

УДК 577.1, 57.088

МЕТОДЫ ВЫЧИСЛИТЕЛЬНОЙ ИНТЕРАКТОМИКИ В ВОПРОСАХ ВЗАИМОДЕЙСТВИЯ ПРОТЕОФОРМ ЧЕЛОВЕКА

Обзор

© 2020 Е.В. Поверенная*, О.И. Киселева, А.С. Иванов, Е.А. Пономаренко

НИИ биомедицинской химии имени В.Н. Ореховича, 119121 Москва, Россия;
электронная почта: k.poverennaya@gmail.com

Поступила в редакцию 25.03.2019

После доработки 16.09.2019

Принята к публикации 07.10.2019

Для человека известно около 20 000 белок-кодирующих генов, которые могут быть транслированы в миллионы уникальных видов белков (протеоформ). Протеоформы, кодируемые одним геном, зачастую отличаются по функции, что соответствует различиям в белковых партнерах. Взаимодействуя между собой, протеоформы образуют сеть, отражающую динамически изменяющиеся клеточные процессы в организме. Нарушение белок-белковых взаимодействий (ББВ) вызывает изменение в топологии сети, что зачастую приводит к возникновению патологических процессов. Изучение протеоформ – это относительно новая задача в протеомике, а потому экспериментальных работ по взаимодействию протеоформ немного. Биоинформатические инструменты позволяют решать ряд задач, комплементарно дополняя и обогащая экспериментальные результаты, в частности, расширяя возможности в исследовании взаимодействия протеоформ.

КЛЮЧЕВЫЕ СЛОВА: белок-белковые взаимодействия, интерактомика, биоинформатика.

DOI: 10.31857/S0320972520010066

Функционирование живой клетки осуществляется за счет множества различных молекулярных взаимодействий между белками, нуклеиновыми кислотами, пептидами, метаболитами и т.д. Несмотря на то что понятие «интерактом» комплексное, объединяющее все виды взаимодействий, наиболее часто его употребляют в отношении белок-белковых взаимодействий (ББВ). Именно белковые взаимодействия играют ключевую роль в биологических процессах [1].

Под взаимодействием белков подразумеваются связь двух и более белковых молекул, обеспечивающая выполнение биологической функции. Различают физические взаимодействия и взаимодействия в составе сети [2]. Физические взаимодействия подразделяются на прямые (бинарные) и комплексные. Группа белков, которые взаимодействуют друг с другом одновременно, называется белковым комплексом.

Существуют различные экспериментальные методы выявления белок-белковых взаимодействий [3]. Основными являются двугибридная

дрожжевая система (yeast two-hybrid, Y2H) [4] и аффинное выделение белкового комплекса с последующим масс-спектрометрическим анализом (affinity purification–mass spectrometry, AP-MS) [5].

Технология Y2H основана на активации экспрессии репортерных генов, которая возникает в результате взаимодействия между целевым белком-«наживкой» и его партнером белком-«добычей» и может быть детектирована.

За 25 лет с момента изобретения первого варианта метода Y2H [6] было разработано множество альтернативных протоколов, позволяющих, например, использовать помимо дрожжей клетки других организмов [7], а также исследовать взаимодействия между белками и нуклеиновыми кислотами: ДНК (one-hybrid, 1H, [8]) и РНК (three-hybrid, 3H, [9]). Преимуществом метода Y2H является простота и возможность автоматизации анализа, что позволяет масштабировать эксперименты.

Метод AP-MS состоит в «вылавливании» белков-партнеров на иммобилизованный белок-«наживку» с последующим протеолизом и масс-спектрометрическим анализом пептидных фрагментов образовавшихся комплексов, смываемых с колонки [10, 11]. Масс-спектрометрическая часть метода позволяет «регистрировать»

Принятые сокращения: AP-MS – аффинное выделение белкового комплекса с последующим масс-спектрометрическим анализом, ББВ – белок-белковые взаимодействия, Y2H – двугибридная дрожжевая система, ПТМ – посттрансляционные модификации.

* Адресат для корреспонденции.

динамично изменяющиеся белковые комплексы, учитывая посттрансляционные модификации (ПТМ), которые могут обладать биологической функцией. Следует также отметить гибкость метода AP-MS. Она выражается в возможности использования нативных антител к эндогенным «наживкам» или в использовании эпителиальных меток, что в первом случае позволяет проводить очистку белков в естественной среде клеточного лизата и учитывать взаимодействия сразу нескольких природных форм белка, а во втором – позволяет проводить эксперимент даже в отсутствие природного антитела [3]. Существует несколько десятков модификаций метода, касающихся не только способа «вылавливания» целевого белка, но и вариантов пробоподготовки и очистки белкового комплекса [5].

Оба метода не лишены недостатков, основной из которых – высокая доля ложных результатов: ложноположительных, при которых реального взаимодействия между белками нет, и ложноотрицательных, при которых существующие ББВ не детектируются. Методы Y2H, сфокусированные на выявлении бинарных белковых взаимодействий, характеризуются высокой долей ложных идентификаций в силу ряда причин (например, неестественно высокой концентрации фьюжн-белков в тестовой системе и неспецифичности аффинных реагентов) [12]. Для методов, основанных на AP-MS, ложноположительные результаты встречаются значительно реже, однако при экстракции и «отмывке» белковых комплексов утрачивается часть взаимодействий. Суммарно доля взаимодействий, установленных методом Y2H, достигает 20%, а методом AP-MS – 55% [11].

С физической точки зрения взаимодействие белков характеризуется константой диссоциации. Для определения кинетики физического взаимодействия белков в большинстве случаев используют оптические биосенсоры [13]. Однако число белковых пар, для которых возможно измерить константы диссоциации, весьма ограничено в силу сложности получения чистых белков в достаточном количестве для проведения экспериментальных работ.

Накопление большого массива интерактомных данных, полученных опытным путем, и их противоречивость, а также изначальная сложность объектов исследования привели к активному развитию биоинформатических методов. В первую очередь методы *in silico* используются для предсказания и описания новых ББВ на основе ранее полученных сведений [14]. Предсказательные алгоритмы базируются на анализе геномных данных, аминокислотных последовательностей, данных о трехмерных структурах

белков и результатах транскриптомных и протеомных экспериментов. Отдельной задачей биоинформатики является верификация сведений о ББВ, заключающаяся в определении взаимодействий в результатах, полученных масс-спектрометрически [15]. Основываясь на предположении, что физически взаимодействующие белки функционально связаны между собой, вычислительные подходы используются также для предсказания функций белков [16].

Кроме того, построение и анализ интерактомных сетей возможны только с помощью вычислительных инструментов, учитывающих комплексность данных, обработка которых требует недостижимых человеку когнитивных способностей. Интеграция экспериментальных и биоинформатических методов, направленная на детальное изучение аспектов взаимодействия белков и их форм в клетке, представляет собой современную интерактомику (рисунок).

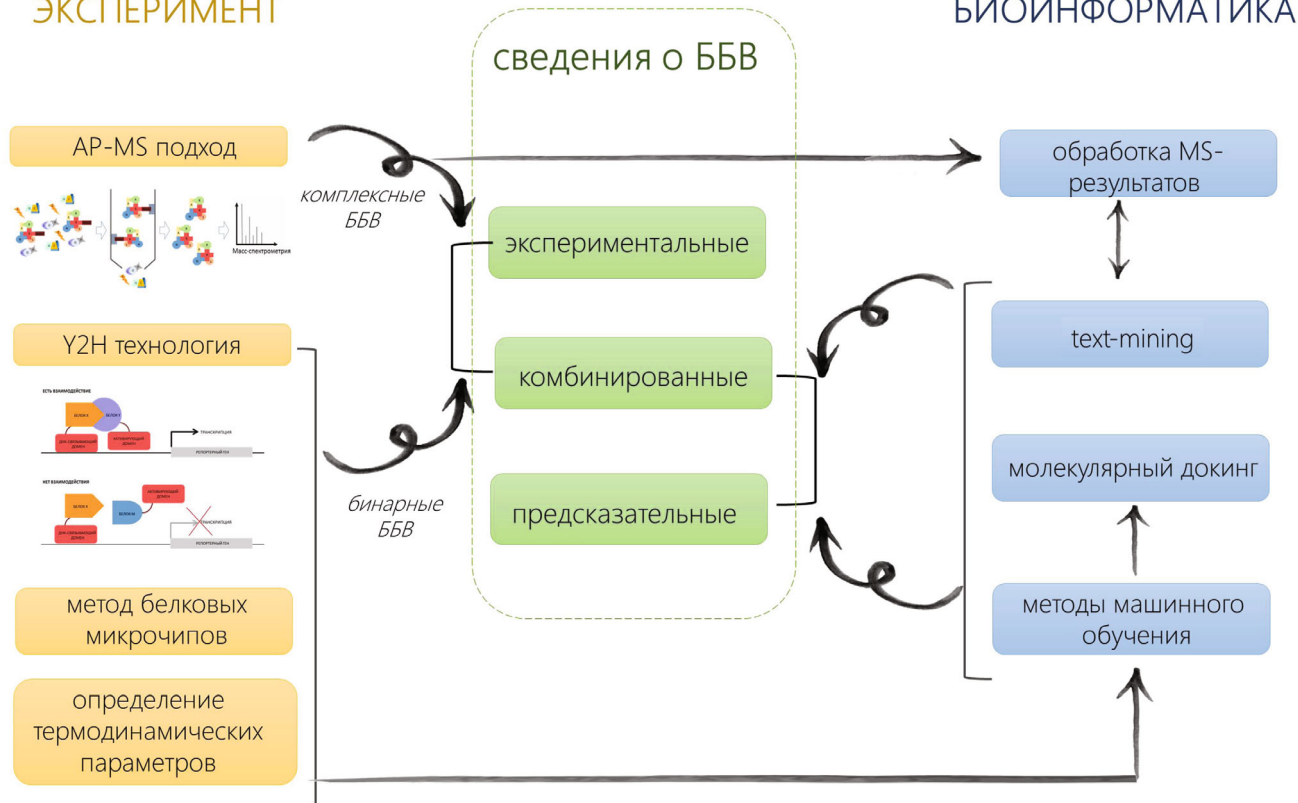
ИНТЕРАКТОМ ЧЕЛОВЕКА

По объему накопленных сведений человек является наиболее изученным организмом, согласно протеомному ресурсу UniProtKB (www.uniprot.org) [17]. Более 90% белок-кодирующих генов человека были подтверждены на протеомном уровне, при этом сведения о ББВ имеются для 93%. Примечательно, что в ряде случаев наличие данных о ББВ является доказательством существования белка.

Доказательство трансляции белка является критичным условием для достоверности белок-белковых взаимодействий. В большинстве предсказательных методов данной информацией пренебрегают, предполагая, что если взаимодействие возможно, то оно имеет место быть, а текущие ограничения связаны с недостаточной чувствительностью аналитических методов [18].

Немаловажным аспектом является описание конкретного типа белковой молекулы – протеоформы, которая может быть образована в результате альтернативного сплайсинга, несинонимичных замен отдельных нуклеотидов, различных ПТМ и ряда других более редких событий [19]. Суммарное количество протеоформ в организме человека окончательно не выяснено: в зависимости от используемого прогноза, учитывающего возможность одновременного возникновения нескольких aberrаций в белковом продукте одного гена, снизу оно ограничено миллионами уникальных белковых молекул, а сверху – может достигать астрономических величин [20, 21].

ЭКСПЕРИМЕНТ



Основные методы экспериментальной и вычислительной интерактомики

Важность учета индивидуальных протеоформ заключается в том, что наличие модификаций и изменений аминокислотного состава белка зачастую критично для выполняемых им функций, и профили белок-белковых взаимодействий для протеоформ, кодируемых одним геном, могут отличаться [22, 23].

Аберрации, приводящие к возникновению протеоформ и изменению их посттрансляционного статуса, связывают с развитием не только онкологических [24], но и других социально-значимых заболеваний [20]: в их числе аутоиммунные [25], метаболические [26] и нейродегенеративные [27] патологии.

Расшифровка протеома, реализуемая в рамках международного проекта «Протеом человека» [28], дала исследователям набор различных методов по выявлению протеоформ [29]. В связи с этим в последние годы появились работы по выявлению белковых взаимодействий для различного типа протеоформ [22, 30, 31]. В большинстве случаев предсказание ББВ для протеоформ осуществляется на основе изменений канонической структуры белка либо же выявлены экспериментальным путем в результате AP-MS экспериментов. Несмотря на развитие тенден-

ции перехода от гено-центричного к протеоформ-центричному интерактому, доминирование сведений о ББВ для гена сохраняется.

В 2006 году было проведено исследование интерактома человека, в котором было определено, что среди полученного количества ББВ (154 000–369 000) большая часть представлена ложноположительными данными и только 10% белковых взаимосвязей являются «истинными» [32]. В 2008 году количество ББВ человека было оценено уже в 650 000 [33]. На текущий момент это самая большая оценка существующего интерактома, однако есть мнения, что данная величина выше реально существующей в гено-центричном контексте [34, 35]. Тем не менее количество выявленных бинарных ББВ увеличивается с каждым годом [34]. В рамках 4-х крупнейших проектов [36–38] по созданию интерактомных карт человека было выявлено >15 000 бинарных взаимодействий, полученных разными экспериментальными методами. Суммарно было выявлено >90 000 ББВ, однако только 10% встречались более чем в одном проекте [39].

Противоречивость данных может быть как результатом технической ошибки, разной чувствительности аналитических методов, так и

наблюдаемыми эффектами биологической вариабельности, обусловленной анализом различных типов биоматериалов, динамичностью и гетерогенностью протеома. Решение данной дилеммы возможно с помощью методов биоинформатики, которая позволяет как предсказывать ББВ для дальнейшей проверки, так и верифицировать получаемые экспериментальные данные для получения новых предсказаний.

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ (ММО)

Развитие постгеномных технологий привело к накоплению данных в области протеомики, геномики и транскриптомики. В связи с этим актуальными становятся биоинформатические методы анализа данных, которые позволяют повысить уровень достоверности интерактома и выявить новые закономерности взаимодействия между белками [40].

Существенную группу среди применяемых алгоритмов поиска ББВ представляют методы машинного обучения. Алгоритмы машинного обучения, к которым относятся метод опорных векторов (support vector machine – SVM), Байесовские сети (Bayesian network), деревья принятия решения (decision tree), цепи Маркова (Markov chain) и др., позволяют оценить вероятность взаимодействия между белками по набору их признаков [41]. Обязательным условием работы таких алгоритмов является наличие обучающей выборки, на основе которой настраивают параметры работы методов.

Метод опорных векторов весьма популярен при исследовании интерактома [2]. Суть подхода заключается в минимизации рисков ложных идентификаций путем поиска гиперплоскости, которая наилучшим образом разделяет два класса объектов (взаимодействующие и не взаимодействующие белки) по набору переменных [42].

Другой подход – использование Байесовских сетей для предсказания ББВ [43]. Байесовские сети позволяют использовать данные различного типа (к примеру, описательные и числовые характеристики), выражая их в виде условных вероятностей. При этом можно компенсировать отсутствие каких-либо признаков у объектов, а также учитывать вес каждого признака в соответствии с его надежностью [44]. Вероятность взаимодействия двух белков рассчитывается как вероятность наблюдения определенного набора признаков для двух белков при условии, что они являются взаимодействующими. Преимущества использования данного метода были показаны в

работе Scott и Barton [45], где уровень ошибки был снижен на то время с 90 до 76%.

Деревья принятия решения используются как один из методов для предсказания ББВ, а также для подтверждения взаимодействий, полученных экспериментальными методами. Каждое дерево состоит из узлов – атрибутов, характеризующих некое свойство белка, ребер – «веток» дерева, на которых записаны значения атрибутов, и «листьев», которым приписаны значения целевой функции, т.е. взаимодействуют белки или нет [2]. Для классификации нового случая необходимо создать такое дерево на основе обучающей выборки, а затем «пройти от корня к листьям» для предсказания взаимодействия между белками.

В методе «случайного леса» (random forest) используется множество таких «деревьев». Итоговое решение принимается путем консенсусного голосования для совокупности всех «деревьев». В интерактоме человека метод «случайного леса» получил большую популярность по сравнению с методом дерева принятия решения [46].

Методы машинного обучения широко используются для разных подходов выявления ББВ: на основе аминокислотной/нуклеотидной последовательности [47], структурной информации [31], филогенетических данных [48], профилей экспрессии [49], анализа интерактомных сетей и т.д. [50]. Кроме того, они получили широкое использование в обработке экспериментальных данных [38] и при интеграции различных типов данных [51]. Модификации и вариации методов различны, в некоторых случаях используются сразу несколько подходов [52].

ПРЕДСКАЗАНИЕ БЕЛКОВЫХ ВЗАИМОДЕЙСТВИЙ НА ОСНОВЕ СТРУКТУРНОЙ ИНФОРМАЦИИ

Возрастающее внимание к изучению протеоформ способствовало открытию «второго дыхания» у методов выявления белковых взаимодействий на основе структурной информации. Данные методы основаны на наблюдении, что белки, имеющие сходную третичную структуру или консервативные участки последовательности, обладают и сходной функцией в клетке, а значит могут взаимодействовать с одинаковыми партнерами [53, 54].

Десятилетия применения рентгеновской кристаллографии позволили получить данные о структуре белковых комплексов. Эти сведения являются богатым источником для изучения принципов взаимодействий белков. На настоящий момент в PDB (Protein Data Bank) [55]

опубликовано >40 000 3D-структур белков человека и их комплексов, и использование принципа гомологии белковых последовательностей позволяет существенно расширить этот набор. В 2002 году Aloy и Russel [56] использовали данные о трехмерной структуре гомологичных белков, чтобы оценить вероятность взаимодействия предлагаемых пар. Учитывая известные 3D-структуры белковых комплексов и гомологичные последовательности для каждого взаимодействующего белка, подходы, основанные на аналогах трехмерных структур, позволяют оценить вероятность возможных взаимодействий между гомологами одного и того же вида [57] или протеоформ, к примеру, в случае одноаминокислотных замен [58] или ПТМ [59].

Предсказание белок-белковых взаимодействий с помощью сопоставления информации о структуре белка реализовано в программе PRISM [60]. При разработке этой системы использовали данные об известных трехмерных структурах, доступные в PDB (~5500 белков, идентичны <50%). Межбелковые контакты выделяли как совокупность атомов, расстояние между ван-дер-ваальсовыми радиусами которых было в пределах 5 Å. Всего была обработана 21 000 контактных площадок и выявлено 3800 характерных интерфейсных поверхностей, обеспечивающих контакты между субъединицами комплекса.

Домены являются наиболее консервативными участками белков и зачастую обеспечивают контактный интерфейс для взаимодействия между субъединицами комплекса. Метод ассоциации (association method, AS) — это один из первых методов предсказания домен-доменных взаимодействий, основанный на оценке количества доменных пар во взаимодействующих белках по отношению к встречаемости каждого домена в паре в независимой выборке [61]. Несколько позже Deng et al. предложили алгоритм, где взаимодействие доменов предсказывается путем оценки максимального правдоподобия данного ББВ [62]. Другими подходами к выявлению белковых взаимодействий на уровне доменов являются: метод линейного программирования (linear programming, LP) и его модификации для выявления бинарных и комплексных взаимодействий [63]; метод опорных векторов; метод моделирования вероятностных сетей (probabilistic network modeling) и др. [64].

На основе домен-доменных взаимодействий для сплайс-форм предсказываются партнеры в виде нативных вариантов белков [31], а также в виде сплайс-форм [65], белковые партнеры для протеоформ с различными ПТМ [66] или одноаминокислотными заменами [67].

МЕТОДЫ АВТОМАТИЧЕСКОГО АНАЛИЗА ТЕКСТОВ НАУЧНЫХ ПУБЛИКАЦИЙ (ТЕХТ-MINING)

Данные о взаимосвязанных белках (ББВ) могут быть получены на основе автоматического анализа текстов научных публикаций (text-mining) [68].

Методы быстрого и надежного извлечения данных из научной литературы — современная необходимость. Совершенствование экспериментальных постгеномных методов ведет к экспоненциальному росту количества статей, предоставляя, с одной стороны, «большие данные» для анализа, а с другой — глобальный вызов разработчикам биоинформатических систем для обработки данных. По состоянию на 2019 год наиболее широко используемая база данных биомедицинской литературы содержит ~30 млн ссылок, при этом для 5 млн из них полнотекстовая версия статьи представлена в открытом доступе в ресурсе *PubMedCentral* (данные на март 2019 года). Автоматический анализ накопленных в форме научных публикаций знаний в области взаимодействий белков во многом является основой для получения новых сведений о молекулярных механизмах возникновения заболеваний человека, а также методах диагностики, профилактики и лечения.

Большинство исследований в области извлечения информации о ББВ из текстов научных публикаций основаны на анализе резюме научных публикаций, поскольку это наиболее доступная часть статьи [68]. Для автоматического выявления в текстах данных о белок-белковых взаимодействиях используют созданные вручную правила поиска или заранее размеченный куратором набор текстов, на основе автоматического анализа которого создаются правила для последующего поиска взаимосвязей [69].

Наиболее простые алгоритмы, используемые для выявления ББВ, базируются на анализе частот совместного упоминания названий белков или генов в заданном фрагменте текста. Поиск такого рода взаимосвязей основан на гипотезе, что, вероятно, два белка взаимодействуют друг с другом, если они встречаются совместно в тексте одной [70] или близких по смыслу публикациях [71]. Очевидным недостатком этого подхода является существенное количество ложноположительных результатов. Снижение количества ложноположительных результатов может достигаться за счет использования дополнительных правил, но воспроизводимость полученных результатов при этом остается невысокой [72].

Следующее поколение методов анализа текстов для выявления ББВ основано на ма-

шинном обучении [69]. Показано, что использование таких методов позволяет существенно снизить количество ложноположительных результатов. Методы динамического программирования, сходные с методами выравнивания последовательностей, могут быть использованы для нахождения шаблонов в размеченных текстах и поиска ББВ [73].

Перспективным считается предсказание взаимодействующих белков с использованием ядерных методов (kernel-based approaches) — класса алгоритмов для распознавания образов [74]. Тем не менее недостатком использования таких методов является отсутствие информации о семантических взаимосвязях между объектами — названиями белков. Решением этой проблемы является использование для построения интерактомной карты триплетов (триплет состоит из названий двух белков и связанного с ними термина, характеризующего взаимосвязь), которые могут быть найдены с использованием байесовских методов машинного обучения.

Зачастую существенного улучшения результатов предсказания ББВ удается достичь за счет комбинирования различных методических подходов [75] или интеграции многомерных данных [76]. Для извлечения триплетов, характеризующих ББВ, в 2018 году в работе, выполненной Yu et al. [69], использовали методы анализа текстов на естественном языке (nature language processing). Взаимосвязь может быть описана произвольным термином или термином, присутствующим в заранее сформированном словаре. Полученные триплеты затем были разделены на «истинные» и «ложные», что позволило достичь существенного преимущества в точности предложенного алгоритма в сравнении с существующими решениями. Для планирования эксперимента или актуализации сведений в базах данных точность аналитического алгоритма предсказания ББВ имеет большее значение, чем высокие показатели воспроизводимости результата [69].

В настоящее время создано внушительное количество систем, позволяющих реконструировать интерактомную карту как в гено-центричном формате, так и находить взаимосвязи между протеоформами. Методы анализа текстов используют в основном для предсказания ББВ между фосфорилированными формами [77].

ОБРАБОТКА РЕЗУЛЬТАТОВ МАСС-СПЕКТРОМЕТРИЧЕСКИХ ЭКСПЕРИМЕНТОВ ПРИ AP-MS

Биоинформатические подходы востребованы и для анализа интерактомных данных, полу-

ченных масс-спектрометрически для комплексов после аффинного выделения. С одной стороны, особенностью таких экспериментов является сложность идентификации белков, заключающаяся в ограниченном количестве получаемых уникальных пептидов, по которым возможно безошибочно определить белок. Кроме того, возможна ошибочная идентификация пептида, возникающая в результате неверной интерпретации масс-спектра [15]. С другой стороны, велика вероятность получения ложноположительных результатов комплексных взаимодействий, обусловленных неспецифическими взаимодействиями из-за высокой концентрации белков в биоматериале, или же ложноотрицательных результатов в силу разрушения белкового комплекса при экстракции [5].

Для решения указанных выше проблем были разработаны различные биоинформатические подходы [78] и базы данных распространенных белков-контраминантов (CRAPome). В первую очередь используют статистические пороги (t -критерий, p -value), полученные на основе масс-спектров для белка-«наживки» и для контрольного образца [79].

Вопрос о статистической значимости оценки белковых взаимодействий на основе масс-спектрометрических данных был решен в рамках программ SAINT (Significance Analysis of INTeractome) [80] и ComPASS (Comparative Proteomics Analysis Software Suite) [81]. В качестве критерия для исключения ложноположительных результатов в программах используется распространенная в протеомике относительная количественная оценка содержания белка в пробе — «счетчик спектров» (spectral count). Она определяется путем подсчета количества масс-спектров фрагментации, относящихся к белку пептидов. Большие значения счетчика характеризуют более высокий уровень содержания белка в исследуемом образце. Близкие по идеологии программы незначительно отличаются эффективностью в пользу SAINT [15].

С помощью SAINT можно не только подтверждать специфические бинарные взаимодействия между белками, но и выявлять новые белковые взаимодействия. Например, при анализе результатов масс-спектрометрических экспериментов для серин/треонин фосфатазы PP5 было установлено ранее неизвестное взаимодействие со стресс-индуцированным фосфопротеином 1 (STIP 1) [82].

Другой метод анализа достоверности ББВ, выявленных с помощью AP-MS — MiST (Mass spectrometry interaction STatistics), основан на интегрированной оценке по нескольким параметрам: содержанию белка (интенсивности пи-

ка), воспроизводимости (повторяемости белкового комплекса) и специфичности (уникальности белка-«добычи» по отношению к другим белкам в комплексе) взаимодействия [83].

Несколько идеологически отличается алгоритм SFINX (Straightforward Filtering INdex) [84], в рамках которого мета-данные анализируются автоматически — нет указания белка-«наживки» в результатах AP-MS, используется не счетчик спектров, а оценка интенсивности сигналов, относящихся к пептидам. Данный метод самостоятельно определяет пороговую величину достоверности в каждом случае на основе биномиального распределения.

Для оценки ложноположительных результатов, возникающих в первую очередь из-за контаминации, были разработаны методы CRAPome [85] и Decontaminator [86], в основе которых лежит оценка частоты встречаемости в различных экспериментах.

В качестве входных данных для методов вычислительной интерактомики используются результаты масс-спектрометрических экспериментов, выполненных после аффинного выделения белкового комплекса. Большое количество таких экспериментов депонируется в централизованные протеомные репозитории. Для многих экспериментов при этом отсутствуют мета-данные, например, в репозитории GPMdb [87], которые бы указывали, что результаты идентификации белков были получены в рамках исследования интерактома. В работе Zhang et al. [88] был предложен метод, позволяющий в автоматическом режиме извлечь из репозитория данные, которые с наибольшей вероятностью относятся к анализу белковых взаимодействий. В основе метода лежит идея о том, что частота совместно встречающихся белков в результатах AP-MS будет высока для обоих белковых партнеров.

Методы панорамной масс-спектрометрии позволяют выявлять протеоформы без дополнительных трудозатрат — путем использования расширенных библиотек и настроек при анализе масс-спектра. На сегодняшний момент экспериментальная база данных ББВ IntACT [89] является единственным ресурсом, поддерживающим публикацию идентифицированных сплайс-форм и их взаимодействий. Необработанные масс-спектры интерактомных экспериментов депонируются непосредственно в протеомные репозитории PRIDE [90], GPMdb [87] и PeptideAtlas [91]. В случае top-down подхода в масс-спектрометрии был разработан биоинформатический способ анализа белковых комплексов, образованных в том числе различными типами протеоформ [30].

БАЗЫ ДАННЫХ

Многие биоинформатические алгоритмы используют опубликованные сведения о ББВ для настройки или определения ложных результатов. Полученные результаты публикуются в виде баз данных. Практически каждый метод имеет свой ресурс, однако в силу того что есть базы данных, агрегаторы, и количество новых интерактомных ресурсов возрастает с каждым годом, некоторые ресурсы перестают поддерживаться. Данные могут обновляться постоянно или с некой периодичностью (версией ресурса), а могут быть статичными.

Ниже приведен список актуальных и наиболее распространенных баз данных ББВ человека (таблица). Можно выделить две группы интерактомных ресурсов — специализирующихся на конкретном организме (например, содержащих данные только для человека) и на конкретном типе данных (предсказанные ББВ, экспериментально подтвержденные ББВ, комбинированные).

Суммарно для человека имеются данные о нескольких миллионах белковых взаимодействий. Многие ресурсы интегрируют различные дополнительные сведения — о клеточной локализации [92], уровне экспрессии [93], свойствах гомологичных белков у других организмов [94] и т.д. Отдельно стоит упоминать базы данных, описывающие межорганизменные взаимодействия (в частности, человека с вирусами) [95]. Тем не менее данные ресурсов совпадают не более чем на 10–20% [39], в связи с чем возникает необходимость разработки новых алгоритмов предсказания и анализа ББВ.

ИНТЕРАКТОМНЫЕ СЕТИ

Использование биоинформатических методов позволяет визуализировать сети (или карты) ББВ. Визуализация дает представление о роли каждого белка в биологических процессах, а также позволяет выявлять новые белковые взаимодействия на основе построенных карт. Примером служит работа Kovács et al. [96], где на основе структурных и эволюционных данных показали, что белки, взаимодействующие через одного партнера, на самом деле в ряде случаев взаимодействуют напрямую.

Перед алгоритмами визуализации интерактомных карт стоит несколько задач: 1) совместимость с различными форматами сведений о ББВ, 2) воспринимаемый формат визуализации с возможностью масштабирования и 3) анализ сетей различными методами (кластеризация, анализ графа и т.д.), включая возможность подклю-

Интерактомные ресурсы, содержащие сведения о ББВ человека

База данных	Ссылка	По типу данных	Количество ББВ для человека	Обновляемость
Специализированная для человека				
BioPlex	http://bioplex.hms.harvard.edu/	экспериментальная	56 000	периодически обновляемая
HAPPI	http://discovery.informatics.uab.edu/HAPPI/	комбинированная	2 922 202	статичная
HPID	http://wilab.inha.ac.kr/hpid/	комбинированная	9000	периодически обновляемая
HPRD	http://www.hprd.org/	комбинированная	41 327	статичная
PepCyber: P-Pep	http://www.pepcyber.org/PPEP/	предсказательная	11 269	статичная
TissueNet	http://netbio.bgu.ac.il/tissuenet/	предсказательная	243 706	периодически обновляемая
UniHI	http://www.unihi.org/	предсказательная	573 995	периодически обновляемая
Содержащая данные для различных организмов				
APID	http://cicblade.dep.usal.es:8080/APID/init.action	комбинированная	385 438	постоянно обновляемая
BioGRID	https://thebiogrid.org/	комбинированная	496 730	постоянно обновляемая
ComPPI	http://ComPPI.LinkGroup.hu	комбинированная	1 311 184	статичная
DIP	https://dip.mbi.ucla.edu/dip/	комбинированная	9141	периодически обновляемая
IID	http://iid.ophid.utoronto.ca/	комбинированная	975 877	периодически обновляемая
IntACT	https://www.ebi.ac.uk/intact/	экспериментальная	~455 000	постоянно обновляемая
I2D	http://ophid.utoronto.ca/ophidv2.204/	комбинированная	296 008	периодически обновляемая
STRING	https://string-db.org/	комбинированная	~4 000 000	постоянно обновляемая
MINT	https://mint.bio.uniroma2.it/	комбинированная	48 352	постоянно обновляемая
3did	http://3did.irbbarcelona.org/	комбинированная	~6000	статичная

ния дополнительной информации из других ресурсов. В качестве программ построения интерактомных карт наибольшую популярность получили Cytoscape, NAViGaTOR, VisAnt, ProViz и др. [97]. Первая интерактомная карта человека появилась в 2004 году [98]. Она была построена по гомологии с данными о белковых взаимодействиях модельных организмов: *Saccharomyces cerevisiae*, *Drosophila melanogaster* и *Caenorhabditis*

elegans. На основе анализа сведений о белковых взаимодействиях модельных организмов были выделены консервативные участки интерактома. Затем появились публикации по построению карты ББВ для человека на основе результатов, полученных высокопроизводительным методом двугибридной дрожжевой системы [99]. В 2007 году была сконструирована первая карта на основе данных метода AP-MS [100]. Сеть состоит

из вершин (белков), связанных ребрами (белок-белковые взаимодействия). По мере накопления сведений об интерактоме возник интерес к исследованию структур сетей.

В зависимости от расстояния между любыми двумя узлами в сети выделяют три структуры интерактомных сетей: «безразмерные» (scale-free network), «маленького мира» (small-world network) и «рандомные» (random network). В сетях «маленького мира» расстояния между узлами небольшие, в отличие от «безразмерных». «Рандомная» структура интерактомных сетей была предложена первой, но оказалась не самой удачной моделью, поскольку не включала в себя узловые белки, или белки-«центры» («hub») [101]. Особенностью узловых белков является повышенное количество взаимодействий по отношению к другим белкам, участвующим в сети.

Ключевая роль узловых белков проявляется в изменении биологического состояния [102]. Взаимодействия одного или нескольких белков-«центров» с компонентами белковой сети бывает достаточно для переключения клеточного процесса. Основная часть белков, участвующих в динамических процессах, экспрессируется независимо от узловых белков, образование которых происходит в течение всего процесса работы комплекса (статистические узловые белки) или обусловлено определенными условиями (динамические узловые белки) [103]. Динамичная модульная структура сети присуща различным видам биологических процессов, включая экспрессию генов, посттрансляционную модификацию и субклеточную организацию.

Анализ сетей ББВ вычислительными и статистическими методами позволяет получать информацию об организации биологических систем [104]. Все больше внимания уделяется исследованию интерактомных карт в медицинском аспекте. Центральная гипотеза, лежащая в основе связи карт ББВ и фенотипа заболеваний, заключается в нарушении локальной или глобальной структуры сетей, приводящей к патологическим изменениям [102].

С помощью исследований интерактомных сетей возможно выявление новых аспектов клеточных процессов: не только при заболеваниях, включающих в том числе наркологические зависимости [105] и взаимодействия с вирусами [106], но и при поиске новых лекарств [107], в том числе антибиотиков [108]. Непосредственно

в выявлении новых механизмов развития заболеваний ключевым фактором становится изменение взаимодействий aberrantных протеоформ [109]. Именно анализ сетей ББВ способствовал развитию изучения протеоформ, которые уже рассматриваются и в клиническом аспекте [110].

Несмотря на активно развивающиеся подходы геномного и постгеномного анализа современная интерактомика невозможна без биоинформатических методов. Высокопроизводительные технологии генерируют огромные массивы молекулярных данных, но расшифровка этих данных подразумевает анализ их взаимодействий между собой. Построение интерактомных карт, равно как и их трактовка, возможны только с помощью методов биоинформатики. Переход к прецизионной медицине требует выявлять особенности «нормы» для каждого индивидуума. При сравнительно небольшом количестве белок-кодирующих генов гетерогенность на уровне протеома у человека значительна.

Для человека, согласно данным UniProt [17], вариатив протеоформ, кодируемых с одного гена, наибольший среди всех организмов, однако количество «нейтральных» и «аномальных» белковых молекул только предстоит узнать. На фоне перехода к прецизионной медицине выявление особенностей «нормы» для каждого индивидуума — это вызов как для протеомики (детекция всех транслируемых протеоформ, в том числе органоспецифичных), так и для интерактомики (выявление критических протеоформ, характерных для развития патологических состояний). В данных вопросах биоинформатика является универсальным инструментом анализа на всех уровнях получения и интерпретации результатов, позволяя, в том числе, компенсировать ограничения экспериментальных методов.

Финансирование. Работа выполнена при поддержке Российского фонда фундаментальных исследований по теме «Построение интерактомной карты человека» (грант № 18-34-00879).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Соблюдение этических норм. Настоящая статья не содержит описания выполненных авторами исследований с участием людей или использованием животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Braun, P., and Gingras, A.-C. (2012) History of protein-protein interactions: from egg-white to complex networks, *Proteomics*, **12**, 1478–1498, doi: 10.1002/pmic.201100563.
2. Qi, Y., Bar-Joseph, Z., and Klein-Seetharaman, J. (2006) Evaluation of different biological data and computational classification methods for use in protein interaction prediction, *Proteins*, **63**, 490–500, doi: 10.1002/prot.20865.

3. Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I., and Stagljari, I. (2015) Fundamentals of protein interaction network mapping, *Mol. Syst. Biol.*, **11**, 848, doi: 10.15252/msb.20156351.
4. Brückner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009) Yeast two-hybrid, a powerful tool for systems biology, *Int. J. Mol. Sci.*, **10**, 2763–2788, doi: 10.3390/ijms10062763.
5. Hakhverdyan, Z., Domanski, M., Hough, L.E., Oroskar, A.A., Oroskar, A.R., Keegan, S., and LaCava, J. (2015) Rapid, optimized interactomic screening, *Nat. Methods*, **12**, 553–560, doi: 10.1038/nmeth.3395.
6. Fields, S., and Song, O. (1989) A novel genetic system to detect protein–protein interactions, *Nature*, **340**, 245–246, doi: 10.1038/340245a0.
7. Riegel, E., Heimbucher, T., Höfer, T., and Czerny, T. (2017) A sensitive, semi-quantitative mammalian two-hybrid assay, *BioTechniques*, **62**, 206–214, doi: 10.2144/000114544.
8. Gaudinier, A., Tang, M., Bågman, A.-M., and Brady, S.M. (2017) Identification of protein–DNA interactions using enhanced yeast one-hybrid assays and a semiautomated approach, *Methods Mol. Biol.*, **1610**, 187–215, doi: 10.1007/978-1-4939-7003-2_13.
9. Glass, F., and Takenaka, M. (2018) The yeast three-hybrid system for protein interactions, *Methods Mol. Biol.*, **1794**, 195–205, doi: 10.1007/978-1-4939-7871-7_12.
10. Dunham, W.H., Mullin, M., and Gingras, A.-C. (2012) Affinity-purification coupled to mass spectrometry: basic principles and strategies, *Proteomics*, **12**, 1576–1590, doi: 10.1002/pmic.201100523.
11. Morris, J.H., Knudsen, G.M., Verschuere, E., Johnson, J.R., Cimermancic, P., Greninger, A.L., and Pico, A.R. (2014) Affinity purification–mass spectrometry and network analysis to understand protein–protein interactions, *Nat. Prot.*, **9**, 2539–2554, doi: 10.1038/nprot.2014.164.
12. Serebriiskii, I.G., and Golemis, E.A. (2001) Two-hybrid system and false positives. Approaches to detection and elimination, *Methods Mol. Biol.*, **177**, 123–134, doi: 10.1385/1-59259-210-4:123.
13. Florinskaya, A., Ershov, P., Mezentsev, Y., Kaluzhskiy, L., Yablokov, E., Medvedev, A., and Ivanov, A. (2018) SPR biosensors in direct molecular fishing: implications for protein interactomics, *Sensors*, **18**, 1616, doi: 10.3390/s18051616.
14. Иванов А.С., Згода В.Г., Арчаков А.И. (2011) Технологии белковой интерактомики, *Биоорганическая химия*, **37**, 8–21, doi: 10.1134/s1068162011010092.
15. Nesvizhskii, A.I. (2012) Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments, *Proteomics*, **12**, 1639–1655, doi: 10.1002/pmic.201100537.
16. Teng, Z., Guo, M., Liu, X., Tian, Z., and Che, K. (2017) Revealing protein functions based on relationships of interacting proteins and GO terms, *J. Biomed. Semantics*, **8**, 27, doi: 10.1186/s13326-017-0139-8.
17. The UniProt Consortium (2008) The Universal Protein Resource (UniProt), *Nucleic Acids Res.*, **36**, Suppl. 1, D190–D195, doi: 10.1093/nar/gkm895.
18. Medvedev, A., Kopylov, A., Buneeva, O., Zgoda, V., and Archakov, A. (2012) Affinity-based proteomic profiling: problems and achievements, *Proteomics*, **12**, 621–637, doi: 10.1002/pmic.201100373.
19. Kiseleva, O., Poverennaya, E., Shargunov, A., and Lisitsa, A. (2017) Proteomic cinderella: customized analysis of bulky MS/MS data in one night, *J. Bioinform. Comput. Biol.*, **16**, doi: 10.1142/S021972001740011X.
20. Aebersold, R., Agar, J.N., Amster, I.J., Baker, M.S., Bertozzi, C.R., et al. (2018) How many human proteoforms are there? *Nat. Chem. Biol.*, **14**, 206–214, doi: 10.1038/nchembio.2576.
21. Ponomarenko, E.A., Poverennaya, E.V., Ilgisonis, E.V., Pyatnitskiy, M.A., Kopylov, A.T., Zgoda, V.G., and Archakov, A.I. (2016) The size of the human proteome: the width and depth, *Int. J. Anal. Chem.*, doi: 10.1155/2016/7436849.
22. Uversky, V.N. (2016) p53 proteoforms and intrinsic disorder: an illustration of the protein structure–function continuum concept, *Int. J. Mol. Sci.*, **17**, doi: 10.3390/ijms17111874.
23. Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013) Function of alternative splicing, *Gene*, **514**, 1–30, doi: 10.1016/j.gene.2012.07.083.
24. Pyatnitskiy, M., Karpov, D., Poverennaya, E., Lisitsa, A., and Moshkovskii, S. (2015) Bringing down cancer aircraft: searching for essential hypomutated proteins in skin melanoma, *PLoS One*, **10**, e0142819, doi: 10.1371/journal.pone.0142819.
25. Plymire, D.A., Wing, C.E., Robinson, D.E., and Patrie, S.M. (2017) Continuous elution proteoform identification of myelin basic protein by superficially porous reversed-phase liquid chromatography and fourier transform mass spectrometry, *Anal. Chem.*, **89**, 12030–12038, doi: 10.1021/acs.analchem.7b02426.
26. Nedelkov, D. (2017) Mass spectrometric studies of apolipoprotein proteoforms and their role in lipid metabolism and type 2 diabetes, *Proteomes*, **5**, 27, doi: 10.3390/proteomes5040027.
27. Lacovich, V., Espindola, S.L., Alloatti, M., Pozo Devoto, V., Cromberg, L.E., Carná, M.E., Giancarlo, F., Gallo, J.M., Bruno, L., Stokin, J.B., Avale, M.E., and Falzone, T.L. (2017) Tau isoforms imbalance impairs the axonal transport of the amyloid precursor protein in human neurons, *J. Neurosci.*, **37**, 58–69, doi: 10.1523/JNEUROSCI.2305-16.2017.
28. Пономаренко Е.А., Поверенная Е.В., Ильгисонис Е.В., Копылов А.Т., Згода В.Г., Лисица А.В., Арчаков А.И. (2017) Перспективы исследования протеома человека, *Вестник Российской академии наук*, **87**, 599–604, doi: 10.1134/S1019331617040049.
29. Киселева О.И., Лисица А.В., Поверенная Е.В. (2018) Протеоформы: методы исследования и клинические перспективы, *Молекулярная биология*, **52**, 394–410, doi: 10.7868/S0026898418030047.
30. Skinner, O.S., Havugimana, P.C., Haverland, N.A., Fornelli, L., Early, B.P., Greer, J.B., and Kelleher, N.L. (2016) An informatic framework for decoding protein complexes by top-down mass spectrometry, *Nat. Methods*, **13**, 237–240, doi: 10.1038/nmeth.3731.
31. Ghadie, M.A., Lambourne, L., Vidal, M., and Xia, Y. (2017) Domain-based prediction of the human isoform interactome provides insights into the functional impact of alternative splicing, *PLoS Comput. Biol.*, **13**, e1005717, doi: 10.1371/journal.pcbi.1005717.
32. Hart, G.T., Ramani, A.K., and Marcotte, E.M. (2006) How complete are current yeast and human protein–interaction networks? *Genome Biol.*, **7**, 120, doi: 10.1186/gb-2006-7-11-120.
33. Stumpf, M.P.H., Thorne, T., de Silva, E., Stewart, R., An, H.J., Lappe, M., and Wiuf, C. (2008) Estimating the size of the human interactome, *Proc. Natl. Acad. Sci. USA*, **105**, 6959–6964, doi: 10.1073/pnas.0708078105.
34. Kotlyar, M., Pastrello, C., Malik, Z., and Jurisica, I. (2019) IID 2018 update: context-specific physical protein–protein interactions in human, model organisms and domesticated species, *Nucleic Acids Res.*, **47**, D581–D589, doi: 10.1093/nar/gky1037.

35. Vidal, M. (2016) How much of the human protein interactome remains to be mapped? *Sci. Signal.*, **9**, eg7, doi: 10.1126/scisignal.aaf6030.
36. Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., and Emili, A. (2015) Panorama of ancient metazoan macromolecular complexes, *Nature*, **525**, 339–344, doi: 10.1038/nature14877.
37. Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., and Mann, M. (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances, *Cell*, **163**, 712–723, doi: 10.1016/j.cell.2015.09.053.
38. Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., and Harper, J.W. (2017) Architecture of the human interactome defines protein communities and disease networks, *Nature*, **545**, 505–509, doi: 10.1038/nature22366.
39. Luck, K., Sheynkman, G.M., Zhang, I., and Vidal, M. (2017) Proteome-scale human interactomics, *Trends Biochem. Sci.*, **42**, 342–354, doi: 10.1016/j.tibs.2017.02.006.
40. Kotlyar, M., Rossos, A.E.M., and Jurisica, I. (2017) Prediction of protein–protein interactions, *Curr. Protoc. Bioinformatics*, **60**, 8.2.1–8.2.14, doi: 10.1002/cpbi.38.
41. Zhang, M., Su, Q., Lu, Y., Zhao, M., and Niu, B. (2017) Application of machine learning approaches for protein–protein interactions prediction, *Med. Chem.*, **13**, 506–514, doi: 10.2174/1573406413666170522150940.
42. Horvatovich, P., Lundberg, E.K., Chen, Y.-J., Sung, T.-Y., He, F., et al. (2015) Quest for missing proteins: update 2015 on chromosome-centric human proteome project, *J. Proteome Res.*, **14**, 3415–3431, doi: 10.1021/pr5013009.
43. Bradford, J.R., Needham, C.J., Bulpitt, A.J., and Westhead, D.R. (2006) Insights into protein–protein interfaces using a Bayesian network prediction method, *J. Mol. Biol.*, **362**, 365–386, doi: 10.1016/j.jmb.2006.07.028.
44. Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., and Gerstein, M. (2003) A Bayesian networks approach for predicting protein–protein interactions from genomic data, *Science*, **302**, 449–53, doi: 10.1126/science.1087361.
45. Scott, M.S., and Barton, G.J. (2007) Probabilistic prediction and ranking of human protein–protein interactions., *BMC bioinformatics*, **8**, 239, doi: 10.1186/1471-2105-8-239.
46. Chatterjee, P., Basu, S., Kundu, M., Nasipuri, M., and Plewczynski, D. (2011) PPI_SVM: prediction of protein–protein interactions using machine learning, domain–domain affinities and frequency tables, *Cell. Mol. Biol. Lett.*, **16**, 264–278, doi: 10.2478/s11658-011-0008-x.
47. Guo, Y., Sheng, Q., Li, J., Ye, F., Samuels, D.C., and Shyr, Y. (2013) Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data, *PLoS One*, **8**, e71462, doi: 10.1371/journal.pone.0071462.
48. Zahiri, J., Bozorgmehr, J., and Masoudi-Nejad, A. (2013) Computational prediction of protein–protein interaction networks: algorithms and resources, *Curr. Genomics*, **14**, 397–414, doi: 10.2174/1389202911314060004.
49. Bartoli, L., Martelli, P.L., Rossi, I., Fariselli, P., and Casadio, R. (2010) The prediction of protein–protein interacting sites in genome-wide protein interaction networks: the test case of the human cell cycle, *Curr. Prot. Pept. Sci.*, **11**, 601–608, doi: 10.2174/138920310794109157.
50. McDowall, M.D., Scott, M.S., and Barton, G.J. (2009) PIPs: human protein–protein interaction prediction database, *Nucleic Acids Res.*, **37**, D651–D656, doi: 10.1093/nar/gkn870.
51. Garzón, J.I., Deng, L., Murray, D., Shapira, S., Petrey, D., and Honig, B. (2016) A computational interactome and functional annotation for the human proteome, *eLife*, **5**, doi: 10.7554/eLife.18715.
52. Dick, K., and Green, J.R. (2018) Reciprocal perspective for Improved protein–protein interaction prediction, *Sci. Rep.*, **8**, 11694, doi: 10.1038/s41598-018-30044-1.
53. Gromiha, M.M., Yugandhar, K., and Jemimah, S. (2017) Protein–protein interactions: scoring schemes and binding affinity, *Curr. Opin. Struct. Biol.*, **44**, 31–38, doi: 10.1016/j.sbi.2016.10.016.
54. Gemovic, B., Sumonja, N., Davidovic, R., Perovic, V., and Veljkovic, N. (2018) Mapping of protein–protein interactions: web-based resources for revealing interactomes, *Curr. Med. Chem.*, **26**, 3890–3910, doi: 10.2174/0929867325666180214113704.
55. Velankar, S., and Kleywegt, G.J. (2011) The Protein Data Bank in Europe (PDBe): bringing structure to biology, *Acta Crystallogr. D*, **67**, 324–330, doi: 10.1107/S090744491004117X.
56. Aloy, P., and Russell, R.B. (2002) The third dimension for protein interactions and complexes, *Trends Biochem. Sci.*, **27**, 633–638.
57. Fang, Y., Sun, M., Dai, G., and Ragain, K. (2016) The intrinsic geometric structure of protein–protein interaction networks for protein interaction prediction, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **13**, 76–85, doi: 10.1109/TCBB.2015.2456876.
58. Tuncbag, N., Keskin, O., Nussinov, R., and Gursoy, A. (2017) Prediction of protein interactions by structural matching: prediction of PPI networks and the effects of mutations on PPIs that combines sequence and structural information, *Methods Mol. Biol.*, **1558**, 255–270, doi: 10.1007/978-1-4939-6783-4_12.
59. Su, M.-G., Weng, J.T.-Y., Hsu, J.B.-K., Huang, K.-Y., Chi, Y.-H., and Lee, T.-Y. (2017) Investigation and identification of functional post-translational modification sites associated with drug binding and protein–protein interactions, *BMC Systems Biology*, **11**, 132, doi: 10.1186/s12918-017-0506-1.
60. Keskin, O., Nussinov, R., and Gursoy, A. (2008) PRISM: protein–protein interaction prediction by structural matching, *Methods Mol. Biol.*, **484**, 505–521, doi: 10.1007/978-1-59745-398-1_30.
61. Sprinzak, E., and Margalit, H. (2001) Correlated sequence-signatures as markers of protein–protein interaction, *J. Mol. Biol.*, **311**, 681–692, doi: 10.1006/jmbi.2001.4920.
62. Deng, M., Mehta, S., Sun, F., and Chen, T. (2002) Inferring domain–domain interactions from protein–protein interactions, *Genome Res.*, **12**, 1540–1548, doi: 10.1101/gr.153002.
63. Hayashida, M., Ueda, N., and Akutsu, T. (2004) A simple method for inferring strengths of protein–protein interactions, *Genome Inform.*, **15**, 56–68.
64. Raghavachari, B., Tasneem, A., Przytycka, T.M., and Jothi, R. (2008) DOMINE: a database of protein domain interactions, *Nucleic Acids Res.*, **36**, D656–D661, doi: 10.1093/nar/gkm761.
65. Tseng, Y.-T., Li, W., Chen, C.-H., Zhang, S., Chen, J.J., Zhou, X., and Liu, C.-C. (2015) IIDB: a database for isoform–isoform interactions and isoform network modules, *BMC Genomics*, **16**, S10, doi: 10.1186/1471-2164-16-S2-S10.
66. Tay, A.P., Pang, C.N.I., Winter, D.L., and Wilkins, M.R. (2017) PTMOacle: a cytoscape app for covisualizing and coanalyzing post-translational modifications in protein interaction networks, *J. Proteome Res.*, **16**, 1988–2003, doi: 10.1021/acs.jproteome.6b01052.
67. Ivanov, A.A., Revenaugh, B., Rusnak, L., Gonzalez-Pecchi, V., Mo, X., Johns, M.A., and Fu, H. (2018) The

- OncoPPI Portal: an integrative resource to explore and prioritize protein–protein interactions for cancer target discovery, *Bioinformatics*, **34**, 1183–1191, doi: 10.1093/bioinformatics/btx743.
68. Skusa, A., Rüegg, A., and Köhler, J. (2005) Extraction of biological interaction networks from scientific literature, *Briefings Bioinform.*, **6**, 263–276.
 69. Yu, K., Lung, P.-Y., Zhao, T., Zhao, P., Tseng, Y.-Y., and Zhang, J. (2018) Automatic extraction of protein–protein interactions using grammatical relationship graph, *BMC Med. Inform. Decis. Mak.*, **18**, 42, doi: 10.1186/s12911-018-0628-4.
 70. Stapley, B.J., and Benoit, G. (2000) Biobibliometrics: information retrieval and visualization from co-occurrences of gene names in Medline abstracts, *Pac. Symp. Biocomput.*, **2000**, 529–540.
 71. Пономаренко Е.А., Лисица А.В., Ильгисонис Е.В., Арчаков А.И. (2010) Создание семантических сетей белков с использованием Pubmed/Medline, *Молекулярная биология*, **44**, 152–161, doi: 10.1134/S0026893310010176.
 72. Lee, J., Kim, S., Lee, S., Lee, K., and Kang, J. (2013) On the efficacy of per-relation basis performance evaluation for PPI extraction and a high-precision rule-based approach, *BMC Med. Inform. Decis. Mak.*, **13** Suppl 1, S7, doi: 10.1186/1472-6947-13-S1-S7.
 73. Huang, M., Zhu, X., Hao, Y., Payan, D.G., Qu, K., and Li, M. (2004) Discovering patterns to extract protein–protein interactions from full texts, *Bioinformatics*, **20**, 3604–3612, doi: 10.1093/bioinformatics/bth451.
 74. Murugesan, G., Abdulkadhar, S., and Natarajan, J. (2017) Distributed smoothed tree kernel for protein–protein interaction extraction from the biomedical literature, *PLoS One*, **12**, e0187379, doi: 10.1371/journal.pone.0187379.
 75. Niu, Y., and Wang, Y. (2015) Protein–protein interaction identification using a hybrid model, *Artif. Intell. Med.*, **64**, 185–193, doi: 10.1016/j.artmed.2015.05.003.
 76. Chang, J.-W., Zhou, Y.-Q., Ul Qamar, M.T., Chen, L.-L., and Ding, Y.-D. (2016) Prediction of protein–protein interactions by evidence combining methods, *Int. J. Mol. Sci.*, **17**, E1946, doi: 10.3390/ijms17111946.
 77. Wang, Q., Ross, K.E., Huang, H., Ren, J., Li, G., Vijay-Shanker, K., and Arighi, C.N. (2017) Analysis of protein phosphorylation and its functional impact on protein–protein interactions via text mining of the scientific literature, *Methods Mol. Biol.*, **1558**, 213–232, doi: 10.1007/978-1-4939-6783-4_10.
 78. Armean, I.M., Lilley, K.S., and Trotter, M.W.B. (2013) Popular computational methods to assess multiprotein complexes derived from label-free affinity purification and mass spectrometry (AP-MS) experiments, *Mol. Cell. Proteomics*, **12**, 1–13, doi: 10.1074/mcp.R112.019554.
 79. Fernández, E., Collins, M.O., Uren, R.T., Kopanitsa, M.V., Komiyama, N.H., Croning, M. D.R., and Grant, S.G.N. (2009) Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins, *Mol. Syst. Biol.*, **5**, 269, doi: 10.1038/msb.2009.27.
 80. Choi, H., Larsen, B., Lin, Z.-Y., Breikreutz, A., Mellacheruvu, D., Fermin, D., and Nesvizhskii, A.I. (2011) SAINT: probabilistic scoring of affinity purification-mass spectrometry data, *Nat. Methods*, **8**, 70–73, doi: 10.1038/nmeth.1541.
 81. Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009) Defining the human deubiquitinating enzyme interaction landscape, *Cell*, **138**, 389–403, doi: 10.1016/j.cell.2009.04.042.
 82. Skarra, D.V., Goudreault, M., Choi, H., Mullin, M., Nesvizhskii, A.I., Gingras, A.-C., and Honkanen, R.E. (2011) Label-free quantitative proteomics and SAINT analysis enable interactome mapping for the human Ser/Thr protein phosphatase 5, *Proteomics*, **11**, 1508–1516, doi: 10.1002/pmic.201000770.
 83. Verschueren, E., Von Dollen, J., Cimermancic, P., Gulbahce, N., Sali, A., and Krogan, N.J. (2015) Scoring large-scale affinity purification mass spectrometry datasets with MiST, *Curr. Protoc. Bioinformatics*, **49**, 8.19.1–8.19.16, doi: 10.1002/0471250953.bi0819s49.
 84. Titeca, K., Meysman, P., Gevaert, K., Tavernier, J., Laukens, K., Martens, L., and Eyckerman, S. (2016) SFINX: straightforward filtering index for affinity purification–mass spectrometry data analysis, *J. Proteome Res.*, **15**, 332–338, doi: 10.1021/acs.jproteome.5b00666.
 85. Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.-P., St-Denis, N.A., Li, T., and Nesvizhskii, A.I. (2013) The CRAPome: a contaminant repository for affinity purification-mass spectrometry data, *Nat. Methods*, **10**, 730–736, doi: 10.1038/nmeth.2557.
 86. Lavallée-Adam, M., Cloutier, P., Coulombe, B., and Blanchette, M. (2011) Modeling contaminants in AP-MS/MS experiments, *J. Proteome Res.*, **10**, 886–895, doi: 10.1021/pr100795z.
 87. Craig, R., Cortens, J.P., and Beavis, R.C. (2004) Open source system for analyzing, validating, and storing protein identification data, *J. Proteome Res.*, **3**, 1234–1242, doi: 10.1021/pr049882h.
 88. Zhang, C., Rogalski, J.C., Evans, D.M., Klockenbusch, C., Beavis, R.C., and Kast, J. (2011) *In silico* protein interaction analysis using the global Proteome Machine Database Research articles, *J. Proteome Res.*, **10**, 656–668.
 89. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., and Hermjakob, H. (2012) The IntAct molecular interaction database in 2012, *Nucleic Acids Res.*, **40**, D841–D846, doi: 10.1093/nar/gkr1088.
 90. Jones, P., Côté, R.G., Cho, S.Y., Klie, S., Martens, L., Quinn, A.F., and Hermjakob, H. (2008) PRIDE: new developments and new datasets, *Nucleic Acids Res.*, **36**, D878–D883, doi: 10.1093/nar/gkm1021.
 91. Deutsch, E.W. (2010) The PeptideAtlas Project, *Methods Mol. Biol.*, **604**, 285–296, doi: 10.1007/978-1-60761-444-9_19.
 92. V̄eres, D.V., Gyurkó, D.M., Thaler, B., Szalay, K.Z., Fazekas, D., Korcsmáros, T., and Csermely, P. (2015) ComPPI: a cellular compartment-specific database for protein–protein interaction network analysis, *Nucleic Acids Res.*, **43**, D485–D493, doi: 10.1093/nar/gku1007.
 93. Basha, O., Barshir, R., Sharon, M., Lerman, E., Kirson, B.F., Hekselman, I., and Yezer-Lotem, E. (2017) The TissueNet v.2 database: a quantitative view of protein–protein interactions across human tissues, *Nucleic Acids Res.*, **45**, D427–D431, doi: 10.1093/nar/gkw1088.
 94. Brown, K.R., and Jurisica, I. (2005) Online Predicted Human Interaction Database, *Bioinformatics*, **21**, 2076–2082, doi: 10.1093/bioinformatics/bti273.
 95. Rozenblatt-Rosen, O., Deo, R.C., Padi, M., Adelmant, G., Calderwood, M.A., Rolland, T., and Vidal, M. (2012) Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins, *Nature*, **487**, 491–495, doi: 10.1038/nature11288.
 96. Kovács, I.A., Luck, K., Spirohn, K., Wang, Y., Pollis, C., Schlabach, S., and Barabási, A.-L. (2019) Network-based prediction of protein interactions, *Nat. Commun.*, **10**, 1240, doi: 10.1038/s41467-019-09177-y.
 97. Agapito, G., Guzzi, P.H., and Cannataro, M. (2013) Visualization of protein interaction networks: problems and solutions, *BMC Bioinformatics*, **14**, Suppl. 1, S1, doi: 10.1186/1471-2105-14-S1-S1.

98. Lehner, B., and Fraser, A.G. (2004) A first-draft human protein–interaction map, *Genome Biol.*, **5**, R63, doi: 10.1186/gb-2004-5-9-r63.
99. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., and Wanker, E.E. (2005) A human protein–protein interaction network: a resource for annotating the proteome, *Cell*, **122**, 957–968, doi: 10.1016/j.cell.2005.08.029.
100. Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., and Figeys, D. (2007) Large-scale mapping of human protein–protein interactions by mass spectrometry, *Mol. Syst. Biol.*, **3**, 89, doi: 10.1038/msb4100134.
101. Taylor, I.W., and Wrana, J.L. (2012) Protein interaction networks in medicine and disease, *Proteomics*, **12**, 1706–1716, doi: 10.1002/pmic.201100594.
102. Li, Q., Chen, W., Song, M., Chen, W., Yang, Z., and Yang, A. (2019) Weighted gene co-expression network analysis and prognostic analysis identifies hub genes and the molecular mechanism related to head and neck squamous cell carcinoma, *Cancer Biol. Ther.*, **20**, 750–759, doi: 10.1080/15384047.2018.1564560.
103. Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., and Wrana, J.L. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome, *Nat. Biotechnol.*, **27**, 199–204, doi: 10.1038/nbt.1522.
104. Sardu, M.E., Gilmore, J.M., Groppe, B.D., Dutta, A., Florens, L., and Washburn, M.P. (2019) Topological scoring of protein interaction networks, *Nat. Commun.*, **10**, 1118, doi: 10.1038/s41467-019-09123-y.
105. Chen, S.-J., Liao, D.-L., Chen, C.-H., Wang, T.-Y., and Chen, K.-C. (2019) Construction and analysis of protein–protein interaction network of heroin use disorder, *Sci. Rep.*, **9**, 4980, doi: 10.1038/s41598-019-41552-z.
106. Ackerman, E.E., Kawakami, E., Katoh, M., Watanabe, T., Watanabe, S., Tomita, Y., and Kawaoka, Y. (2018) Network-guided discovery of influenza virus replication host factors, *mBio*, **9**, doi: 10.1128/mBio.02002-18.
107. Macalino, S.J.Y., Basith, S., Clavio, N.A.B., Chang, H., Kang, S., and Choi, S. (2018) Evolution of *in silico* strategies for protein–protein interaction drug discovery, *Molecules*, **23**, 1963, doi: 10.3390/molecules23081963.
108. Miho, E., Roškar, R., Greiff, V., and Reddy, S.T. (2019) Large-scale network analysis reveals the sequence space architecture of antibody repertoires, *Nat. Commun.*, **10**, 1321, doi: 10.1038/s41467-019-09278-8.
109. Soetkamp, D., Raedschelders, K., Mastali, M., Sobhani, K., Bairey Merz, C.N., and Van Eyk, J. (2017) The continuing evolution of cardiac troponin I biomarker analysis: from protein to proteoform, *Expert Rev. Proteomics*, **14**, 973–986, doi: 10.1080/14789450.2017.1387054.
110. Van der Burgt, Y.E.M., and Cobbaert, C.M. (2018) Proteoform analysis to fulfill unmet clinical needs and reach global standardization of protein measurands in clinical chemistry proteomics, *Clin. Lab. Med.*, **38**, 487–497, doi: 10.1016/j.cll.2018.05.001.

METHODS OF COMPUTATIONAL INTERACTOMICS FOR INVESTIGATION OF INTERACTIONS OF HUMAN PROTEOFORMS

Review

E. V. Poverennaya*, O. I. Kiseleva, A. S. Ivanov, and E. A. Ponomarenko

Institute of Biomedical Chemistry, 119121 Moscow, Russia; E-mail: k.poverennaya@gmail.com

Received March 25, 2019

Revised September 16, 2019

Accepted October 7, 2019

The human genome contains ca. 20,000 protein-coding genes, which could be translated into millions of unique protein species (proteoforms). Proteoforms coded by a single gene often differ in function, which implies differences in their protein partners. By interacting with each other, proteoforms create a network reflecting dynamics of cellular processes in an organism. Perturbations of protein–protein interactions cause changes in the network topology, which often trigger pathological processes. The study of proteoforms is a relatively new research area in proteomics, that is why there are relatively few experimental studies on interactions of proteoforms. Bioinformatics tools allow facilitating various tasks providing valuable complementary information to experimental data such as, for example, expanding the possibilities in investigation of interactions of proteoforms.

Keywords: protein–protein interactions, interactomics, bioinformatics