project, and chromosomes. *J. Proteome Res.*, **10**(1), 210. DOI: 10.1021/pr101099h
3. Zgoda V.G., Kopylov A.T., Tikhonova O.V., Moisa A.A., Pyndyk N.V., Farafonova T.E., Novikova S.E., Lisitsa A.V., Ponomarenko E.A., Poverennaya E.V., Radko S.P., Khmeleva S.A., Kurbatov L.K., Filimonov A.D., Bogolyubova N.A., Ilgisonis E.V., Chernobrovkin A.L., Ivanov A.S., Medvedev A.E., Mezentshev Y.V., Moshkovskii S.A., Naryzhny S.N., Ilina E.N., Kostjukova E.S., Alexeev D.G., Tyakht A.V., Govorun V.M., Archakov A.I. (2013) Chromosome 18 transcriptome profiling and targeted proteome mapping in depleted plasma, liver tissue and HepG2 cells. *J. Proteome Res.*, **12**(1), 123. DOI: 10.1021/pr300821n
4. Kopylov A.T., Zgoda V.G., Lisitsa A.V., Archakov A.I. (2013) Combined use of irreversible binding and MRM technology for low- and ultralow copy-number protein detection and quantitation. *Proteomics*, **13**(5), 727. DOI: 10.1002/pmic.201100460
5. Anderson N.L., Polanski M., Pieper R., Gatlin T., Tirumalai R.S., Conrads T.P., Veenstra T.D., Adkins J.N., Pounds J.G., Fagan R., Lobley A. (2004) The human plasma proteome: A nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics*, **3**(4), 311. DOI: 10.1074/mcp.M300127-MCP200
6. Michalski A., Cox J., Mann M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.*, **10**(4), 1785. DOI: 10.1021/pr101060v
7. Vavilov N., Ilgisonis E., Lisitsa A., Ponomarenko E., Farafonova T., Tikhonova O., Zgoda V., Archakov A. (2022) Number of detected proteins as the function of the sensitivity of proteomic technology in human liver cells. *Curr. Protein Pept. Sci.*, **23**(4), 290. DOI: 10.2174/1389203723666220526092941
8. Vavilov N.E., Zgoda V.G., Tikhonova O.V., Farafonova T.E., Shushkova N.A., Novikova S.E., Yarygin K.N., Radko S.P., Ilgisonis E.V., Ponomarenko E.A., Lisitsa A.V., Archakov A.I. (2020) Proteomic analysis of Chr 18 proteins using 2D fractionation. *J. Proteome Res.*, **19**(12), 4901. DOI: 10.1021/acs.jproteome.0c00856
9. Deinichenko K., Krasnov G., Radko S., Ptitsyn K., Shapovalova V., Timoshenko O., Khmeleva S., Kurbatov L., Kiseleva Y., Ilgisonis E., Pyatnitskiy M., Poverennaya E., Kiseleva O., Vakhrushev I., Tsvetkova A., Buromski I., Markin S., Zgoda V., Archakov A., Lisitsa A., Ponomarenko E. (2021) Human CHR18: “Stakhanovite” genes, missing and uPE1 proteins in liver tissue and HepG2 cells. *Biomedical Chemistry: Research and Methods*, **4**(1), e00144. DOI: 10.18097/BMCRM00144
10. Furey A., Moriarty M., Bane V., Kinsella B., Lehane M. (2013) Ion suppression: A critical review on causes, evaluation, prevention and applications. *Talanta*, **115**, 104. DOI: 10.1016/j.talanta.2013.03.048
11. Mei H., Hsieh Y., Nardo C., Xu X., Wang S., Ng K., Korfmacher W.A. (2003) Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: Application to drug discovery. *Rapid Commun. Mass Spectrom.*, **17**(1), 97. DOI: 10.1002/rcm.876
12. Muller L., Fornecker L., van Dorsselaer A., Cianfèrani S., Carapito C. (2016) Benchmarking sample preparation/digestion protocols reveals tube-gel being a fast and repeatable method for quantitative proteomics. *Proteomics*, **16**(23), 2953. DOI: 10.1002/pmic.201600288
13. Ramus C., Hovasse A., Marcellin M., Hesse A.M., Mouton-Barbosa E., Bouyssie D., Vaca S., Carapito C., Chaoui K., Bruley C., Garin J., Cianfèrani S., Ferro M., van Dorsselaer A., Burlet-Schiltz O., Schaeffer C., Couté Y., Gonzalez de Peredo A. (2016) Benchmarking quantitative label-free LC-MS data processing workflows using a complex spiked proteomic standard dataset. *J. Proteomics*, **132**, 51. DOI: 10.1016/j.jprot.2015.11.011

14. Matuszewski B.K., Constanzer M.L., Chavez-Eng C.M. (2003) Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. *Anal. Chem.*, **75**(13), 3019. DOI:10.1021/ac020361s
15. Kodo K., Nishizawa T., Furutani M., Arai S., Yamamura E., Joo K., Takahashi T., Matsuoka R., Yamagishi H. (2009) GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc. Natl. Acad. Sci. USA*, **106**(33), 13933. DOI: 10.1073/pnas.0904744106
16. Maekawa M., Ishizaki T., Boku S., Watanabe N., Fujita A., Iwamatsu A., Obinata T., Ohashi K., Mizuno K., Narumiya S. (1999) Signaling from Rho to the actin cytoskeleton through protein kinases ROCK and LIM-kinase. *Science*, **285**(5429), 895. DOI: 10.1126/science.285.5429.895
17. D  leris P., Trost M., Topisirovic I., Tanguay P.L., Borden K.L., Thibault P., Meloche S. (2011) Activation loop phosphorylation of ERK3/ERK4 by group I p21-activated kinases (PAKs) defines a novel PAK-ERK3/4-MAPK-activated protein kinase 5 signaling pathway. *J. Biol. Chem.*, **286**(8), 6470. DOI: 10.1074/jbc.M110.181529
18. Koch C.M., Chiu S.F., Akbarpour M., Bharat A., Ridge K.M., Barmore E.T., Winter D.R. (2018) A beginner's guide to analysis of RNA sequencing data. *Am. J. Respir. Cell Mol. Biol.*, **59**(2), 145. DOI: 10.1165/rcmb.2017-0430TR
19. Reimegard J., Tarbier M., Danielsson M., Schuster J., Baskaran S., Panagiotou S., Dahl N., Friedlander M.R., Gallant C.J. (2021) A combined approach for single-cell mRNA and intracellular protein expression analysis. *Commun. Biol.*, **4**(1), 624. DOI: 10.1038/s42003-021-02142-w
20. Shyu A.B., Greenberg M.E., Belasco J.G. (1989) The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. *Genes Dev.*, **3**(1), 60. DOI: 10.1101/gad.3.1.60
21. Treisman R. (1985) Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. *Cell*, **42**(3), 889. DOI: 10.1016/0092-8674(85)90285-5
22. Mathieson T., Franken H., Kosinski J., Kurzawa N., Zinn N., Sweetman G., Poeckel D., Ratnu V.S., Schramm M., Becher I., Steidel M., Noh K.M., Bergamini G., Beck M., Bantscheff M., Savitski M.M. (2018) Systematic analysis of protein turnover in primary cells. *Nat. Commun.*, **9**(1), 689. DOI: 10.1038/s41467-018-03106-1
23. Bevilacqua A., Ceriani M.C., Canti G., Asnaghi L., Gherzi R., Brewer G., Papucci L., Schiavone N., Capaccioli S., Nicolin A. (2003) Bcl-2 protein is required for the adenine/uridine-rich element (ARE)-dependent degradation of its own messenger. *J. Biol. Chem.*, **278**(26), 23451. DOI: 10.1074/jbc.M210620200
24. Yang E., van Nimwegen E., Zavolan M., Rajewsky N., Schroeder M., Magnasco M., Darnell J.E. Jr. (2003) Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. *Genome Res.*, **13**(8), 1863. DOI: 10.1101/gr.1272403
25. Ponomarenko E.A., Poverennaya E.V., Ilgisonis E.V., Pyatnitskiy M.A., Kopylov A.T., Zgoda V.G., Lisitsa A.V., Archakov A.I. (2016) The size of the human proteome: The width and depth. *Int. J. Anal. Chem.*, **2016**, 7436849. DOI: 10.1155/2016/7436849

Поступила в редакцию: 27. 04. 2024.
После доработки: 10. 06. 2024.
Принята к печати: 11. 06. 2024.

DETECTION OF LOW-COPY PROTEINS IN PROTEOMIC STUDIES: ISSUES AND SOLUTIONS

A.I. Archakov, N.E. Vavilov, V.G. Zgoda*

Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; *e-mail: n.vavilov95@gmail.com

Detection of low-copy proteins in complex biological samples is one of the most important issues of modern proteomics. The main reason for inefficient detection of low protein concentrations is the insufficient sensitivity of mass spectrometric detectors and the high dynamic range of protein concentrations. In this study we have investigated the possibilities and limitations of a targeted mass spectrometric analysis using the reconstructed system of standard proteins UPS1 (Universal Proteomic Standard 1) as an example. The study has shown that the sensitivity of the method is affected by the concentration of target proteins of the UPS1 system, as well as by a high level of biological noise modelled by proteins of whole *E. coli* cell lysate. The limitations of the method have been overcome by concentrating and pre-fractionating the sample peptides in a reversed phase chromatographic system under alkaline elution conditions. Proteomic analysis of the biological sample (proteins of the human hepatocellular carcinoma cell line HepG2 encoded by genes of human chromosome 18) showed an increase in the sensitivity of the method as compared to the standard targeted mass spectrometric analysis. This culminated in registration of 94 proteins encoded by genes located on human chromosome 18.

The whole English version is available at <http://pbmc.ibmc.msk.ru>.

Key words: proteomics; mass spectrometry; HepG2; human chromosome 18 gene products

Funding. The work was carried out within the framework of the Program of Fundamental Scientific Research in the Russian Federation for a long-term period (2021–2030) (No. 122030100168-2).

Received: 27.04.2024; revised: 10.06.2024; accepted: 11.06.2024.