

ХРОМОСОМОЦЕНТРИЧНЫЙ ПОДХОД К ПРЕОДОЛЕНИЮ ТРУДНОСТЕЙ В РЕАЛИЗАЦИИ МЕЖДУНАРОДНОГО ПРОЕКТА «ПРОТЕОМ ЧЕЛОВЕКА»

А. И. АРЧАКОВ¹, В. Г. ЗГОДА¹, А. Т. КОПЫЛОВ¹, С. Н. НАРЫЖНЫЙ^{1,2},
А. Л. ЧЕРНОБРОВКИН¹, Е. А. ПОНОМАРЕНКО¹, А. В. ЛИСИЦА¹

¹Научно-исследовательский институт биомедицинской химии им. В. Н. Ореховича РАМН;

²Петербургский институт ядерной физики Национального исследовательского центра «Курчатовский институт»

Международный проект «Протеом человека» (ППЧ), являющийся логическим продолжением проекта «Геном человека», начат 23 сентября 2010 года. В соответствии с геноцентричным подходом, целью ППЧ является подготовка каталога всех протеинов человека и расшифровка сети их взаимодействий. Трудности реализации ППЧ возникают вследствие того, что сам объект исследования — протеом — гораздо сложнее генома. Основной проблемой является недостаточная чувствительность протеомных методов, не позволяющая детектировать низко- и ультранизкокопийные протеины. Плохая воспроизводимость протеомных методов и отсутствие так называемого «золотого стандарта» является второй основной сложностью в реализации ППЧ. Третья проблема — это динамичный характер протеома, его нестабильность во времени. В статье рассматриваются возможные варианты преодоления этих сложностей, препятствующих успешному выполнению ППЧ.

Ключевые слова: проект «Геном человека», проект «Протеом человека», предел чувствительности, масс-спектрометрический метод мониторинга множественных реакций.

В 2010 году международная организация «Протеом человека» (Human Proteome Organization, HUPO) официально открыла глобальный проект «Протеом человека» (ППЧ). Первоочередной задачей проекта является выявление хотя бы одного протеина, кодируемого каждым из 20300 генов человека [1]. С помощью количественной масс-спектрометрии (MS) и аффинных реагентов в рамках проекта планируется охарактеризовать внутриклеточную локализацию протеинов и основные модифицированные формы [2]. Хромосомоцентричная часть ППЧ (Chromosome-centric, C-HPP) направлена на расширение знаний о протеоме человека путем глубокого понимания роли каждого гена. Фундаментальные аспекты, связанные с возникновением и развитием патологических состояний и выявлением новых биологических маркеров, исследуются в рамках второй части ППЧ, ориентированной на заболевания (Biology/Disease, B/D-HPP). В последнее десятилетие в рамках реализации инициатив HUPO по исследованию протеома различных тканей и органов (протеом плазмы, протеом печени, протеом мозга человека и др. [77]) были заложены основы для этой части проекта.

С 2008 года международной организацией «Протеом человека» организован и проведен

ряд научных совещаний, посвященных вопросам определения целей, подходов и методических рекомендаций к реализации ППЧ [78, 3]. Планируется, что результатом выполнения I этапа проекта (длительностью 6 лет) с использованием масс-спектрометрических подходов будут идентифицированы основные (немодифицированные) протеины человека в различных тканях; II этап ППЧ (длительностью 5 лет) направлен на идентификацию модифицированных протеинов, т.е. продуктов реализации однонуклеотидных полиморфизмов, а также протеинов, возникающих в ходе альтернативного сплайсинга или посттрансляционных модификаций.

Проект «Протеом человека» сравнивают с проектом «Геном человека» (ПГЧ), который был успешно завершен более десяти лет назад [4, 5]. Конечной целью ПГЧ было декодирование всего генома, последовательность которого контекстно-независима и индивидуальна для человека. Секвенирование целого генома в начале 1990-х казалось нереальной задачей, и на проведение работ было потрачено в общей сложности около 3 млрд. долл. США (1 основание было секвенировано за 1 долл. США) [6, 7]. Реализация геномного проекта привела к развитию технологий [8], и сейчас за 1 доллар США может быть секвенировано уже

100 млн. оснований [9]. Секвенирование генома человека привело к получению продукта, востребованного на медицинском рынке и представляющего большую ценность для фундаментальных исследований. Тем не менее, наши сегодняшние знания о функции генома ограничиваются в основном экспрессируемой частью, которая составляет не более 1,5% расшифрованной последовательности ДНК [10]. Создание аналогичного каталога протеинов в протеомике может стать хорошей основой для развития новых диагностических подходов и терапии [5].

По сравнению с геномными технологиями начала 1990-х, современные протеомные технологии более перспективны [13], поскольку в одном эксперименте могут быть выявлены до 10 тыс. протеинов [14]. Тем не менее, реализация ППЧ сопряжена с рядом сложностей, которые будут рассмотрены далее. Эти сложности — недостаточная чувствительность протеомных технологий, отсутствие «золотого стандарта» и нестабильность протеома во времени — связаны с фундаментальными различиями геномики и протеомики.

Недостаточная чувствительность протеомных технологий. Предел чувствительности аналитических методов

В геномных технологиях с помощью полимеразно-цепной реакции (ПЦР, 15) одна молекула в биологическом образце может быть амплифицирована для повышения ее концентрации до предела обнаружения. В протеомике ситуация противоположная, поскольку пока нет технологии для репликации молекул протеина, то основной проблемой реализации ППЧ является недостаточная чувствительность высокопроизводительных аналитических методов [16].

Наиболее чувствительные масс-спектрометрические детекторы могут обнаруживать аналит, присутствующий в концентрации до 10^{-18} молей на литр (М) [81]. Такая чувствительность оказывается достаточной для детекции протеинов, присутствующих в следовых количествах (менее 1000 молекул в 1 мкл образца), при этом в контексте биологических образцов предел чувствительности составляет 10^{-8} – 10^{-10} М, вследствие общей сложности и высокого динамического диапазона протеинов в биоматериале [17].

При исследовании протеома протеины-мишени заранее не определены, поэтому пробоподготовка и фракционирование, влияю-

щие на количество выявленных протеинов, даже более значимы, чем физические свойства масс-спектрометрических детекторов [18]. В 3D-экспериментах, которые обеспечивают всестороннее исследование протеома, протеины в нативном биологическом образце сначала разделяются гель-фильтрацией [14] или изоэлектрическим фокусированием [19], после чего каждая фракция гидролизует трипсином и образующиеся при этом пептиды анализируются с помощью 2D-LC-MS/MS. Такой метод идентификации протеинов, известный также как метод «многомерной идентификации» (MudPIT), привел к быстрому переходу от обнаружения 100 полипептидов [20] к обнаружению тысяч протеиновых молекул.

При исследовании протеома дрожжей с использованием LC-MS/MS и ловушки масс-спектрометра ионного циклотронного резонанса с преобразованием Фурье, были обнаружены примерно 2000 различных протеинов, при этом самый низкокопийный протеин присутствовал в 100 копиях протеина на клетку [21]. Учитывая, что клетка *Saccharomyces cerevisiae* составляет примерно 30 мкм^3 ($\sim 3 \times 10^{-14}$ л) [82], предел обнаружения одного протеина в клетке соответствует концентрации 10^{-8} М. Пять лет спустя, группа под руководством Манна с использованием комбинированного подхода на базе Orbitrap/MudPIT идентифицировала 10000 протеинов в клеточных лизатах: миозин 5B (NeXtProt NX_Q9ULV0) был обнаружен на уровне примерно пяти копий на 1 клетку HeLa, что соответствует концентрации 5×10^{-13} М [14]. Зависимость количества идентифицированных протеинов от технологии пробоподготовки наблюдается также в протеомных экспериментах, выполненных по стратегии «top-down» [23, 24].

Вышеприведенные примеры показывают, что разделение протеинов является решающим фактором для проведения панорамных протеомных исследований, но при этом в настоящее время только 70% основных (мастерных) протеинов могут быть определены на протеомном уровне с использованием масс-спектрометрических подходов [1]; масштабное разделение увеличивает расход биоматериала, приводя к неизбежной потере аналитов. Альтернативный подход к повышению чувствительности MS основан на таргетных (целевых) методах.

Таргетная (целевая) протеомика

Предварительная информация для ППЧ может быть получена с использованием тех-

нологии MudPIT; детальная каталогизация протеома (последовательный анализ протеинов каждого гена) проводится с помощью таргетной (целевой) масс-спектрометрии [3]. Таргетный подход, основанный на масс-спектрометрическом методе мониторинга множественных реакций (MPM), обеспечивает чувствительность до 10^{-14} М, позволяя обнаружить десятки протеиновых копий в дрожжевой клетке [25]. 47 различных протеинов в концентрации 1 мкг/мл (10^{-9} М) были обнаружены в плазме с помощью мультиплексного MPM [26]. Группе под руководством Карра удалось достичь предела чувствительности до 1 нг/мл (10^{-12} М) с использованием меток стабильными изотопами (SID) и MPM идентификации протеинов, добавленных в плазму крови человека [27]. Эта же группа применила метод MPM для количественного определения среднекопийных протеиновых маркеров повреждения миокарда, в том числе тропонинов [28]. Whiteaker с коллегами использовали антипептидные антитела, чтобы достичь предела обнаружения 10^{-12} М на 10 мкл плазмы и показали, что предел обнаружения в 10^{-15} М (пг/мл) возможен при обогащении пептидов из больших объемов плазмы (1 мл) [30]. Хотя метод SID-MPM развивался как метод количественного анализа, он может быть использован и для выявления модифицированных протеинов в биологических образцах [31].

Начиная с измерения нескольких пептидных аналитов, таргетный подход быстро развивается. Сегодня компания Biognosys (Цюрих, Швейцария) предлагает количественный анализ 100 протеинов в одном MPM-методе, оптимизируя время удерживания протеотипических пептидов [83]. Данная компания оказывает содействие хранилищу MPM данных [84], которые могут быть использованы для выбора так называемых «чрезвычайно успешных» (т.е. подходящих для MS) синтетических пептидов, чтобы разработать MPM-метод для анализа таргетных протеинов. Это достижение показывает, что потенциал MPM-метода далеко не исчерпан и его применение может обеспечить, фактически, неограниченные возможности для ППЧ.

Поскольку экспериментальный этап хромосомотцентричной части ППЧ связан с выявлением мастерных протеинов, то метод MPM может быть применен для выявления одного за другим протеинов в биологическом образце [33]. При таком подходе используются методы биоинформатики для разработки *in silico* списка переходов, которые выбираются в качестве

пиков с высоким уровнем ответа и воспроизводимости из перечня MS/MS фрагментов для исследуемого протеотипического пептида [26, 34]. Исходные переходы применяются для анализа биопробы в статическом режиме MPM с использованием трех- и четырех-LC/MS. Оценка измерений может проводиться с использованием автоматизированных статистических инструментов [35], а идентификация пептида основана на значительном совпадении между списком исходных переходов и спектрами образцов [85]. Повторение статических экспериментов MPM для нескольких (от двух до пяти) протеотипических пептидов протеина позволяет компилировать несколько групп переходов с соответствующими временами удерживания в протеинспецифический динамический MPM-метод. Эти белокспецифические переходы затем объединяются в мультиплексный анализ и проверяются с применением биопробы, которая уже использовалась для разработки отдельных методов.

Преодоление термодинамического барьера с использованием необратимого связывания

В нашей лаборатории MPM использовали для определения протеотипических пептидов, закодированных 18-й хромосомой немодифицированных протеинов в плазме крови, печени и экстрактах клеток линии HepG2 [36]. В настоящее время детектировано около 80% мастерных протеинов, кодируемых на этой хромосоме [86] при пределе обнаружения 10^{-15} М.

В нашем подходе мы исходим из предположения, что протеомные технологии ограничены обнаружением миллионов (в лучшем случае, тысяч) копий протеинов в биологическом материале. Теоретически успех в протеомике будет достигнут только тогда, когда концентрационный предел обнаружения понизится до уровня обнаружения одной молекулы в 1 л биоматериала (10^{-24} М), достигая обратного числа Авогадро (одна молекула на 1 л раствора) [16]. Предполагалось, что чувствительность детектирования протеина может быть увеличена с помощью необратимого химического связывания протеинов из больших объемов биоматериала. В серии экспериментов с применением атомно-силовой микроскопии и с использованием химического фишинга аналита с фотоактивированными антителами, нам удалось достичь уровня детекции 10^{-16} М против уровня 10^{-12} М, который наблюдается при обратимом связывании [37]. Дальнейшие эксперименты показали, что необратимое связывание протеинов с гранулами BrCN-сефарозы,

позволяет регистрировать протеотипические пептиды добавленных протеинов до уровня концентрации — 10^{-18} М, т. е. одна копия протеина в 1 мкл плазмы крови.

Для решения проблем чувствительности и успешной реализации ППЧ необходимо опираться на исследования транскриптома, в которых результаты выражаются не в единицах концентрации, а в количествах копий мРНК [38, 39]. Работая в диапазоне сверхнизких концентраций, удобно оперировать копиями протеина, а не концентрациями, так как это позволяет сравнить результаты транскриптомных и протеомных экспериментов [40]. Хотя концентрации протеина являются общепринятыми в протеомике [41], при чувствительности в моль/л, соответствующее количество копий протеина может быть легко подсчитано как произведение LOD_{CV}/NRA , при объеме образца, V , и LOD — чувствительности метода измерения, N_{RA} — обратное число Авогадро, $\sim 10^{-24}$ [16]. Например, концентрация 10^{-18} М соответствует одной молекуле протеина в 1 мкл, тогда как такая же концентрация в HepG2-клетке с объемом 10^{-13} л (каждая клетка имеет диаметр ~ 8 мкм) [42], приравнивается к 10^{-7} молекул в клетке, т. е. одна молекула протеина на 10 миллиардов клеток. Поскольку одна копия протеина в объеме HepG2-клетки соответствует концентрации 10^{-13} М, достигаемой современными масс-спектрометрическими методами, то для клеточной протеомики дальнейшее повышение чувствительности не требуется. Эта гипотеза была бы верна, если бы в масс-спектрометр можно было загрузить одну клетку; в реальности требуются тысячи и даже миллиарды клеток для получения протеинов в количествах, достаточных для масс-спектрометрических измерений. В обычном исследовании с использованием двумерного электрофореза Zappacosta с коллегами показали, что необходимо минимум 200 фмолей протеина (2×10^8 клеток для детекции протеина, присутствующего в 100 копиях/клетку) [43]. Для исследования протеома одной клетки, необходимо достичь чувствительности уровня одной молекулы протеина на десятки тысяч или даже миллионов клеток. Такая чувствительность позволяет произвести инвентаризацию протеома отдельного органа, например: 1 г печени содержит 129 миллионов клеток [44], поэтому одна копия протеина на 10 миллионов клеток нередкое явление в масштабе органа.

В завершение обсуждения чувствительности протеомных методов, следует отметить, что число видов протеинов, присутствующих в конкретном клеточном материале и биоло-

гических жидкостях, остается неизвестным. В клетке предполагается понижение динамического диапазона концентраций протеина, в противоположность плазме крови, которая имеет исключительный динамический диапазон — более 10 порядков [17]. Высокий динамический диапазон означает, что плазма представляет собой сложную смесь разнообразных протеинов из различных тканей и клеток [45]. При том, что любой протеин теоретически может присутствовать в плазме, экспериментально в ней обнаружено всего несколько тысяч протеинов [46]. Мы сообщали об экспериментальном подходе к оценке числа видов протеинов, присутствующих в плазме [37], согласно которому протеины в биопробе разделялись 2D-электрофорезом и визуализировались окрашиванием различной чувствительности в диапазоне от 10^{-6} до 10^{-12} М. Используя этот метод, подсчитано, что протеом плазмы теоретически состоит примерно из 10^6 различных типов протеинов.

Отсутствие «золотого стандарта»

Соответствие между чувствительностью аналитического метода и количеством детектируемых протеинов является основой для разработки нового подхода к протеомной стандартизации. На основании экспертной оценки, хромосомо-центричная парадигма (по крайней мере на первом этапе) позволит обойти проблеме низкой воспроизводимости и ошибочных идентификаций протеина.

Традиционно проблема стандартизации в протеомике основывается на нескольких очевидных соображениях [13, 47]. В соответствии с концепцией масштабирования протеомики для сохранения воспроизводимости и чувствительности необходимо руководствоваться стандартными рабочими процедурами (standard operating procedures — SOP) и преаналитическим контролем процессов [48]. SOP должны объединить отбор образцов и условия хранения клинических образцов. Например, для спинномозговой жидкости была показана существенная потеря всех А- β пептидов при замораживании образцов [49]. Стандартизация образцов ткани, полученных посмертно или с помощью биопсии, является еще более сложной задачей, поскольку изменения в ткани могут происходить уже через 20 мин после забора образца. Для решения проблемы изменчивости тканей, Espina и Muller предложили универсальный фиксирующий раствор для стабилизации и сохранения протеинов, нуклеиновых кислот и тканевой структуры [50].

Сильная вариабельность наблюдается и при аналитическом этапе обработки образцов, где для достижения более высокой чувствительности и увеличения покрытия протеома исследователи изменяют уже опубликованные протоколы. Влияние пробоподготовки, префракционирования, хроматографии и методов MS при анализе образцов ткани и клеточных лизатов были рассмотрены в работе Altelaar и Heck [51]. Использование сортировки флуоресцентно активированных клеток, лазерной микродиссекции, предварительного разделения методом SDS-PAGE и прямого LC-MS анализа в эксперименте позволяет варьировать такими параметрами как необходимое количество биоматериала и число обнаруженных протеинов. В зависимости от применяемого протеомного метода и от количества обработанного биоматериала, число идентифицируемых протеинов изменилось с 1006 до 6664 видов (протеина).

Таргетные методы на базе MRM (multiple reaction monitoring) обладают высокой воспроизводимостью, что существенно облегчает решение вопроса стандартизации их применения. В национальном институте рака (Мерилэнд, США), было проведено исследование «MultiLab», целью которого являлась оценка воспроизводимости и надежности SID-MRM-MS метода для определения количества пептидов и протеинов в плазме человека [52]. Семь различных протеинов использовались в качестве добавленных зондов в девяти концентрациях; каждая концентрация зонда измерялась в трех технических повторностях. Внутри- и межлабораторная изменчивость результатов измерений оценивалась по коэффициенту вариации (CV) со значением ниже 25%. На основе полученных результатов приведенного исследования был сделан вывод, что таргетный подход обеспечивает выполнение воспроизводимого измерения протеинов, присутствующих в концентрациях от высоких до средних (выше 10^{-8} M) в нефракционированной плазме с линейным динамическим диапазоном в три порядка. Однако, несмотря на последующий строгий SOP, несколько сигнатур пептидов не были обнаружены в MRM-анализе, указывая на то, что этот метод (как и другие) зависит от обработки проб, включая фракционирование, обессоливание и неполное ферментативное расщепление [52].

Различия в MS/MS поисковых алгоритмах и базах данных аминокислотных последовательностей являются источником общих несоответствий в перечне идентификаций протеинов. В сравнительном исследовании

оценивались пять наиболее популярных алгоритмов поиска MS/MS спектров протеинов в плазме крови с использованием лабораторного справочно-информационного фонда (СИФ) против базы данных последовательности Международного индекса протеина (IPI) [53]. В зависимости от поискового механизма, было обнаружено 526 пептидных идентификаций и лишь 335 (63%) из них были точно определены на основе консенсуса четырех алгоритмов, хотя все исходные спектры были сгенерированы в одной лаборатории. Последующее MultiLab исследование было выполнено с использованием контрольного образца (эквимоллярная смесь из 20 протеинов) и MS/MS спектров, полученных ведущими протеомными лабораториями [54]. Несмотря на относительно простую смесь протеинов, присутствующих в сравнительно высоких концентрациях (10^{-10} – 10^{-12} M, в зависимости от разведения сухих контрольных образцов), только семь из 27 лабораторий правильно определили все 20 протеинов. Централизованный анализ результатов позволил сделать вывод, что большинство лабораторий получают высококачественные MS данные для выявления всех 20 протеинов и выявленные несоответствия, скорее всего, вызваны различиями между алгоритмами анализа и базами данных.

Нестабильность базы данных последовательностей является еще одним источником несоответствий в идентификациях протеинов, полученных с помощью MS. Проблема рассматривалась в отношении четырех самых популярных баз данных: IPI, UniProtKB, NCBIInr и Ensemble [55], где наибольшей стабильностью обладает UniProtKB, тогда как изменения идентификаторов протеина в IPI отмечается примерно в 10% случаев.

Основной пример использования стандартизации можно наблюдать в проекте ПППЧ (Проект «Протеом плазмы крови человека»), в рамках которого был разработан подход для решения проблем, возникающих при сборе, интеграции, анализе и распространении полученных данных во всем мире [56]. Суть данного подхода состоит в развитии модели данных, инструктирующих выполнение принятых рекомендаций для получения и дальнейшей публикации данных идентификации пептидов и протеинов [57]. В соответствии с моделью данных был создан репозиторий, удовлетворяющий конкретным требованиям проекта. Такое строгое управление данными в результате позволило получить в рамках ПППЧ список из 3024 многопептидных идентификаций протеи-

нов в плазме крови человека, который стал справочником для различных типов анализа [58].

Наработки ПППЧ и результаты исследования опытного образца, полученные группой под руководством Bell, рекомендованного HUPO, показывают, что стандартизация может стать «узким местом» для ППЧ [46, 54]. По нашему мнению, участники международного проекта ППЧ должны сосредоточиться на решении проблемы стандартизации, используя имеющиеся возможности протеома [59], а не стандартизируя методы исследования протеома, как это предлагается сегодня.

Хромосоמו-центричный подход дает возможность пересмотреть концепцию стандартизации, так как при его использовании следует учитывать только чувствительность аналитического метода и общее количество протеина, взятого для анализа из биологического образца. Данные, представленные в работе Арчакова с соавт. [37] для 2DE, позволили нам предложить следующую формулу (см. дополнительные данные 3), выражающую число протеиновых видов в образце (N) как функцию двух параметров: LOD (limit of detection) и количество протеина в пробе (Q), введенные для выполнения аналитической процедуры; выглядят следующим образом:

$$N = \frac{a * Q}{LOD^b}, \quad (1)$$

где коэффициенты a и b зависят от типа биоматериала (например, $a = 6,4 \times 10^{-3}$ и $b = 0,47$ для плазмы крови, или $a = 4,3 \times 10^{-5}$ и $b = 0,79$ для HepG2 клеток). Конечно, формула также зависит от типа протеомной технологии.

LOD была предложена в качестве меры чувствительности протеомных технологий, выраженной в виде нижней границы определяемой концентрации протеина [16]. На основании формулы (1), можно сделать несколько предположений: например, понижение LOD приводит к идентификации большего числа видов протеина, а именно:

$$\text{если } LOD_1 < LOD_2, \text{ то } N_1 > N_2.$$

В противном случае, увеличивая количество биоматериала, можно получить больше видов протеина:

$$\text{если } Q_1 > Q_2, \text{ то } N_1 > N_2.$$

Ожидается также, что, используя метод с максимальным физически релевантным LOD (т. е. одна молекула протеина на 1 л), можно рассчитать общее число видов протеина при

условии отсутствия ограничений в количестве образца, а именно:

$$N \rightarrow N_{\max} \text{ at } LOD \rightarrow 0 \text{ и } Q \rightarrow \infty. \quad (2)$$

Методы, используемые для идентификации протеинов при хромосоמוцентричном подходе, тоже можно сравнивать, указывая LOD и количество протеина в образце Q . Мера стандартизации может быть корректно формализована как вероятность аналитического метода для расшифровки N протеинов в определенном типе биоматериала. Точность эксперимента будет выражена количественно в виде отклонения полученного значения N от ожидаемого значения, рассчитанного по формуле (1), учитывая значения LOD (характеристика аналитического канала информации) и Q (характеристика плана эксперимента).

Мы ожидаем, что хромосоמו-центричный подход даст нам возможность рассматривать проблему «золотого» стандарта под новым углом зрения. Например, число репрезентативных основных протеинов, кодируемых генами хромосомы 18, в настоящее время составляет около 280 [60]. На начальном этапе, не следует сосредотачиваться на выявлении всех протеинов, кодируемых хромосомой, поскольку не все последовательности экспрессируются и чувствительность современных методов ограничена [16, 41]. При использовании стандартных аналитических методов (на основе использования антител/MS-технологий), протокол должен обеспечивать воспроизводимое детектирование в конкретном образце биоматериала, по крайней мере, порядка 80% протеинов, кодируемых определенной хромосомой.

Протеомика – ситуационная наука

Принято считать, что протеом здорового человека динамически изменяется с течением времени. На основании этого можно предположить зависимость результатов экспериментов, как правило, от внешних факторов, влияющих на субъект исследования [47]. Удивительно, но до настоящего времени систематические протеомные исследования для подтверждения степени индивидуальных вариаций уровня протеина в плазме крови или в тканях человека – не проводились. Идея индивидуальной изменчивости протеома была заимствована из молекулярной физиологии и спортивной медицины, где изучался ответ здоровых людей, испытуемых на внешние раздражители. Большинство выводов было сделано по результатам исследований регуляторных молекул: изменения цистатионина плазмы после завтрака [61],

изменения в системе фактора I роста инсулина в условиях острого дефицита энергии [62], изменения в моделях подкласса липопротеинов низкой плотности после диеты с пониженным содержанием жира у мужчин [63] и многие другие. В исследованиях, связанных с измерением одиночного протеина плазмы – транстиретина, было показано, что его уровень может измениться в ответ на 30-минутное воздействие высокочастотного электромагнитного поля, создаваемого сотовыми телефонами [64]. Однако авторы согласились с тем, что наблюдаемые изменения в уровне протеинов плазмы также могли быть вызваны стрессом, который добровольцы испытали во время забора крови. Изучение влияния микрогравитации на здоровых добровольцев, показал удвоение уровня экспрессии $\beta 2$ -интегринов в нейтрофилах периферической крови через 65 дней после положения тела головой вниз, а после окончания гипокинезии – вернулся к исходному уровню [65]. Степень экспрессии протеина в ткани тоже может меняться: например, уровень определенных протеинов, полученных из пинеальной железы крысы показывал более чем пятикратные циркадные изменения [66]. С другой стороны, результаты высокопроизводительных LC-MS/MS экспериментов на мышах не показали существенных различий в протеоме мозга, обусловленных старением [67].

Немногие имеющиеся факты об изменениях транскриптома отдельного человека предполагают соответствующую динамичность протеома. Данных об мультипротеиновых исследованиях изменений протеома у млекопитающих опубликовано немного. На основании обзора исследований с применением 2-DE для анализа плазмы крови, был сделан вывод, что изменения яркости светового пятна в здоровой контрольной группе сопоставимы с изменением при болезни в сравнении с контрольной оценкой [69]. Corzett с соавт. изучили изменения протеома плазмы крови человека с использованием 2-D DIGE и отметили, что CV изменялся от 10 до 90% со средним значением 23% при субъект-объектном анализе и лишь 21 пятно протеинов имело CV > 50% [70]. Индивидуальное изменение характеризовалось средней CV 10%, которое сопоставимо с техническими изменениями по сообщению авторов [71].

Мета-анализ литературы выявил разницу на два порядка в оценке копийности протеинов плазмы для здоровых людей [72]. В 150 многокопийных протеинах плазмы с диапазоном концентраций от 10^{-3} до 10^{-8} M, уровень

большинства протеинов изменялся в два–три раза, но при этом присутствовали и высокодинамичные виды, такие как лизоцим, который изменялся от 10^{-6} до 10^{-8} M.

Основываясь на вышесказанном, третья проблема ППЧ становится очевидной: протеомика зависит от ситуации, где только одна последовательность генома может быть (условно) представлена в виде контекстно-независимой единицы. Последовательность генома является цифровым продуктом, который может быть предложен человеку и, однажды предоставленный, он сохраняет свое значение на длительное время. Например, американская компания 23andMe (Калифорния, США) продвигает высокоплотное генотипирование отдельных геномов, чтобы обеспечить заказчика своевременным обновлением информации, связанной с медицинским воздействием вновь открытых мутаций у населения [88].

Такой подход неприменим к протеому человека, так как невозможно получить протеом в виде постоянного снимка, неизменного в своем значении. Вместо этого участники ППЧ должны сосредоточиться на разработке технологии «кадрового фильма» [73]. Такой «фильм» будет информативен, если хромосомотричный подход предполагает углубленное изучение aberrантных функций, продиктованное полиморфизмами, мутациями, изоформами или модификацией состояния протеинов [13]. Применяя таргетный подход, при котором особое внимание уделяется ограниченному набору разработанных протеинов, можно будет проследивать временной паттерн протеома [74].

На самом деле ожидаемая изменчивость протеома способствует развитию пилотной фазы ППЧ, так как некоторые недостающие протеины могут быть выявлены путем сбора образцов внутри и между субъектами. Проблему ППЧ, связанную с изменчивостью протеома, можно решить обычным путем, если начать с отбора контрольной группы лиц, тщательно проверенных современными клиническими методами. Для русской части проекта [36], мы использовали группу здоровых добровольцев, выдержавших жесткие условия аэрокосмической программы [75]. Таким способом можно вывести средний протеом плазмы, характерный для «абсолютного здоровья». Альтернативный подход состоит в осмотре относительно здоровых людей, не имеющих явных патологических показаний [71,76]. Средний протеом плазмы крови человека можно суммировать по данным этой группы, аналогично многим исследованиям, включая крупномасштабный ПППЧ [46].

При любом из этих подходов должна быть выполнена осевая выборка с использованием индивидуального «внутреннего контроля» [74]. Сканирование временных изменений позволит проводить более чувствительные эксперименты, которые могут привести к открытию биомаркеров более эффективно, чем при простом сравнении одиночных образцов, полученных при исследовании групп сильно непохожих людей [73, 76]. С развитием MRM методов для нефракционированной плазмы [52] мультиплексный анализ протеинов-мишеней будет методом выбора для исследования изменчивости протеома при достаточно высоком уровне чувствительности.

Итак, три проблемы – «узкие места» в протеомных технологиях, а именно: чувствительность, отсутствие «золотого стандарта» и динамичность в копиях протеинов, должны быть решены в рамках реализации ППЧ. В зависимости от успехов в преодолении этих «узких мест», выполнение ППЧ позволит внести существенный практический вклад в развитие медицины.

Работа выполнена в рамках программы «Протеом человека» Российской академии медицинских наук при поддержке Министерства образования и науки РФ (соглашение № 8274).

ХРОМОСОМОЦЕНТРИЧНИЙ ПІДХІД ДО ПОДОЛАННЯ ТРУДНОЩІВ У РЕАЛІЗАЦІЇ МІЖНАРОДНОГО ПРОЕКТУ «ПРОТЕОМ ЛЮДИНИ»

А. І. Арчаков¹, В. Г. Згода¹, А. Т. Копилов¹, С. Н. Наріжний^{1,2}, А. Л. Чернобровкін¹, Є. А. Пономаренко¹, А. В. Лисиця¹

¹Науково-дослідний інститут біомедичної хімії ім. В. М. Ореховича РАН;

²Петербурзький інститут ядерної фізики Національного дослідницького центру «Курчатовський інститут»

Міжнародний проект «Протеом людини» (ППЛ), який є логічним продовженням проекту «Геном людини», розпочато 23 вересня 2010 року. Згідно з геноцентричним підходом метою ППЛ є підготовка каталогу всіх протеїнів людини і розшифрування шляхів їх взаємодії. Труднощі реалізації ППЧ виникають внаслідок того, що сам об'єкт дослідження – протеом – набагато складніше геному. Головною проблемою є недостатня чутливість протеомних методів, яка не дозволяє детектувати низько- і ультранизькокопійовані протеїни. По-

гане відтворювання протеомних методів і відсутність так званого «золотого стандарту» є другою основною складністю в реалізації ППЛ. Третя проблема – це динамічний характер протеома, його нестабільність у часі. У статті розглядаються можливі варіанти подолання цих складностей, які є перешкодою успішному виконанню ППЛ.

Ключові слова: проект «Геном людини», проект «Протеом людини», межа чутливості, мас-спектрометричний метод моніторингу множинних реакцій.

CHROMOSOMOCENTRIC APPROACH TO OVERCOMING DIFFICULTIES IN IMPLEMENTATION OF INTERNATIONAL PROJECT HUMAN PROTEOME

A. I. Archakov¹, V. G. Zgoda¹, A. T. Kopylov¹, S. N. Naryzhny^{1,2}, A. L. Chernobrovkin¹, E. A. Ponomarenko¹, A. V. Lisitsa¹

¹V. N. Orekhovich Scientific-Research Institute of Biomedical Chemistry, Russian Academy of Medical Sciences;

²Petersburg Nuclear Physics Institute at National Research Centre Kurchatov Institute

The international project *Human Proteome* (PHP), being a logical continuation of the project *Human Genome*, was started on September 23, 2010. In correspondence with the genocentric approach, the PHP aim is to prepare a catalogue of all human proteins and to decipher a network of their interactions. The PHP implementation difficulties arise because the research subject itself – proteome – is much more complicated than genome. The major problem is the insufficient sensitivity of proteome methods that does not allow detecting low- and ultralow-copy proteins. Bad reproducibility of proteome methods and the lack of so-called «gold standard» is the second major complicacy in PHP implementation. The third problem is the dynamic character of proteome, its instability in time. The paper deals with possible variants of overcoming these complicacies, preventing from successful implementation of PHP.

Key words: project Human Genome, project Human Proteome, sensitivity limit, mass-spectrometric method of monitoring of multiple responses.

1. *Legrain P., Aebersold R., Archakov A. et al. // Mol. Cell. Proteomics. – 2011. – 10. – M111.009993.*

2. Paik Y-K., Jeong S-K., Omenn G. S. et al. // Nat. Biotechnol. – 2012. – **30**. – P. 221–223.
3. Paik Y-K., Omenn G. S., Uhlen M. et al. // J. Proteome Res. – 2012. – **11**. – P. 2005–2013.
4. *The big ome* // Nature. – 2008. – **452**. – P. 913–914.
5. *A gene-centric human proteome project: HUPO – the Human Proteome organization* // Mol. Cell. Proteomics. – 2010. – **9**. – P. 427–429.
6. Landier E. S., Linton L. M., Birren B. et al. // Nature. – 2001. – **409**. – P. 860–921.
7. Venter J. C., Adams M. D., Myers E. W. et al. // Science (New York, NY). – 2001. – **291**. – P. 1304–1351.
8. Voelkerding K. V., Dames S. A., Durtschi J. D. // Clin. Chem. – 2009. – **55**. – P. 641–658.
9. Snyder M., Du J., Gerstein M. // Genes Dev. – 2010. – **24**. – P. 423–431.
10. Tejedor J. R., Valcárcel J. // Nature. – 2010. – **465**. – P. 45–46.
11. Via M., Gignoux C., Burchard E. G. // Genome Medicine. – 2010. – **2**. – P. 3.
12. Rosenbloom K. R., Dreszer T. R., Pheasant M. et al. // Nucleic Acids Res. – 2010. – **38**. – D620–5.
13. Nilsson T., Mann M., Aebersold R. et al. // Nat. Methods. – 2010. – **7**. – P. 681–685.
14. Nagaraj N., Wisniewski J. R., Geiger T. et al. // Mol. Syst. Biol. – 2011. – **7**. – P. 548.
15. Saiki R. K., Gelfand D. H., Stoffel S. et al. // Science (New York, NY). 1988. – **239**. – P. 487–491.
16. Archakov A. I., Ivanov Y. D., Lisitsa A. V., Zgoda V. G. // Proteomics. – 2007. – **7**. – P. 4–9.
17. Anderson N. L., Anderson N. G. // Mol. Cell. Proteomics. – 2002. – **1**. – P. 845–867.
18. Zgoda V. G., Moshkovskii S. A., Ponomarenko E. A. et al. // Proteomics. – 2009. – **9**. – P. 4102–4105.
19. Malmström J., Lee H., Nesvizhskii A. I. et al. // J. Proteome Res. – 2006. – **5**. – P. 2241–2249.
20. Washburn M. P., Wolters D., Yates J. R. // Nat. Biotechnol. – 2001. – **19**. – P. 242–247.
21. de Godoy L. M. F., Olsen J. V., de Souza G. A. et al. // Genome Biol. – 2006. – **7**. – R50.
22. Klose J. // Electrophoresis. – 2009. – **30**, Suppl. 1. P. S142–9.
23. Nedelkov D., Kiernan U. A., Niederkofler E. E. et al. // Mol. Cell. Proteomics. – 2006. – **5**. – P. 1811–1818.
24. Tran J. C., Zamdborg L., Ahlf D. R. et al. // Nature. – 2011. – **480**. – P. 254–258.
25. Kiyonami R., Schoen A., Prakash A. et al. // Mol. Cell. Proteomics. – 2011. – **10**. – M110.002931.
26. Anderson L., Hunter C. L. // Mol. Cell. Proteomics. – 2006. – **5**. – P. 573–588.
27. Keshishian H., Addona T., Burgess M. et al. // Mol. Cell. Proteomics. – 2007. – **6**. – P. 2212–2229.
28. Keshishian H., Addona T., Burgess M. et al. // Mol. Cell. Proteomics. – 2009. – **8**. – P. 2339–2349.
29. Stahl-Zeng J., Lange V., Ossola R. et al. // Mol. Cell. Proteomics. – 2007. – **6**. – P. 1809–1817.
30. Whiteaker J. R., Zhao L., Anderson L., Paulovich A. G. // Mol. Cell. Proteomics. – 2010. – **9**. – P. 184–196.
31. Unwin R. D., Griffiths J. R., Leverenz M.K. et al. // Mol. Cell. Proteomics. – 2005. – **4**. – P. 1134–1144.
32. Unwin R. D., Griffiths J. R., Whetton A. D. // Nat. Protoc. – 2009. – **4**. – P. 870–877.
33. Berle M., Kroksveen A. C., Haaland O. A. et al. // Fluids and Barriers of the CNS. – 2011. – **8**. – P. 19.
34. Sherwood C. A., Eastham A., Lee L. W. et al. // J. Proteome Res. – 2009. – **8**. – P. 4396–4405.
35. Reiter L., Rinner O., Picotti P. et al. // Nat. Methods. – 2011. – **8**. – P. 430–435.
36. Archakov A., Aseev A., Bykov V. et al. // Proteomics. – 2011. – **11**. – P. 1853–1856.
37. Archakov A., Ivanov Y., Lisitsa A., Zgoda V. // Proteomics. – 2009. – **9**. – P. 1326–1343.
38. Wang Z., Gerstein M., Snyder M. // Nat. Rev. Genet. – 2009. – **10**. – P. 57–63.
39. Taniguchi Y., Choi P. J., Li G-W. et al. // Science. – 2010. – **329**. – P. 533–538.
40. Schwanhüsser B., Busse D., Li N. et al. // Nature. – 2011. – **473**. – P. 337–342.
41. Anderson N. L., Anderson N. G., Pearson T. W. et al. // Mol. Cell. Proteomics. 2009. – **8**. – P. 883–886.
42. Dandoy P., Meunier C. F., Michiels C., Su B-L. // PLoS One. – 2011. – **6**. – e20983.
43. Zappacosta F., Huddleston M. J., Karcher R. L. et al. // Anal. Chem. – 2002. – **74**. – P. 3221–3231.
44. Marcos R., Monteiro R. A. F., Rocha E. // Liver Int. – 2006. – **26**. – P. 116–124.
45. Shen Y., Kim J., Strittmatter E. F. et al. // Proteomics. – 2005. – **5**. – P. 4034–4045.
46. Omenn G. S., States D. J., Adamski M. et al. // Proteomics. – 2005. – **5**. – P. 3226–3245.
47. Hamacher M., Meyer H. E. // Expert Rev. Proteomics. – 2005. – **2**. – P. 1–3.
48. Migneault I., Hunter J. M. // Bioanalysis. – 2009. – **1**. – P. 149–1163.
49. Bibl M., Esselmann H., Otto M. et al. // Electrophoresis. – 2004. – **25**. – P. 2912–2918.
50. Espina V., Mueller C. // Methods Mol. Biol. – 2012. – **823**. – P. 49–57.

51. Altelaar A. M., Heck A. J. // *Curr. Opin. Chem. Biol.* – 2012. – **16**. – P. 206–213.
52. Addona T. A., Abbatiello S. E., Schilling B. et al. // *Nat. Biotechnol.* – 2009. – **27**. – P. 633–641.
53. Kapp E. A., Schütz F., Connolly L. M. et al. // *Proteomics.* – 2005. – **5**. – P. 3475–3490.
54. Bell A. W., Deutsch E. W., Au C. E. et al. // *Nat. Methods.* – 2009. – **6**. – P. 423–430.
55. Griss J., Côté R. G., Gerner C. et al. // *Mol. Cell. Proteomics.* – 2011. – **10**. – M111.008490.
56. Adamski M., Blackwell T., Menon R. et al. // *Proteomics.* – 2005. – **5**. – P. 3246–3261.
57. Carr S., Aebersold R., Baldwin M. et al. // *Mol. Cell. Proteomics.* – 2004. – **3**. – P. 531–533.
58. Klie S., Martens L., Vizcaino J. A. et al. // *J. Proteome Res.* – 2008. – **7**. – P. 182–191.
59. Geiger T., Wehner A., Schaab C. et al. // *Mol. Cell. Proteomics.* – 2012. – **11**. – M111.014050.
60. *The Universal Protein Resource (UniProt) in 2010* // *Nucleic Acids Res.* – 2010. – **38**. – P. D142–8.
61. Guttormsen A. B., Solheim E., Refsum H. // *Am. J. Clin. Nutr.* – 2004. – **79**. – P. 76–79.
62. Alemany J. A., Nindl B. C., Kellogg M. D. et al. // *J. Appl. Physiol.* – 2008. – **105**. – P. 58–64.
63. Dreon D. M., Fernstrom H. A., Miller B., Krauss R. M. // *FASEB J.* – 1994. – **8**. – P. 121–126.
64. Yuasa K., Arai N., Okabe S. et al. // *Clin. Neurophysiol.* – 2006. – **117**. – P. 900–905.
65. Choukèr A., Thiel M., Baranov V. et al. // *J. Appl. Physiol.* – 2001. – **90**. – P. 1736–1743.
66. Møller M., Sparre T., Bache N. et al. // *Proteomics.* – 2007. – **7**. – P. 2009–2018.
67. Walther D. M., Mann M. // *Mol. Cell. Proteomics.* – 2011. – **10**. – M110.004523.
68. Sonna L. A., Fujita J., Gaffin S. L., Lilly C. M. // *J. Appl. Physiol.* – 2002. – **92**. – P. 1725–1742.
69. Trifonova O., Larina I., Grigoriev A. et al. // *Expert Rev. Proteomics.* – 2010. – **7**. – P. 431–438.
70. Corzett T. H., Fodor I. K., Choi M. W. et al. // *J. Biomed. Biotechnol.* – 2010. – **2010**. – 258494.
71. Corzett T. H., Fodor I. K., Choi M. W. et al. // *J. Proteome Res.* – 2006. – **5**. – P. 2611–2619.
72. Hortin G. L., Sviridov D., Anderson N. L. // *Clin. Chem.* – 2008. – **54**. – P. 1608–1616.
73. Chen R., Mias G. I., Li-Pook-Than J. et al. // *Cell.* – 2012. – **148**(6). – P. 1293–1307.
74. Li G-W., Xie X. S. // *Nature.* – 2011. – **475**. – P. 308–315.
75. Bogomolov V. V., Castrucci F., Comtois J-M. et al. // *Aviat. Space Environ. Med.* – 2007. – **78**. – P. 1162–1169.
76. Jackson D., Herath A., Swinton J. et al. // *Proteomics .Clin. Appl.* – 2009. – **3**. – P. 394–407.
77. <http://www.hupo.org/>
78. <http://c-hpp.org/meeting/>
79. <http://hupo.org/research/hpp/>
80. www.fixingproteomics.org/
81. http://www.chem-agilent.com/cimg/6490_Brochure.pdf
82. http://biochemie.web.med.uni-muenchen.de/Yeast_Biol/02%20Yeast%20Cell%20Architecture%20and%20Function.pdf
83. www.biognosys.ch
84. www.mrmbase.ch
85. http://www.chem.agilent.com/Library/applications/5990-8912en_lo.pdf
86. www.kb18.ru
87. <http://www.icm.jhu.edu/theses/theses/yungphd.pdf>
88. <https://www.23andme.com/>