

©Коллектив авторов

IN SILICO И IN CELLULO ПОДХОДЫ ДЛЯ ФУНКЦИОНАЛЬНОЙ АННОТАЦИИ СПЛАЙС-ФОРМ БЕЛКОВ ЧЕЛОВЕКА

О.И. Киселева, В.А. Арзуманян, И.Ю. Курбатов, Е.В. Поверенная*

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

Изяснение механизмов сплайсинга пре-мРНК не перестаёт интересовать ученых даже спустя более полувека с момента открытия того факта, что кодирующие участки в генах прерываются некодирующими. Превалирующее большинство генов человека имеют несколько вариантов мРНК, которые, в свою очередь, кодируют структурно и функционально разные варианты белков — в тканезависимой манере и с привязкой к конкретным этапам развития организма. Нарушение паттернов сплайсинга смещает баланс функционально различающихся белков в живой системе, искажает нормальные молекулярные пути и может спровоцировать возникновение и развитие патологий. За последние два десятилетия выполнено множество исследований в различных областях наук о жизни для более глубокого понимания механизмов сплайсинга и степени его влияния на функционирование живых систем. Целью данного обзора было суммирование экспериментальных и вычислительных подходов, используемых для выяснения функций сплайс-опосредованных белковых продуктов одного гена: на основе собственного опыта, накопленного в лаборатории интерактомики протеоформ Института биомедицинской химии, и лучших мировых практик.

Ключевые слова: функциональная аннотация; альтернативный сплайсинг; сплайс-варианты; протеоформы; гетерогенность протеома; мультиомные исследования

DOI: 10.18097/PBMC20247005315

ВВЕДЕНИЕ

Завершение проекта “Геном человека” окончательно расставило все точки над *i* применительно к центральной догме молекулярной биологии, продемонстрировав масштабы нарушения принципа “один ген — один белок”. Вот уже более четверти века общепринято, что альтернативный сплайсинг (АС) в высших эукариотах увеличивает разнообразие вариантов сочетания последовательностей экзонов и тем самым расширяет спектр белковых продуктов (рис. 1), кодируемых относительно небольшим набором генов [1]. Более 95% мультиэкзонных генов человека продуцируют более одного (конститутивного) сплайс-варианта [2]. С развитием высокопроизводительных методов анализа нуклеотидных и аминокислотных последовательностей у широкого круга исследователей появилась возможность анализировать молекулярные профили интересующих их объектов не только на уровне мастерного белка (т.е. обобщённого образа белковых продуктов гена [3]), но и на уровне конкретных протеоформ.

1. СВЯЗЬ АЛЬТЕРНАТИВНОГО СПЛАЙСИНГА С ЗАБОЛЕВАНИЯМИ

“Тонкая настройка” всей биологической системы за счёт АС способна менять её молекулярный состав. Например, титин, имеющий наиболее длинную аминокислотную последовательность (достигает 38138 аминокислотных остатков), меняет свою превалирующую сплайс-форму по ходу взросления человека [4]. Такие модификации

изменяют длину белка и относительную жёсткость, а потому влияют на напряжение желудочка в покое и ассоциируются с приобретёнными формами сердечной недостаточности. Транскрипционные и посттранскрипционные изменения, которые увеличивают длину и растяжимость титина, делая саркомер длиннее и мягче, связаны с систолической дисфункцией и расширением левого желудочка. Модификации титина, которые укорачивают и сам белок, и саркомер, связаны с диастолической дисфункцией [5].

В последние годы в исследованиях молекулярной гетерогенности наметился тренд на поиск взаимосвязи паттернов сплайсинга с возникновением и развитием заболеваний. В общем случае, такие исследования носят панорамный характер и нацелены на формирование наиболее полных транскриптомных и трансляционных профилей исследуемых объектов [6–9]. Доподлинно известно, что нарушения механизмов сплайсинга, в зависимости от их масштабности, могут запускать продуцирование неработоспособных белков [10]. Такие белки, имеющие изменённую функцию, искажают отлаженные процессы дифференцировки, роста, межклеточной коммуникации и апоптоза, иначе говоря, нарушают нормальное функционирование всего организма и способны привести к его гибели [11]. Особенно ярко связь aberrаций сплайсинга показана на примере “визитных карточек” рака: каждой отличительной чертой, характерной для онкозаболеваний, соответствует случай нарушенного сплайсинга (рис. 2). В 2022 году ставшие классическими десять типичных признаков рака (пролиферативный сигналинг, репликационное бессмертие, индукция ангиогенеза, уклонение от супрессоров роста и т.д.)

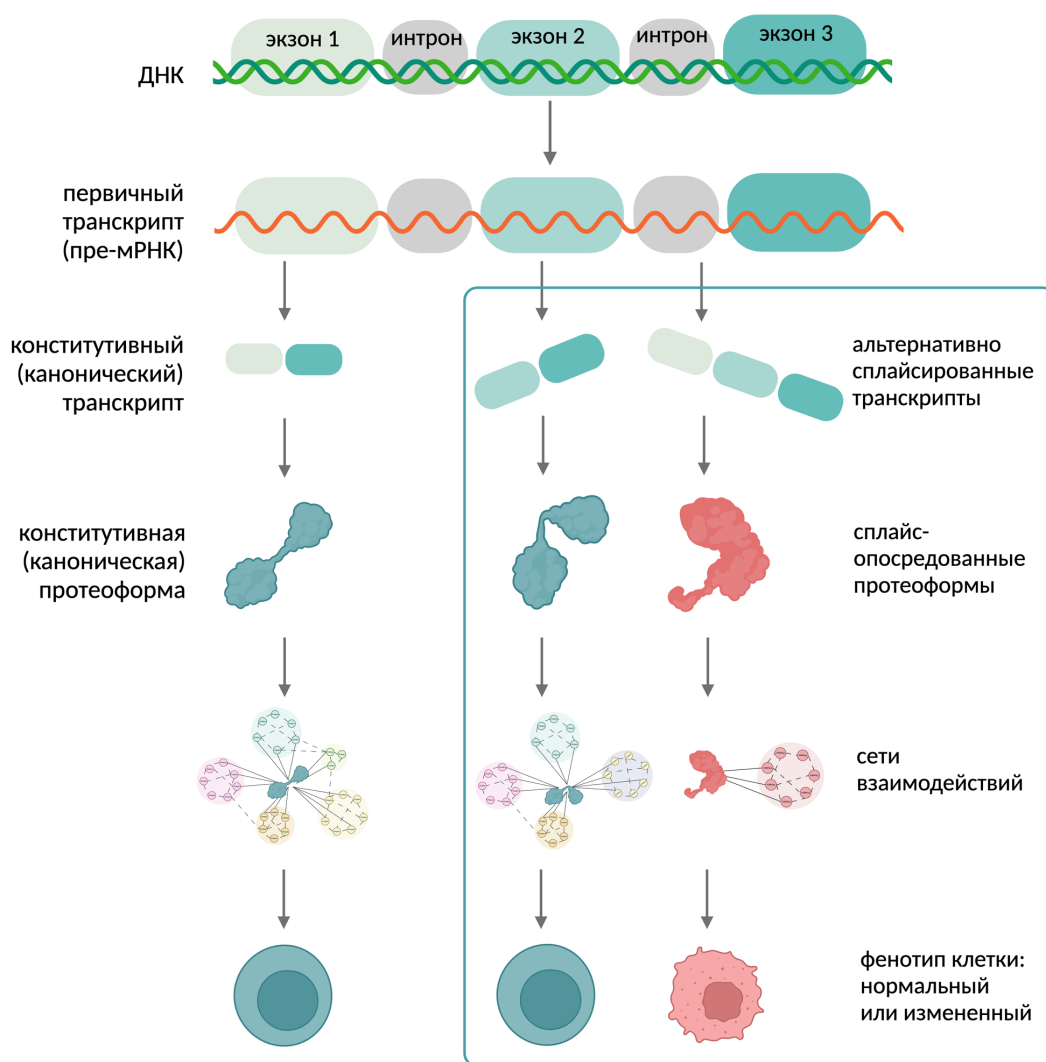


Рисунок 1. Этапы передачи генетической информации, формирующие итоговый фенотип объекта. В результате альтернативного сплайсинга (АС) один ген может продуцировать несколько сплайс-вариантов, представляющих собой различную комбинацию экзонов (реже — с включениями интронов) и транскрибирующихся с разных 3'- и 5'-сайтов сплайсинга.

дополнились ещё четырьмя: эпигенетическим перепрограммированием [12], полиморфным микробиомом [13], старением клеток [14, 15], а также разблокировкой пластичности фенотипа [16].

Сложно подобрать пример онкологического заболевания, для которого не было бы выявлено случаев нарушения нормальных паттернов сплайсинга. Накоплены знания об aberrациях сплайсинга как в солидных опухолях (в мозге [17], печени [18], коже [19], почках [20], лёгких [21], молочной железе [22], шейке матки [23], яичниках [24], простате [25]), так и при онкогематологии. Большинство случаев патологического АС объясняется генетическими мутациями в сайтах конститутивного сплайсинга или нарушениями уровня экспрессии регуляторных факторов сплайсосомы [26].

Массив подобных работ позволил обнаружить эффект свитчинга (англ. switching), описывающего смену или “переключение” преобладающего по экспрессии сплайс-варианта гена при переходе от нормальной ткани к патологически изменённой [27, 28]. Данное

явление определяется на основе изменений пропорций сплайс-вариантов и/или их дифференциальной экспрессии всех экспрессируемых транскриптов. Накоплено достаточно фактов, подтверждающих, что изменение характера представленности сплайс-вариантов одного гена влияет на развитие и функционирование живой клетки в норме и при патологии [29–31]. В общем случае такие выводы делаются на основе анализа обогащения наборов генов или метаболитов, а не детального исследования причинно-следственных связей. Кроме того, зачастую при анализе свитчинга и в целом при интерпретации результатов транскриптомного профилирования рассматривают все возможные варианты транскриптов, часть из которых не транслируется в аминокислотную последовательность [32]. Как мы показали ранее, фокусирование исключительно на белок-продуцирующих транскриптах ожидаемо сокращает массив случаев переключения преобладающей сплайс-формы, но не сводит его к десяткам случаев [28], косвенно подтверждая гипотезу о разных функциональных свойствах протеоформ.

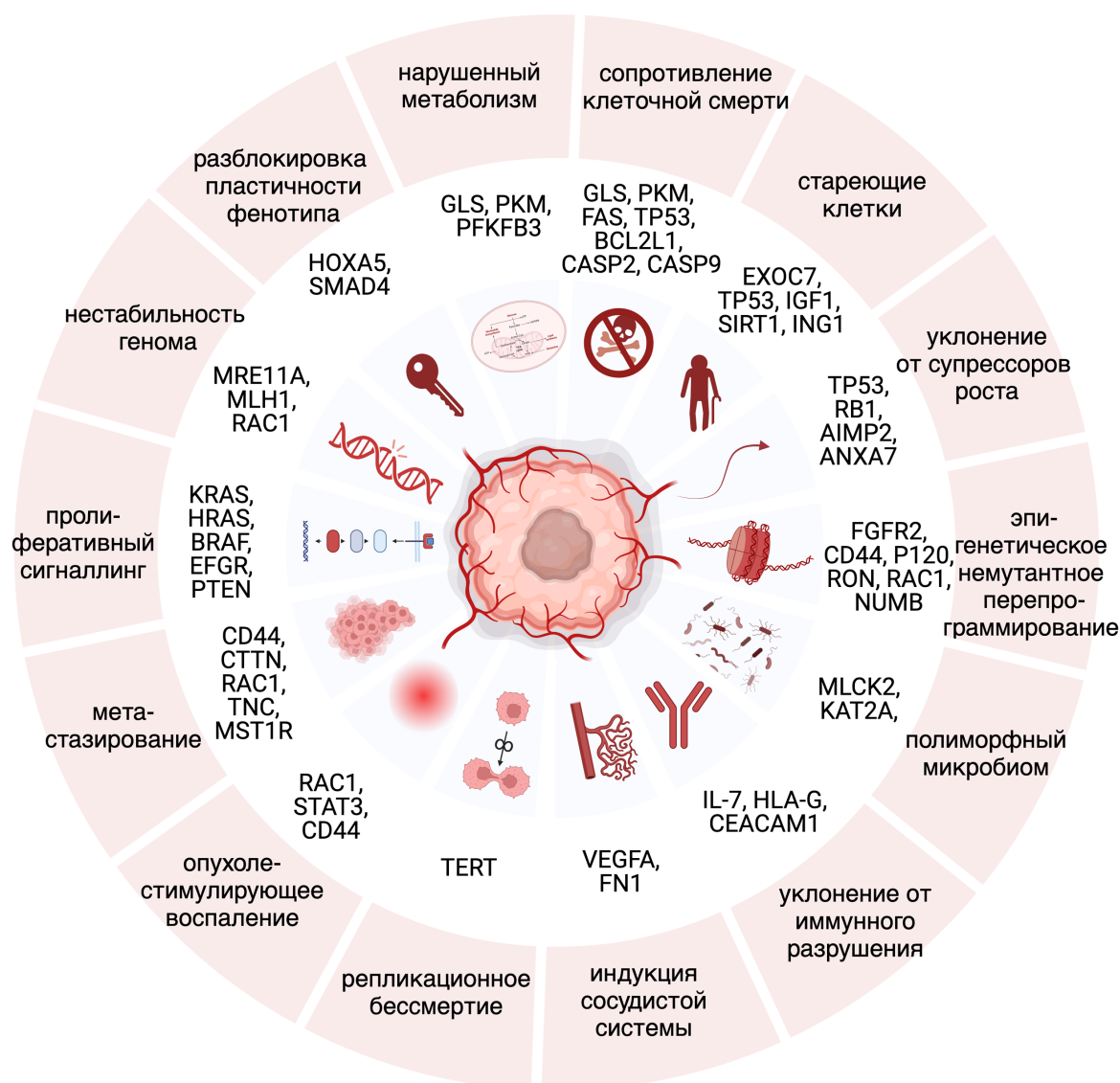


Рисунок 2. Место альтернативного сплайсинга в реализации текущей версии “визитных карточек” рака — концепции, призванной снизить сложность онкофенотипов до условного набора принципов, общих для возникновения и развития различных видов раковых заболеваний. Иллюстрация построена с учётом эволюции концепции на протяжении последних 20 лет [26–28] и приводит знаковые примеры генов, сплайсинг которых достоверно ассоциирован с возникновением и развитием онкологических заболеваний.

Страховой от соблазнительной “*post hoc, ergo propter hoc*” ошибки при поиске биологического смысла в данных секвенирования, касающихся переключения альтернативного сплайсинга, является переход от количества данных к их качественному (функциональному) осмыслению. Чтобы в полной мере использовать знания о паттернах сплайсинга как источник диагностических, прогностических, предиктивных и терапевтических инструментов, надо не только заметить численные закономерности, но ещё и определить функции кодируемых одним геном сплайс-вариантов.

2. ФУНКЦИОНАЛЬНАЯ ГЕТЕРОГЕННОСТЬ СПЛАЙС-ВАРИАНТОВ

По-разному сплайсированные белковые продукты демонстрируют разную ферментативную активность,

локализуются в разных компартментах клетки [33] и зачастую ведут себя как отдельные белки, а не как взаимозаменяемые варианты друг друга. Они могут проявлять доминантно-негативные эффекты по отношению к другим формам, кодируемым тем же геном, экспрессироваться в большей или меньшей степени, чем конститутивный вариант, или даже обладать противоположными функциями.

Самый яркий, но далеко не единственный пример гена с функционально разными сплайс-вариантами — это BCL2L1 из семейства регуляторов апоптоза. Сплайс-варианты этого гена различаются из-за наличия альтернативных 5' сайтов сплайсинга во втором экзоне, причём короткая форма (BCL-XS) запускает процессы клеточной смерти, а длинная (BCL-XL), напротив, обладает антиапоптотической функцией и зачастую активируется при раке [34]. Выявленная закономерность перспективна с практической точки

зрения: антисенс-терапия, направленная на изменение соотношения между двумя протеоформами BCL₂, позволяет повысить чувствительность клеток к индуцированному химиотерапевтическими препаратами апоптозу [35]. Аналогичная ситуация обстоит с представителем суперсемейства рецепторов факторов некроза опухолей TNFR2, который также кодирует два варианта сплайсинга с антагонистическими функциями. Канонические вариант рецептора TNFR2 опосредует апоптоз, индуцированный TNF- α , а более короткий вариант DS-TNFR2, лишённый аминокислотной последовательности, кодируемой седьмым и восьмым экзонами, напротив, апоптоз блокирует [35]. Ещё один пример — сверхэкспрессирующийся в раковых клетках ген IG20, который кодирует как минимум шесть сплайс-форм (IG20, IG20-PA или IG20pa, DENN или MADD, DENN-SV, KIAA0358 и IG20-SV4), отличающихся по своим апоптотическим свойствам [36, 37].

Приведённые примеры функционального различия отдельных сплайс-вариантов выглядят впечатляюще, но подобная глубина проработки и выяснения роли сплайсинга в формировании здорового или искажённого фенотипа скорее исключительна и достигнута лишь для нескольких десятков белок-кодирующих генов, причём зачастую на уровне транскриптов и транслятов. Потребовалось несколько лет и несколько итеративных подходов к исследованию источников гетерогенности омикс-слоев, чтобы добиться единомыслия в вопросах альтернативного сплайсинга. Здесь показателен путь лаборатории структурной и вычислительной биологии Испанского национального центра изучения рака. Исходно, в 2015 году на основе результатов восьми крупномасштабных протеомных экспериментов и анализа депонированных в базы данных было высказано предположение, что АС не играет значительной роли в формировании белкового разнообразия [38]. С скромный набор детектированных альтернативно сплайсированных белков натолкнул исследователей на мысль о том, что большинство белок-кодирующих генов, по-видимому, производят лишь один — канонический — белковый продукт [39], а большинство альтернативных вариантов не выдерживают селективного давления и вряд ли вообще функционально значимы. Несколько лет спустя те же исследователи опровергли свои предыдущие утверждения, анализируя альтернативный сплайсинг на уровне белка [40], который для трети генов проявляется тканеспецифичным образом. Совсем недавно эта же группа, основываясь на результатах крупных протеомных экспериментов, разработала биоинформатический инструмент TRIFID для предсказания функциональной значимости сплайс-форм и вернулась к утверждению, что 85% альтернативных вариантов транскриптов скорее всего незначимы [41]. Такие возвратно-поступательные движения при исследовании гетерогенности на разных омикс-уровнях побуждают проявлять осторожность при исследовании адаптивности АС и попытках установить соответствие между разнообразием вариантов мРНК и кодируемых ими белков.

3. СЛОЖНОСТИ ИССЛЕДОВАНИЯ ФУНКЦИЙ НА УРОВНЕ ПРОТЕОМА

Полученные на уровне транскриптома выводы с трудом переносятся на протеом, особенно это относится к попыткам установить количественные закономерности. Коэффициент Пирсона для корреляции между содержанием транскрипта и соответствующего белка обычно не превышает 0,5, как было показано в работах по анализу имеющихся взаимосвязей для изучения возможности построения модели представленности белков на основе транскриптомных и транслятомных данных [42–45]. Дополнительные усилия по моделированию эффекта регуляции синтеза белка после сплайсинга пре-мРНК смогли объяснить 30% различий в соотношениях белка и мРНК [43]. Другим достижением в вопросе поиска согласованности между содержанием транскрипта и белка стало довольно естественное замечание того, что превышение определённых уровней экспрессии транскрипта является хорошим предиктором экспрессии белка [46]. Тем не менее, очевидно, что для полноты формирования представления о функционировании живой системы нельзя пренебречь протеомной компонентой.

Пептидоцентричная природа протеомных данных зачастую не позволяет различить отдельные сплайс-формы [43–45, 47] из-за того, что специфичные для конкретной протеоформы пептиды не всегда удаётся выделить и надёжно детектировать [48]. Для повышения качества покрытия белкового сиквенса используют коктейли протеаз [49], процессинг данных в режиме *de novo* [50, 51] и ортогональные технологии секвенирования [52, 53]; однако даже сумма всех этих усилий не позволяет в полной мере оценить количество протеоформ.

Сегодня всё ещё остаётся много белых пятен на карте системной биологии, но общий консенсус состоит в том, что движущей силой живых систем прежде всего являются белки. Изучение отдельных вариантов белков уже достигло существенных успехов [54–58]. С нашей стороны было предложено усилить результаты масс-спектрометрического профилирования белков клеточной линии HepG2, распределившихся по ячейкам двумерного геля в соответствии с их физико-химическими свойствами [59], за счёт применения кастомизированной поисковой библиотеки. Такая библиотека, построенная на данных транскриптомного секвенирования исследуемого образца клеточной линии, даёт возможность наиболее точно сформировать поисковое пространство ожидаемых протеоформ. С одной стороны, это достигается благодаря учёту специфичных для конкретного образца сплайс-вариантов и последовательностей с точечными заменами. С другой стороны, объём поискового пространства сверху удаётся ограничить благодаря игнорированию последовательностей, продуцирование которых, согласно транскриптомным данным, не следует ожидать в исследуемом образце. Интегрированный анализ в клетках HepG2 позволил увеличить число выявленных протеоформ на 76% по сравнению со стандартным панорамным

профилированием без предварительного фракционирования на двумерном геле. Данный эффект достигается в результате синергии двух факторов: во-первых, снижения сложности биологической смеси, во-вторых — дополнительных знаний о физико-химических свойствах протеоформ [52, 53].

Применение оригинальных масс-спектрометрических подходов [60] и технологий антигенового обогащения [61] приблизило протеомику к ответу на вопрос: как финальная последовательность и структура формирует функционал белковых вариантов и каким образом отдельные варианты белков влияют на жизнеспособность всей живой системы?

4. ПЕРЕХОД ОТ ДАННЫХ ДИФФЕРЕНЦИАЛЬНОЙ ЭКСПРЕССИИ К ФУНКЦИОНАЛЬНОЙ АННОТАЦИИ

Популяризация методов секвенирования последовательностей РНК и возможность анализировать представленность генов и отдельных сплайс-вариантов позволили накопить данные, демонстрирующие разницу их экспрессии в разных условиях. Опубликованы сотни работ, сопоставляющие представленность генных продуктов в нормальной и опухолевой тканях и оценивающие изменения молекулярных профилей после нокаута или нокдауна [62, 63]. Результаты таких экспериментов могут быть использованы повторно, в частности, для выявления новых функций белков [33]. Обнаружение дифференциально экспрессированных генов для определения функций белков — стандартный путь транскриптомных работ. Транскриптомика достигла зрелости как область науки, и протоколы получения и обработки транскриптомных данных оптимизированы, надёжны и эффективны [64]. Это позволяет разрабатывать вычислительные методы для систематического исследования функции белка на уровне изоформ [65–68].

Выяснение функции белков — это главный “активационный барьер” в формировании систематического представления об устройстве живых систем. Высокопроизводительное секвенирование становится всё более доступным, данные множатся, однако качественный переход от сведений об экспрессии отдельно взятого гена к пониманию роли конкретной белковой молекулы ещё не произошёл. Несмотря на значительные успехи в каталогизации протеома, не решены фундаментальные вопросы относительно ролей отдельных белков в сложном функционирующем протеомном механизме. Причина этого — значительное различие в структурах и размерах информационного пространства транскриптома и протеома [69].

5. ПРАКТИКА ОПРЕДЕЛЕНИЯ ФУНКЦИЙ БЕЛКА

В условиях отсутствия общепризнанного стандарта определения функции белка, большая часть проводимых функциональных аннотаций носит предсказательный характер за счёт широкого использования методов биоинформатики

в приложении к большим омикс-данным и имеющейся, по сути, фрагментарной информации о регуляции клеточных процессов. Существенный разрыв между методами экспериментальной и вычислительной биохимии по трудоёмкости, дороговизне и экспрессности объясняет преобладание биоинформатических предсказаний над эмпирическими доказательствами при определении функций белков. В наших предыдущих ретроспективных исследованиях на примере базы данных neXtProt, известной полнотой и надёжностью публикуемых сведений о белках человека, мы проанализировали эволюцию терминологии, применяющуюся для описания функции белка [70]. Мы заметили, что накопленные аннотации в преобладающем большинстве случаев достигнуты вычислительными методами, но даже лучшие представители биоинформатических инструментов зачастую дают неудовлетворительные результаты, когда речь идет об аннотации неканонических вариантов. Часто применяется постулат вины “за компанию” (англ. *guilty by association*): на основе результатов масс-спектрометрии с аффинной очисткой (англ. *affinity purification – mass spectrometry, AP-MS*) и технологии двугибридного анализа (англ. *yeast-two-hybrid, Y2H*) функцию белка пытаются картировать на биохимические процессы, изучая контакты или “рукопожатия” целевых белков. Ещё одна проблема, которую мы заметили при исследовании трендов функциональной аннотации белков, касается репозитория с ручной проверкой депонируемых данных. Датафреймы (то есть табличные системы архитектуры данных по типу “наблюдения – переменные”) таких репозиториях (например, neXtProt) не оптимизированы для эффективного хранения информации в протеоформ-центричном режиме. Сейчас neXtProt предоставляет сведения приблизительно для 10 тыс. сплайс-вариантов, 10% из которых по-разному локализуются в клетке и имеют различные функции внутри одного гена [71].

В рамках существующих экспериментальных подходов к функциональной аннотации белков глобально можно выделить два направления: 1) нокаут или изменение уровня экспрессии интересующего гена для выявления искажающихся молекулярных путей на основе данных анализа одного или нескольких омикс-уровней и 2) анализ интерактома.

5.1. Потеря или приобретение функции (англ. *loss-of-function* или *gain-of-function*)

Подавление экспрессии гена, очевидно, изменяет биологические процессы, в которые вовлечён кодируемый им белок. Эти изменения нелегко определить и/или однозначно интерпретировать: имеющиеся сведения о молекулярных путях фрагментарны, а сами пути нелинейны и зачастую дублируют ряд этапов в клеточных процессах. Дополнительная сложность в подобных исследованиях привносится за счёт направленного изменения экспрессии — внесения точечных мутаций, нарушающих рамку считывания, или

использования явления интерференции. В настоящий момент наиболее популярный способ нокаута, а также нокада или нокина (в ситуациях, когда нокаут невозможен или, наоборот, экспрессия слишком низкая) — это применение генетического редактирования на основе CRISPR-Cas9 системы [72], дающее более предсказуемый и стабильный результат, чем интерферирующие микроРНК. Как мы ранее показали, возможности методов на основе CRISPR-Cas9 для изучения свойств белков и их разнообразия в протеомных исследованиях впечатляют, однако для исследования сплайс-форм такие технологии ещё широко не применялись [72].

В немногочисленных работах, описывающих опыт приложения методов генетического редактирования в отношении сплайс-форм, доказывалась важность их присутствия для сохранения функции или фенотипа исследуемого объекта. Например, нокаут сплайс-форм, кодируемых геном *Reep6*, показал, что именно канонический вариант *Reep6.1* критичен для функционирования палочек в составе сетчатки [73]. При этом, вторая сплайс-форма важна для сохранения фертильности у самцов мышей, причём в семенниках экспрессируются оба варианта. Сравнительный протеомный и фенотипический анализ клеточных линий ES-2 и OVCAR-8 с нокаутированными сплайс-формами рецептора TGFβ позволил описать разные роли этих протеоформ в развитии рака яичников [74]. В работе по оценке специфичных для рака желудка сплайс-форм через нокаут промоторов определили, что у супрессора опухоли ZFNХ3 сплайс-формы имеют противоположные функции, аналогично вариантам BCL2 [75]. Любопытно, что в качестве терапии мышечной дистрофии Дюшенна — генетического заболевания, обусловленного нарушением трансляции дистрофина из-за мутаций, приводящих к сдвигу рамки считывания — разрабатываются методы по исключению экзонов с такими ошибками для экспрессии более короткого варианта белка. При подавлении экспрессии канонического варианта показана эффективность в частичном сохранении функции у альтернативно сплайсированного дистрофина [76, 77].

В своей работе по исследованию функции белка митохондриального комплекса TOMM34 мы проводили оценку изменений до/после его нокаута на транскриптомном, протеомном и метаболомном уровнях. Одним из критериев выбора TOMM34 (помимо его малоизученности) стало наличие только одной транслируемой мРНК, согласно UniProt, что позволило описать функциональную роль конкретной аминокислотной последовательности [63]. При нокауте TOMM34 нам потребовалось внести пять мутаций в первый экзон для полного подавления его экспрессии. При фокусировке на конкретную сплайс-форму, помимо технических сложностей, увеличиваются риски нарушения рамки считывания и дополнительных off-target эффектов, что может в свою очередь повлиять на каскад молекулярных событий [78].

5.2. Функциональная аннотация через интерактом

Большинство сплайс-вариантов внутри одного гена имеют менее 50% общих партнёров по межмолекулярным взаимодействиям [33]. Примечательно, что партнёры, взаимодействующие с конкретным сплайс-вариантом, как правило, экспрессируются в высокой степени тканеспецифично и принадлежат к отдельным функциональным модулям [34]. Именно ассоциированность белковых партнёров с определёнными функциями позволяет предположить, что и исследуемый белок также с ней связан (т.е. применяя концепцию *guilty by association*). В случае сплайс-форм дополнительная сложность заключается в том, что функциональная аннотация проводится по совокупной информации для протеоформ, кодируемых одним геном.

В настоящий момент сформированы интерактомные профили 80% генов человека [79, 80], и наметился тренд на выявление белок-белковых взаимодействий (ББВ) для сплайс-форм. Экспериментально бинарные взаимодействия можно изучать двугибридными методами за счёт использования характеристичных последовательностей в качестве целевого белка или методом AP-MS, предоставляющим сведения о молекулярном комплексе. Одна из наиболее впечатляющих работ по определению ББВ для сплайс-форм была выполнена в 2016 году, в рамках которой с помощью Y2H были определены интерактомные профили для 366 из более тысячи исследуемых сплайс-форм [36]. Авторы работы показали, что даже на небольшой выборке интерактомная сеть увеличивается в 3,2 раза по сравнению с геноцентричным подходом.

В AP-MS экспериментах идентификация белковых партнёров осуществляется масс-спектрометрически, то есть по пептидам, уникально картирующимся на аминокислотную последовательность конкретной протеоформы [81]. Одним из первых ресурсов, содержащих данные о ББВ для сплайс-форм, стала база данных IntACT [82], в которую депонируются результаты интерактомных экспериментов, в том числе, выполненных методами AP-MS. Несмотря на то, что подходы на основе AP-MS дают возможность не только использовать сплайс-форму как целевой белок, но и описывать конкретные протеоформы в получаемых комплексах, акцента на сплайс-формы не делается [83]. Иллюстративным примером является один из крупнейших проектов BioPlex: несмотря на упоминание сплайс-форм, данные предоставляются в геноцентричном формате [84].

Накопленные результаты AP-MS дают возможность определять новые взаимодействия, включая сплайс-специфичные ББВ. Так, мы переанализировали массив масс-спектрометрических данных проекта BioPlex 2.0. По статистически значимой частоте совместной встречаемости удалось определить 287474 взаимодействия, на основе которых для 391 белка впервые была спрогнозирована функция, а для 31 гена показана разница в интерактомных профилях кодируемых ими сплайс-форм [85].

Хотя ББВ являются основой клеточных процессов, немалую роль играют и белок-метаболические взаимодействия (БМВ), способы выявления которых были нами описаны в работе [86]. Ценность БМВ для распутывания интерактома сплайс-форм можно проиллюстрировать на примере 505 сплайс-форм: их уникальная роль в биохимических процессах заметна при анализе взаимодействий с малыми молекулами, но не различается при применении стандартных протоколов анализа [87].

Немалую часть интерактомных методов представляют биоинформатические алгоритмы. Они сопоставляют последовательности и переносят интерактомные профили и функции с изученных объектов на неизученные или интегрируют различные типы данных для улучшения аннотации интерактомов. В своей работе пятилетней давности мы описали существующие подходы к аннотации протеоформ [88], с того времени основные принципы не изменились. Тем не менее, в настоящее время всё больше работ посвящены разработке методов интегрирования данных о ББВ со сведениями об экспрессии/трансляции для выделения специфичных взаимодействий непосредственно для сплайс-форм. Например, метод DIGGER, основанный на совокупном анализе интерактомов, данных о взаимодействии доменов и аминокислотных остатков, а также экспрессии, позволяет извлекать специфичные для сплайс-форм подсети [89]. В более продвинутом методе LINDA для расшифровки интерактомов сплайс-форм используют дополнительно данные о транскрипционных факторах, а также результаты DIGGER [90]. Примером ещё более специализированного ресурса для предсказания функций сплайс-вариантов может случить NEASE (метод сетевого обогащения для событий AC) [91], который при выделении подсетей для сплайс-форм учитывает не только интерактомные данные и количественные изменения, но и структурные особенности белков по данным DOMINE [92], 3did [93], Eukaryotic Linear Motif [94] и PDB [95].

5.3. Биоинформатические методы и инструменты

Биоинформатические методы для функциональной аннотации протеоформ развиваются не только в рамках интерактомного анализа. За 20 лет активных омиксных исследований удалось выстроить архитектуру различных баз знаний. Среди них как узкоспециальные — исключительно про сплайсинг [96–99], так и всеобъемлющие, в числе которых UniProt [100], neXtProt [101], RefSeq (NCBI) [102], Ensembl-GENCODE (EMBL) [103] и MANE [104]. Эти ресурсы суммируют результаты сотен и тысяч экспериментов, способных различить сплайс-варианты одного гена на разных омикс-уровнях. Каждый из перечисленных ресурсов с аннотациями аминокислотных и нуклеотидных последовательностей человека приводит сведения о более чем 100 тысячах белок-кодирующих транскриптов [105, 106]. При этом, пулы рапортируемых данных сильно отличаются: так, в 2018 году было показано, что базы данных

ENCODE и RefSeq согласуются лишь на 1/6 часть от всех представленных транскриптов [105].

Накопленные знания и экспериментальные данные способствовали развитию биоинформатических методов оценки функции белков. Для исследования функции генов обычно применяют выравнивание последовательностей и анализ экспрессии. Главная идея при использовании этих подходов состоит в предположении о том, что консервативные для многих биологических видов последовательности скорее функциональны [107]. Подобной логикой пользуются применительно и к сплайс-вариантам: сохранность определённой последовательности (например, белка, кодируемого альтернативным экзоном) у видов, которые эволюционно отдалились друг от друга десятки миллионов лет назад, указывает на её функциональную значимость [108]. Аналогично, чем больше накоплено доказательств экспрессии того или иного транскрипта в разных объектах, тем вероятнее его функциональность. Функциональную аннотацию осложняет тот факт [109], что экспрессия большинства сплайс-вариантов у человека весьма низкая (часто на уровне биологического шума [110]), и детектируются они не повсеместно, а тканеспецифически [105].

Практически любой функциональный анализ сводится к сопоставлению данных об известной роли интересующего белка в клеточных процессах. Среди разных ресурсов, описывающих и каталогизирующих функциональную аннотацию генов и их продуктов, наиболее популярной является терминологическая система Gene Ontology (GO) [111]. Система включает в себя 45000 терминов и объединяет их в три субонтологии: молекулярные функции (т.е. термины, описывающие функции белков, например, киназную активность); биологические процессы (для описания последовательности событий, происходящих в клетке или организме, в которые вовлечены гены и кодируемые ими белки, например, клеточное деление или иммунный ответ), и клеточные компоненты (термины, описывающие локализацию белков: ядро, мембрана и т.д.). GO предлагает функциональные аннотации только на уровне генов, не детализируя информацию для отдельных сплайс-вариантов [112]. К настоящему моменту было разработано несколько алгоритмов: iMILP [67], mi-SVM [113], WLRM [114], IsoResolve [115], DeepIsoFun [116] и DIFFUSE [117], призванные детализировать термины GO по отношению к сплайс-формам, основанные на многовариантном обучении. Текущее ограничение таких методов заключается в невысокой точности прогнозирования, в том числе из-за сложности учёта при анализе иерархической структуры и обширной семантики терминов GO. Более того, в большинстве алгоритмов предусмотрено, что только одна сплайс-форма отвечает за реализацию функции гена, хотя на практике для выполнения этой функции могут взаимодействовать одновременно несколько вариантов [70]. В более прогрессивном алгоритме IsofunGO [116] акцент сделан на обнаружении различных аннотаций для отдельных изоформ.

Для определения функции сплайс-вариантов также используют технологии предсказания трёхмерной структуры белков, подавая на вход последовательность РНК или белка. Популярная программа AlphaFold2 позволяет в автоматическом режиме предсказать с высокой точностью 3D-структуры белков [108, 118, 119], в том числе сплайс-форм. Предполагается, что белковые молекулы, последовательность которых сворачивается в упорядоченную структуру (около 68% протеома человека [118, 120]), вероятнее всего функциональны, и напротив, плохо упакованные молекулы скорее всего не работоспособны. Подобный подход был испытан на пуле данных, суммирующем результаты более 10 тысяч транскриптомных экспериментов [108], с помощью программы ColabFold [121], которая является оптимизированной версией AlphaFold. Иллюстрацией определения функции сплайс-формы является разбор гена ASMT (кодирующего N-ацетилсеротонин-О-метилтрансферазу), который участвует в биосинтезе мелатонина. По совокупной информации транскриптомного профилирования шишковидной железы, биопсия которой была проведена ночью в связи с особенностями синтеза в ней мелатонина [122], было определено, что более чем в 10 раз превышающая другие сплайс-формы экспрессия характерна для варианта, показавшего более стабильную сборку белка.

Научное сообщество долгое время полагалось на два метода обнаружения и предсказания функции белков: RNA-seq и выравнивание последовательностей ДНК и белков для обнаружения эволюционной консервативности. В настоящее

время надежда возлагается на вычислительные методы на основе GO, предсказания структуры белков и интерактивных взаимодействий. Стоит признать, что это только первые шаги. Разработка вычислительных методов обнаружения функций сплайс-вариантов останется областью активных исследований на долгие годы.

ЗАКЛЮЧЕНИЕ

Экспериментаторы склонны рекуррентно изучать одни и те же белки. На долю всего лишь 30 белков головного мозга человека приходится 2/3 всей научной литературы, посвящённой анализу протеома мозга. Эффект Матфея, согласно которому “богатые становятся богаче, а бедные — беднее”, находит отражение и в вопросе функциональной аннотации белков. Такому положению дел способствует скорее наличие финансирования и имеющиеся технологические возможности, а не фундаментальная или практическая ценность исследования некоторого набора и так хорошо изученных белков. Ситуация естественным образом усугубляется при переходе от мастерных белков, которые мы определяем как совокупность белковых продуктов одного гена без конкретизации протеоформ [69], к конкретным белковым вариантам.

Изучив степень проработки вопроса по функциональной аннотации отдельных сплайс-форм, мы предлагаем вариант дорожной карты (рис. 3) и надеемся, что упорядоченные усилия всех участников сообщества позволят существенно продвинуться в понимании устройства живых систем.

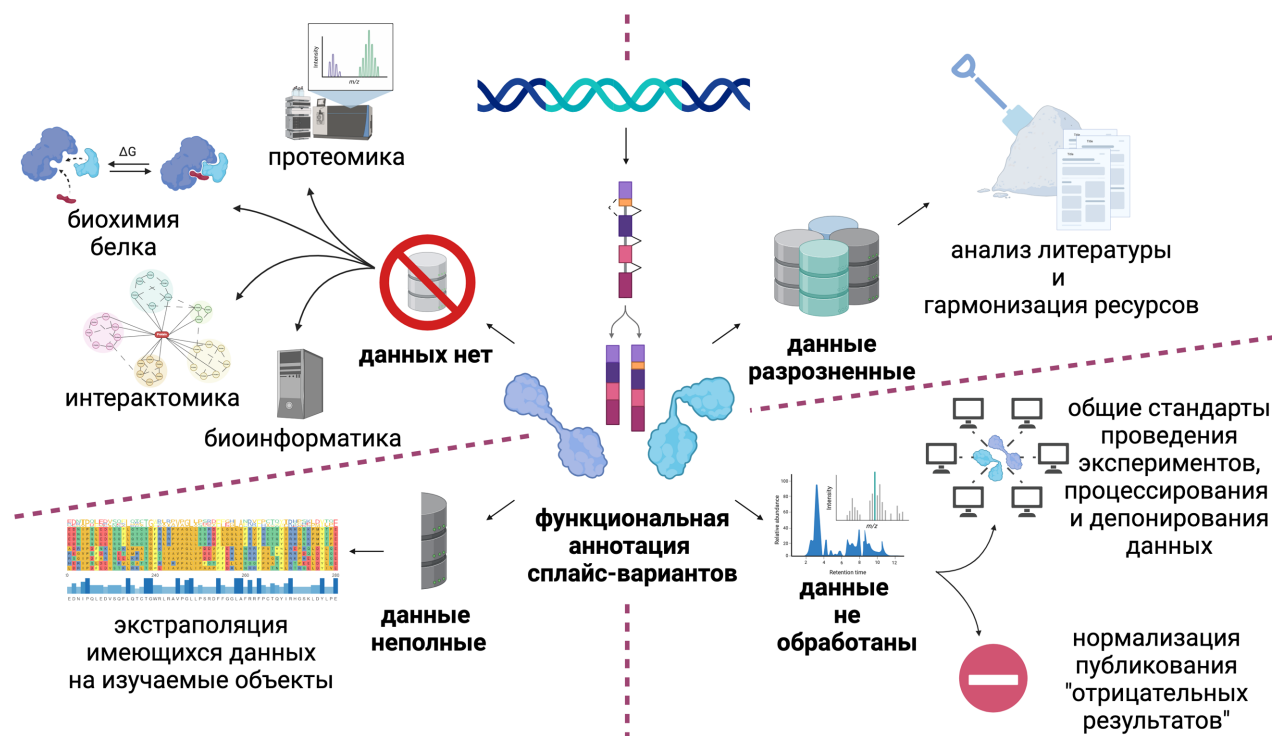


Рисунок 3. Предлагаемая дорожная карта по функциональной аннотации гетерогенного протеома.

ФИНАНСИРОВАНИЕ

Данное исследование поддержано грантом РНФ № 21-74-10061.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

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

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

ЛИТЕРАТУРА

1. Graveley B.R. (2001) Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet.*, **17**(2), 100–107. DOI: 10.1016/S0168-9525(00)02176-4
2. Lee Y., Rio D.C. (2015) Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.*, **84**, 291–323. DOI: 10.1146/annurev-biochem-060614-034316
3. Archakov A., Aseev A., Bykov V., Grigoriev A., Govorun V., Ivanov V., Khlunov A., Lisitsa A., Mazurenko S., Makarov A.A., Ponomarenko E., Sagdeev R., Skryabin K. (2011) Gene-centric view on the human proteome project: The example of the Russian roadmap for chromosome 18. *Proteomics*, **11**(10), 1853–1856. DOI: 10.1002/pmic.201000540
4. Tharp C.A., Haywood M.E., Sbaizero O., Taylor M.R.G., Mestroni L. (2019) The giant protein titin's role in cardiomyopathy: Genetic, transcriptional, and post-translational modifications of TTN and their contribution to cardiac disease. *Front. Physiol.*, **10**, 1436. DOI: 10.3389/fphys.2019.01436
5. Tharp C., Mestroni L., Taylor M. (2020) Modifications of titin contribute to the progression of cardiomyopathy and represent a therapeutic target for treatment of heart failure. *J. Clin. Med.*, **9**(9), 2770. DOI: 10.3390/jcm9092770
6. Vitting-Seerup K., Sandelin A. (2019) IsoformSwitchAnalyzeR: Analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*, **35**(21), 4469–4471. DOI: 10.1093/bioinformatics/btz247
7. Vitting-Seerup K., Sandelin A. (2017) The landscape of isoform switches in human cancers. *Mol. Cancer Res.*, **15**(9), 1206–1220. DOI: 10.1158/1541-7786.MCR-16-0459
8. Nowicka M., Robinson M.D. (2016) DRIMSeq: A Dirichlet-multinomial framework for multivariate count outcomes in genomics. *F1000Research*, **5**, 1356. DOI: 10.12688/f1000research.8900.2
9. Anders S., Reyes A., Huber W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res.*, **22**(10), 2008–2017. DOI: 10.1101/gr.133744.111
10. Liu Q., Fang L., Wu C. (2022) Alternative splicing and isoforms: From mechanisms to diseases. *Genes*, **13**(3), 401. DOI: 10.3390/genes13030401
11. Sterne-Weiler T., Sanford J.R. (2014) Exon identity crisis: Disease-causing mutations that disrupt the splicing code. *Genome Biol.*, **15**, 201. DOI: 10.1186/gb4150
12. Pradella D., Naro C., Sette C., Ghigna C. (2017) EMT and stemness: Flexible processes tuned by alternative splicing in development and cancer progression. *Mol. Cancer*, **16**, 8. DOI: 10.1186/s12943-016-0579-2
13. Zou C., Zan X., Jia Z., Zheng L., Gu Y., Liu F., Han Y., Xu C., Wu A., Zhi Q. (2023) Crosstalk between alternative splicing and inflammatory bowel disease: Basic mechanisms, biotechnological progresses and future perspectives. *Clin. Transl. Med.*, **13**(11), e1479. DOI: 10.1002/ctm2.1479
14. Georgilis A., Klotz S., Hanley C.J., Herranz N., Weirich B., Moranchó B., Leote A.C., d'Artista L., Gallage S., Seehawer M., Carroll T., Dharmalingam G., Wee K.B., Mellone M., Pombo J., Heide D., Guccione E., Arribas J., Barbosa-Morais N.L., Heikenwalder M., Thomas G.J., Zender L., Gil J. (2018) PTBP1-mediated alternative splicing regulates the inflammatory secretome and the pro-tumorigenic effects of senescent cells. *Cancer Cell*, **34**(1), 85–102.e9. DOI: 10.1016/j.ccell.2018.06.007
15. Deschênes M., Chabot B. (2017) The emerging role of alternative splicing in senescence and aging. *Aging Cell*, **16**(5), 918–933. DOI: 10.1111/ace1.12646
16. Yuan S., Norgard R.J., Stanger B.Z. (2019) Cellular plasticity in cancer. *Cancer Discov.*, **9**(7), 837–851. DOI: 10.1158/2159-8290.CD-19-0015
17. Babic I., Anderson E.S., Tanaka K., Guo D., Masui K., Li B., Zhu S., Gu Y., Villa G.R., Akhavan D., Nathanson D., Gini B., Mareninov S., Li R., Camacho C.E., Kurdistan S.K., Eskin A., Nelson S.F., Yong W.H., Cavenee W.K., Cloughesy T.F., Christofk H.R., Black D.L., Mischel P.S. (2013) EGFR mutation-induced alternative splicing of Max contributes to growth of glycolytic tumors in brain cancer. *Cell Metab.*, **17**(6), 1000–1008. DOI: 10.1016/j.cmet.2013.04.013
18. Duriez M., Mandouri Y., Lekbaby B., Wang H., Schnuriger A., Redelsperger F., Guerrera C.I., Lefevre M., Fauveau V., Ahodantin J., Quetier I., Chhuon C., Gourari S., Boissonnas A., Gill U., Kennedy P., Debzi N., Sitterlin D., Maini M.K., Kremsdorf D., Soussan P. (2017) Alternative splicing of hepatitis B virus: A novel virus/host interaction altering liver immunity. *J. Hepatol.*, **67**(4), 687–699. DOI: 10.1016/j.jhep.2017.05.025
19. Jensen M.A., Wilkinson J.E., Krainer A.R. (2014) Splicing factor SRSF6 promotes hyperplasia of sensitized skin. *Nat. Struct. Mol. Biol.*, **21**(2), 189–197. DOI: 10.1038/nsmb.2756
20. Sokół E., Kędzierska H., Czuby A., Rybicka B., Rodzik K., Tański Z., Bogusławska J., Piekietko-Witkowska A. (2018) MicroRNA-mediated regulation of splicing factors SRSF1, SRSF2 and hnRNP A1 in context of their alternatively spliced 3'UTRs. *Exp. Cell Res.*, **363**(2), 208–217. DOI: 10.1016/j.yexcr.2018.01.009
21. Sheng J., Zhao Q., Zhao J., Zhang W., Sun Y., Qin P., Lv Y., Bai L., Yang Q., Chen L., Qi Y., Zhang G., Zhang L., Gu C., Deng X., Liu H., Meng S., Gu H., Liu Q., Coulson J.M., Li X., Sun B., Wang Y. (2018) SRSF1 modulates PTPMT1 alternative splicing to regulate lung cancer cell radioresistance. *EBioMedicine*, **38**, 113–126. DOI: 10.1016/j.ebiom.2018.11.007
22. Xie R., Chen X., Chen Z., Huang M., Dong W., Gu P., Zhang J., Zhou Q., Dong W., Han J., Wang X., Li H., Huang J., Lin T. (2019) Polypyrimidine tract binding protein 1 promotes lymphatic metastasis and proliferation of bladder cancer via alternative splicing of MEIS2 and PKM. *Cancer Lett.*, **449**, 31–44. DOI: 10.1016/j.canlet.2019.01.041

23. Liu F., Dai M., Xu Q., Zhu X., Zhou Y., Jiang S., Wang Y., Ai Z., Ma L., Zhang Y., Hu L., Yang Q., Li J., Zhao S., Zhang Z., Teng Y. (2018) SRSF10-mediated IL1RAP alternative splicing regulates cervical cancer oncogenesis via mIL1RAP-NF- κ B-CD47 axis. *Oncogene*, **37**(18), 2394–2409. DOI: 10.1038/s41388-017-0119-6
24. Iborra S., Hirschfeld M., Jaeger M., Zur Hausen A., Braicu I., Sehoul J., Gitsch G., Stickeler E. (2013) Alterations in expression pattern of splicing factors in epithelial ovarian cancer and its clinical impact. *Int. J. Gynecol. Cancer*, **23**(6), 990–996. DOI: 10.1097/IGC.0b013e31829783e3
25. Fan L., Zhang F., Xu S., Cui X., Hussain A., Fazli L., Gleave M., Dong X., Qi J. (2018) Histone demethylase JMJD1A promotes alternative splicing of AR variant 7 (AR-V7) in prostate cancer cells. *Proc. Natl. Acad. Sci. USA*, **115**(20), E4584–E4593. DOI: 10.1073/pnas.1802415115
26. Zhang Y., Qian J., Gu C., Yang Y. (2021) Alternative splicing and cancer: A systematic review. *Signal Transduct. Target. Ther.*, **6**, 78. DOI: 10.1038/s41392-021-00486-7
27. Sebestyén E., Zawisza M., Eyra E. (2015) Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res.*, **43**(3), 1345–1356. DOI: 10.1093/nar/gku1392
28. Dolgalev G., Poverennaya E. (2023) Quantitative analysis of isoform switching in cancer. *Int. J. Mol. Sci.*, **24**(12), 10065. DOI: 10.3390/ijms241210065
29. Khan F., Anelo O.M., Sadiq Q., Effah W., Price G., Johnson D.L., Ponnusamy S., Grimes B., Morrison M.L., Fowke J.H., Hayes D.N., Narayanan R. (2023) Racial differences in androgen receptor (AR) and AR splice variants (AR-SVs) expression in treatment-naïve androgen-dependent prostate cancer. *Biomedicines*, **11**(3), 648. DOI: 10.3390/biomedicines11030648
30. Bonnal S.C., López-Oreja I., Valcárcel J. (2020) Roles and mechanisms of alternative splicing in cancer — implications for care. *Nat. Rev. Clin. Oncol.*, **17**(8), 457–474. DOI: 10.1038/s41571-020-0350-x
31. West S., Kumar S., Batra S.K., Ali H., Ghersi D. (2019) Uncovering and characterizing splice variants associated with survival in lung cancer patients. *PLoS Comput. Biol.*, **15**(10), e1007469. DOI: 10.1371/journal.pcbi.1007469
32. Tress M.L., Abascal F., Valencia A. (2017) Alternative splicing may not be the key to proteome complexity. *Trends Biochem. Sci.*, **42**(2), 98–110. DOI: 10.1016/j.tibs.2016.08.008
33. Kelemen O., Convertini P., Zhang Z., Wen Y., Shen M., Falaleeva M., Stamm S. (2013) Function of alternative splicing. *Gene*, **514**, 1–30. DOI: 10.1016/j.gene.2012.07.083
34. Yang X., Coulombe-Huntington J., Kang S., Sheynkman G.M., Hao T., Richardson A., Sun S., Yang F., Shen Y.A., Murray R.R., Spirohn K., Begg B.E., Duran-Frigola M., MacWilliams A., Pevzner S.J., Zhong Q., Wanamaker S.A., Tam S., Ghamsari L., Sahni N., Yi S., Rodriguez M.D., Balcha D., Tan G., Costanzo M., Andrews B., Boone C., Zhou X.J., Salehi-Ashtiani K., Charleatoux B., Chen A.A., Calderwood M.A., Aloy P., Roth F.P., Hill D.E., Iakoucheva L.M., Xia Y., Vidal M. (2016) Widespread expansion of protein interaction capabilities by alternative splicing. *Cell*, **164**(4), 805–817. DOI: 10.1016/j.cell.2016.01.029
35. Lainez B., Fernandez-Real J.M., Romero X., Esplugues E., Cacete J.D., Ricart W., Engel P. (2004) Identification and characterization of a novel spliced variant that encodes human soluble tumor necrosis factor receptor 2. *Int. Immunol.*, **16**(1), 169–177. DOI: 10.1093/intimm/dxh014
36. Kurada B.R.V.V.S.N., Li L.C., Mulherkar N., Subramanian M., Prasad K.V., Prabhakar B.S. (2009) MADD, a splice variant of IG20, is indispensable for MAPK activation and protection against apoptosis upon tumor necrosis factor- α treatment. *J. Biol. Chem.*, **284**(20), 13533–13541. DOI: 10.1074/jbc.M808554200
37. Efimova E.V., Al-Zoubi A.M., Martinez O., Kaithamana S., Lu S., Arima T., Prabhakar B.S. (2004) IG20, in contrast to DENN-SV, (MADD splice variants) suppresses tumor cell survival, and enhances their susceptibility to apoptosis and cancer drugs. *Oncogene*, **23**(5), 1076–1087. DOI: 10.1038/sj.onc.1207210
38. Ezkurdia I., Rodriguez J.M., Carrillo-de Santa Pau E., Vázquez J., Valencia A., Tress M.L. (2015) Most highly expressed protein-coding genes have a single dominant isoform. *J. Proteome Res.*, **14**(4), 1880–1887. DOI: 10.1021/pr501286b
39. Tress M.L., Abascal F., Valencia A. (2017) Most alternative isoforms are not functionally important. *Trends Biochem. Sci.*, **42**(6), 408–410. DOI: 10.1016/j.tibs.2017.04.002
40. Rodriguez J.M., Pozo F., di Domenico T., Vazquez J., Tress M.L. (2020) An analysis of tissue-specific alternative splicing at the protein level. *PLoS Comput. Biol.*, **16**(10), e1008287. DOI: 10.1371/journal.pcbi.1008287
41. Pozo F., Martinez-Gomez L., Walsh T.A., Rodriguez J.M., di Domenico T., Abascal F., Vazquez J., Tress M.L. (2021) Assessing the functional relevance of splice isoforms. *NAR Genom. Bioinform.*, **3**(2), lqab044. DOI: 10.1093/nargab/lqab044
42. Ponomarenko E.A., Krasnov G.S., Kiseleva O.I., Kryukova P.A., Arzumanyan V.A., Dolgalev G.V., Ilgisonis E.V., Lisitsa A.V., Poverennaya E.V. (2023) Workability of mRNA sequencing for predicting protein abundance. *Genes*, **14**(11), 2065. DOI: 10.3390/genes14112065
43. Eraslan B., Wang D., Gusic M., Prokisch H., Hallström B.M., Uhlén M., Asplund A., Pontén F., Wieland T., Hopf T., Hahne H., Kuster B., Gagneur J. (2019) Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues. *Mol. Syst. Biol.*, **15**(2), e8513. DOI: 10.15252/msb.20188513
44. Liu Y., González-Porta M., Santos S., Brazma A., Marioni J.C., Aebersold R., Venkitaraman A.R., Wickramasinghe V.O. (2017) Impact of alternative splicing on the human proteome. *Cell Rep.*, **20**(5), 1229–1241. DOI: 10.1016/j.celrep.2017.07.025
45. Tay A.P., Pang C.N.I., Twine N.A., Hart-Smith G., Harkness L., Kassem M., Wilkins M.R. (2015) Proteomic validation of transcript isoforms, including those assembled from RNA-seq data. *J. Proteome Res.*, **14**(9), 3541–3554. DOI: 10.1021/pr5011394
46. Vogel C., Marcotte E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.*, **13**(4), 227–232. DOI: 10.1038/nrg3185
47. Kosti I., Jain N., Aran D., Butte A.J., Sirota M. (2016) Cross-tissue analysis of gene and protein expression in normal and cancer tissues. *Sci. Rep.*, **6**, 24799. DOI: 10.1038/srep24799
48. Tran T.T., Bollineni R.C., Strozynski M., Koehler C.J., Thiede B. (2017) Identification of alternative splice variants using unique tryptic peptide sequences for database searches. *J. Proteome Res.*, **16**(7), 2571–2578. DOI: 10.1021/acs.jproteome.7b00126

49. Wang X., Codreanu S.G., Wen B., Li K., Chambers M.C., Liebler D.C., Zhang B. (2018) Detection of proteome diversity resulted from alternative splicing is limited by trypsin cleavage specificity. *Mol. Cell. Proteomics*, **17**(3), 422–430. DOI: 10.1074/mcp.RA117.000155
50. Karunratanakul K., Tang H.-Y., Speicher D.W., Chuangsuwanich E., Sriswasdi S. (2019) Uncovering thousands of new peptides with sequence-mask-search hybrid *de novo* peptide sequencing framework. *Mol. Cell. Proteomics*, **18**(12), 2478–2491. DOI: 10.1074/mcp.TIR119.001656
51. Bogdanow B., Zauber H., Selbach M. (2016) Systematic errors in peptide and protein identification and quantification by modified peptides. *Mol. Cell. Proteomics*, **15**(8), 2791–2801. DOI: 10.1074/mcp.M115.055103
52. Kiseleva O., Ponomarenko E., Poverennaya E. (2020) Empowering shotgun mass spectrometry with 2DE: A HepG2 study. *Int. J. Mol. Sci.*, **21**(11), 3813. DOI: 10.3390/ijms21113813
53. Поверенная Е.В., Киселева О.И., Пономаренко Е.А., Нарыжный С.Н., Згода В.Г., Лисица А.В. (2017) Мультиомная стратегия исследования протеома клеточной линии гепатоцеллюлярной карциномы HepG2. *Биомедицинская химия*, **63**(5), 373–378. [Poverennaya E.V., Kiseleva O.I., Ponomarenko E.A., Naryzhny S.N., Zgoda V.G., Lisitsa A.V. (2017) Multioomics study of HepG2 cell line proteome. *Biomeditsinskaya Khimiya*, **63**(5), 373–378.] DOI: 10.18097/PBMC20176305373
54. Киселева О.И., Лисица А.В., Поверенная Е.В. (2018) Протеоформы: методы исследования и клинические перспективы. *Молекулярная биология*, **52**(3), 394–410. [Kiseleva O.I., Lisitsa A.V., Poverennaya E.V. (2018) Proteoforms: Methods of analysis and clinical prospects. *Mol. Biol. (Mosk)*, **52**(3), 394–410.] DOI: 10.7868/S0026898418030047
55. Smith L.M., Agar J.N., Chamot-Rooke J., Danis P.O., Ge Y., Loo J.A., Paša-Tolić L., Tsybin Y.O., Kelleher N.L., Consortium for Top-Down Proteomics (2021) The human proteoform project: Defining the human proteome. *Sci Adv.*, **7**(46), eabk0734. DOI: 10.1126/sciadv.abk0734
56. Smith L.M., Kelleher N.L. (2018) Proteoforms as the next proteomics currency. *Science*, **359**(6380), 1106–1107. DOI: 10.1126/science.aat1884
57. Carbonara K., Andonovski M., Coorssen J.R. (2021) Proteomes are of proteoforms: Embracing the complexity. *Proteomes*, **9**(3), 38. DOI: 10.3390/proteomes9030038
58. Forgrave L.M., Wang M., Yang D., de Marco M.L. (2022) Proteoforms and their expanding role in laboratory medicine. *Pract. Lab. Med.*, **28**, e00260. DOI: 10.1016/j.plabm.2021.e00260
59. Naryzhny S. (2016) Towards the full realization of 2DE power. *Proteomes*, **4**(4), 33. DOI: 10.3390/proteomes4040033
60. Fornelli L., Toby T.K., Schachner L.F., Doubleday P.F., Srzentić K., deHart C.J., Kelleher N.L. (2018) Top-down proteomics: Where we are, where we are going? *J. Proteomics*, **175**, 3–4. DOI: 10.1016/j.jprot.2017.02.002
61. Chang A., Leutert M., Rodriguez-Mias R.A., Villén J. (2023) Automated enrichment of phosphotyrosine peptides for high-throughput proteomics. *J. Proteome Res.*, **22**(6), 1868–1880. DOI: 10.1021/acs.jproteome.2c00850
62. Romashin D., Rusanov A., Arzumanyan V., Varshaver A., Poverennaya E., Vakhrushev I., Netrusov A., Luzgina N. (2024) Exploring the functions of mutant p53 through TP53 knockout in HaCaT keratinocytes. *Curr. Issues Mol. Biol.*, **46**(2), 1451–1466. DOI: 10.3390/cimb46020094
63. Poverennaya E.V., Pyatnitskiy M.A., Dolgalev G.V., Arzumanyan V.A., Kiseleva O.I., Kurbatov I.Y., Kurbatov L.K., Vakhrushev I.V., Romashin D.D., Kim Y.S., Ponomarenko E.A. (2023) Exploiting multi-omics profiling and systems biology to investigate functions of TOMM34. *Biology*, **12**(2), 198. DOI: 10.3390/biology12020198
64. Rosati D., Palmieri M., Brunelli G., Morrione A., Iannelli F., Frullanti E., Giordano A. (2024) Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review. *Comput. Struct. Biotechnol. J.*, **23**, 1154–1168. DOI: 10.1016/j.csbj.2024.02.018
65. Li W., Liu C.-C., Kang S., Li J.-R., Tseng Y.-T., Zhou X.J. (2016) Pushing the annotation of cellular activities to a higher resolution: Predicting functions at the isoform level. *Methods*, **93**, 110–118. DOI: 10.1016/j.ymeth.2015.07.016
66. Tseng Y.-T., Li W., Chen C.-H., Zhang S., Chen J.J., Zhou X.J., Liu C.-C. (2015) IIIDB: A database for isoform-isoform interactions and isoform network modules. *BMC Genomics*, **16**(Suppl 2), S10. DOI: 10.1186/1471-2164-16-S2-S10
67. Li W., Kang S., Liu C.-C., Zhang S., Shi Y., Liu Y., Zhou X.J. (2014) High-resolution functional annotation of human transcriptome: Predicting isoform functions by a novel multiple instance-based label propagation method. *Nucleic Acids Res.*, **42**(6), e39. DOI: 10.1093/nar/gkt1362
68. Cruz L.M., Trefflich S., Weiss V.A., Castro M.A.A. (2017) Protein function prediction. *Methods Mol. Biol.*, **1654**, 55–75. DOI: 10.1007/978-1-4939-7231-9_5
69. Ponomarenko E.A., Poverennaya E.V., Ilgisonis E.V., Pyatnitskiy M.A., Kopylov A.T., Zgoda V.G., Lisitsa A.V., Archakov A.I. (2016) The size of the human proteome: The width and depth. *Int. J. Anal. Chem.*, **2016**, 7436849. DOI: 10.1155/2016/7436849
70. Ilgisonis E.V., Pogodin P.V., Kiseleva O.I., Tarbeeva S.N., Ponomarenko E.A. (2022) Evolution of protein functional annotation: Text mining study. *J. Pers. Med.*, **12**(3), 479. DOI: 10.3390/jpm12030479
71. Zahn-Zabal M., Lane L. (2020) What will neXtProt help us achieve in 2020 and beyond? *Expert Rev. Proteomics*, **17**(2), 95–98. DOI: 10.1080/14789450.2020.1733418
72. Dolgalev G., Poverennaya E. (2021) Applications of CRISPR-Cas technologies to proteomics. *Genes*, **12**(11), 1790. DOI: 10.3390/genes12111790
73. Liang Q., Wu N., Zaneveld S., Liu H., Fu S., Wang K., Bertrand R., Wang J., Li Y., Chen R. (2021) Transcript isoforms of Reep6 have distinct functions in the retina. *Hum. Mol. Genet.*, **30**(21), 1907–1918. DOI: 10.1093/hmg/ddab157
74. Jacobs Catane L., Moshel O., Smith Y., Davidson B., Reich R. (2021) Splice-variant knock-out of TGFβ receptors perturbs the proteome of ovarian carcinoma cells. *Int. J. Mol. Sci.*, **22**(23), 12647. DOI: 10.3390/ijms222312647
75. Davies R., Liu L., Taotao S., Tuano N., Chaturvedi R., Huang K.K., Itman C., Mandoli A., Qamra A., Hu C., Powell D., Daly R.J., Tan P., Rosenbluh J. (2021) CRISPRi enables isoform-specific loss-of-function screens and identification of gastric cancer-specific isoform dependencies. *Genome Biol.*, **22**, 47. DOI: 10.1186/s13059-021-02266-6

76. Amoasii L., Hildyard J.C.W., Li H., Sanchez-Ortiz E., Mireault A., Caballero D., Harron R., Stathopoulou T.-R., Massey C., Shelton J.M., Bassel-Duby R., Piercy R.J., Olson E.N. (2018) Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science*, **362**(6410), 86–91. DOI: 10.1126/science.aau1549
77. Long C., Amoasii L., Mireault A.A., McAnally J.R., Li H., Sanchez-Ortiz E., Bhattacharyya S., Shelton J.M., Bassel-Duby R., Olson E.N. (2016) Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science*, **351**(6271), 400–403. DOI: 10.1126/science.aad5725
78. Dours-Zimmermann M.T., Zimmermann D.R. (2012) A novel strategy for a splice-variant selective gene ablation: The example of the versican V0/V2 knockout. *Methods Mol. Biol.*, **836**, 63–85. DOI: 10.1007/978-1-61779-498-8_5
79. Dimitrakopoulos G.N., Klapa M.I., Moschonas N.K. (2022) How far are we from the completion of the human protein interactome reconstruction? *Biomolecules*, **12**(1), 140. DOI: 10.3390/biom12010140
80. Huttlin E.L., Bruckner R.J., Navarrete-Perea J., Cannon J.R., Baltier K., Gebreab F., Gygi M.P., Thornock A., Zarraga G., Tam S., Szpyt J., Gassaway B.M., Panov A., Parzen H., Fu S., Golbazi A., Maenpaa E., Stricker K., Guha Thakurta S., Zhang T., Rad R., Pan J., Nusinow D.P., Paulo J.A., Schweppe D.K., Vaites L.P., Harper J.W., Gygi S.P. (2021) Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell*, **184**(11), 3022–3040.e28. DOI: 10.1016/j.cell.2021.04.011
81. Nesvizhskii A.I. (2012) Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics*, **12**(10), 1639–1655. DOI: 10.1002/pmic.201100537
82. Hermjakob H., Montecchi-Palazzi L., Lewington C., Mudali S., Kerrien S., Orchard S., Vingron M., Roechert B., Roepstorff P., Valencia A., Margalit H., Armstrong J., Bairoch A., Cesareni G., Sherman D., Apweiler R. (2004) IntAct: An open source molecular interaction database. *Nucleic Acids Res.*, **32**(Database issue), D452–D455. DOI: 10.1093/nar/gkh052
83. Frommelt F., Fossati A., Uliana F., Wendt F., Xue P., Heusel M., Wollscheid B., Aebersold R., Ciuffa R., Gstaiger M. (2024) DIP-MS: Ultra-deep interaction proteomics for the deconvolution of protein complexes. *Nat. Methods*, **21**(4), 635–647. DOI: 10.1038/s41592-024-02211-y
84. Huttlin E.L., Ting L., Bruckner R.J., Gebreab F., Gygi M.P., Szpyt J., Tam S., Zarraga G., Colby G., Baltier K., Dong R., Guarani V., Vaites L.P., Ordureau A., Rad R., Erickson B.K., Wühr M., Chick J., Zhai B., Kolippakkam D., Mintseris J., Obar R.A., Harris T., Artavanis-Tsakonas S., Sowa M.E., de Camilli P., Paulo J.A., Harper J.W., Gygi S.P. (2015) The BioPlex network: A systematic exploration of the human interactome. *Cell*, **162**(2), 425–440. DOI: 10.1016/j.cell.2015.06.043
85. Poverennaya E., Kiseleva O., Romanova A., Pyatnitskiy M. (2020) Predicting functions of uncharacterized human proteins: From canonical to proteoforms. *Genes*, **11**(6), 677. DOI: 10.3390/genes11060677
86. Kurbatov I., Dolgalev G., Arzumanyan V., Kiseleva O., Poverennaya E. (2023) The knowns and unknowns in protein-metabolite interactions. *Int. J. Mol. Sci.*, **24**(4), 4155. DOI: 10.3390/ijms24044155
87. Hernández Sánchez L.F., Burger B., Castro Campos R.A., Johansson S., Njølstad P.R., Barsnes H., Vaudel M. (2023) Extending protein interaction networks using proteoforms and small molecules. *Bioinformatics*, **39**(10), btad598. DOI: 10.1093/bioinformatics/btad598
88. Поверенная Е.В., Киселева О.И., Иванов А.С., Пономаренко Е.А. (2020) Методы вычислительной интерактомики в вопросах взаимодействия протеоформ человека. *Биохимия*, **85**(1), 80–92. DOI: 10.31857/S0320972520010066 [Poverennaya E.V., Kiseleva O.I., Ivanov A.S., Ponomarenko E.A. (2020) Methods of computational interactomics for investigating interactions of human proteoforms. *Biochemistry (Moscow)*, **85**(1), 68–79. DOI: 10.1134/S000629792001006X]
89. Louadi Z., Yuan K., Gress A., Tsoy O., Kalinina O.V., Baumbach J., Kacprowski T., List M. (2021) DIGGER: Exploring the functional role of alternative splicing in protein interactions. *Nucleic Acids Res.*, **49**(D1), D309–D318. DOI: 10.1093/nar/gkaa768
90. Gjerga E., Naarmann-de Vries I.S., Dieterich C. (2023) Characterizing alternative splicing effects on protein interaction networks with LINDA. *Bioinformatics*, **39**(Suppl 1), i458–i464. DOI: 10.1093/bioinformatics/btad224
91. Louadi Z., Elkjaer M.L., Klug M., Lio C.T., Fenn A., Illes Z., Bongiovanni D., Baumbach J., Kacprowski T., List M., Tsoy O. (2021) Functional enrichment of alternative splicing events with NEASE reveals insights into tissue identity and diseases. *Genome Biol.*, **22**, 327. DOI: 10.1186/s13059-021-02538-1
92. Yellaboina S., Tasneem A., Zaykin D.V., Raghavachari B., Jothi R. (2011) DOMINE: A comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.*, **39**(Database issue), D730–D735. DOI: 10.1093/nar/gkq1229
93. Mosca R., Céol A., Stein A., Olivella R., Aloy P. (2014) 3did: A catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.*, **42**(Database issue), D374–D379. DOI: 10.1093/nar/gkt887
94. Kumar M., Gouw M., Michael S., Sámano-Sánchez H., Pancsa R., Glavina J., Diakogianni A., Valverde J.A., Bukirova D., Čalyševa J., Palopoli N., Davey N.E., Chemes L.B., Gibson T.J. (2020) ELM — the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.*, **48**(D1), D296–D306. DOI: 10.1093/nar/gkz1030
95. Berman H.M. (2008) The protein data bank: A historical perspective. *Acta Crystallogr. A*, **64**(Pt 1), 88–95. DOI: 10.1107/S0108767307035623
96. Zhang Y., Yao X., Zhou H., Wu X., Tian J., Zeng J., Yan L., Duan C., Liu H., Li H., Chen K., Hu Z., Ye Z., Xu H. (2022) OncoSplicing: An updated database for clinically relevant alternative splicing in 33 human cancers. *Nucleic Acids Res.*, **50**(D1), D1340–D1347. DOI: 10.1093/nar/gkab851
97. Li Q., Lai H., Li Y., Chen B., Chen S., Li Y., Huang Z., Meng Z., Wang P., Hu Z., Huang S. (2021) RJunBase: A database of RNA splice junctions in human normal and cancerous tissues. *Nucleic Acids Res.*, **49**(D1), D201–D211. DOI: 10.1093/nar/gkaa1056
98. Ling J.P., Wilks C., Charles R., Leavey P.J., Ghosh D., Jiang L., Santiago C.P., Pang B., Venkataraman A., Clark B.S., Nellore A., Langmead B., Blackshaw S. (2020) ASCOT identifies key regulators of neuronal subtype-specific splicing. *Nat. Commun.*, **11**, 137. DOI: 10.1038/s41467-019-14020-5

99. Tian J., Wang Z., Mei S., Yang N., Yang Y., Ke J., Zhu Y., Gong Y., Zou D., Peng X., Wang X., Wan H., Zhong R., Chang J., Gong J., Han L., Miao X. (2019) CancerSplicingQTL: A database for genome-wide identification of splicing QTLs in human cancer. *Nucleic Acids Res.*, **47**(D1), D909–D916. DOI: 10.1093/nar/gky954
100. UniProt Consortium (2022) UniProt: The universal protein knowledgebase in 2023. *Nucleic Acids Res.*, **51**(D1), D523–D531. DOI: 10.1093/nar/gkac1052
101. Lane L., Argoud-Puy G., Britan A., Cusin I., Duek P.D., Evalet O., Gateau A., Gaudet P., Gleizes A., Masselot A., Zwahlen C., Bairoch A. (2012) NeXtProt: A knowledge platform for human proteins. *Nucleic Acids Res.*, **40**(Database issue), D76–D83. DOI: 10.1093/nar/gkr1179
102. O'Leary N.A., Wright M.W., Brister J.R., Ciufu S., Haddad D., McVeigh R., Rajput B., Robbertse B., Smith-White B., Ako-Adjei D., Astashyn A., Badredin A., Bao Y., Blinkova O., Brover V., Chetvernin V., Choi J., Cox E., Ermolaeva O., Farrell C.M., Goldfarb T., Gupta T., Haft D., Hatcher E., Hlavina W., Joardar V.S., Kodali V.K., Li W., Maglott D., Masterson P., McGarvey K.M., Murphy M.R., O'Neill K., Pujar S., Rangwala S.H., Rausch D., Riddick L.D., Schoch C., Shkeda A., Storz S.S., Sun H., Thibaud-Nissen F., Tolstoy I., Tully R.E., Vatsan A.R., Wallin C., Webb D., Wu W., Landrum M.J., Kimchi A., Tatusova T., di Cuccio M., Kitts P., Murphy T.D., Pruitt K.D. (2016) Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**(D1), D733–D745. DOI: 10.1093/nar/gkv1189
103. Frankish A., Diekhans M., Jungreis I., Lagarde J., Loveland J.E., Mudge J.M., Sisu C., Wright J.C., Armstrong J., Barnes I., Berry A., Bignell A., Boix C., Carbonell Sala S., Cunningham F., di Domenico T., Donaldson S., Fiddes I.T., García Girón C., Gonzalez J.M., Grego T., Hardy M., Hourlier T., Howe K.L., Hunt T., Izuogu O.G., Johnson R., Martin F.J., Martínez L., Mohanan S., Muir P., Navarro F.C.P., Parker A., Pei B., Pozo F., Riera F.C., Ruffier M., Schmitt B.M., Stapleton E., Suner M.M., Sycheva I., Uszczyńska-Ratajczak B., Wolf M.Y., Xu J., Yang Y.T., Yates A., Zerbino D., Zhang Y., Choudhary J.S., Gerstein M., Guigó R., Hubbard T.J.P., Kellis M., Paten B., Tress M.L., Flicek P. (2021) GENCODE 2021. *Nucleic Acids Res.*, **49**(D1), D916–D923. DOI: 10.1093/nar/gkaa1087
104. Morales J., Pujar S., Loveland J.E., Astashyn A., Bennett R., Berry A., Cox E., Davidson C., Ermolaeva O., Farrell C.M., Fatima R., Gil L., Goldfarb T., Gonzalez J.M., Haddad D., Hardy M., Hunt T., Jackson J., Joardar V.S., Kay M., Kodali V.K., McGarvey K.M., McMahon A., Mudge J.M., Murphy D.N., Murphy M.R., Rajput B., Rangwala S.H., Riddick L.D., Thibaud-Nissen F., Threadgold G., Vatsan A.R., Wallin C., Webb D., Flicek P., Birney E., Pruitt K.D., Frankish A., Cunningham F., Murphy T.D. (2022) A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature*, **604**(7905), 310–315. DOI: 10.1038/s41586-022-04558-8
105. Pertea M., Shumate A., Pertea G., Varabyou A., Breitwieser F.P., Chang Y.-C., Madugundu A.K., Pandey A., Salzberg S.L. (2018) CHES: A new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.*, **19**, 208. DOI: 10.1186/s13059-018-1590-2
106. Howe K.L., Achuthan P., Allen J., Allen J., Alvarez-Jarreta J., Amodé M.R., Armean I.M., Azov A.G., Bennett R., Bhai J., Billis K., Boddu S., Charkhchi M., Cummins C., da Rin Fioretto L., Davidson C., Dodiya K., El Houdaigui B., Fatima R., Gall A., Garcia Giron C., Grego T., Guijarro-Clarke C., Haggerty L., Hemrom A., Hourlier T., Izuogu O.G., Juettemann T., Kaikala V., Kay M., Lavidas I., Le T., Lemos D., Gonzalez Martinez J., Marugán J.C., Maurel T., McMahon A.C., Mohanan S., Moore B., Muffato M., Oheh D.N., Paraschas D., Parker A., Parton A., Prosovetskaia I., Sakthivel M.P., Salam A.I.A., Schmitt B.M., Schuilenburg H., Sheppard D., Steed E., Szpak M., Szuba M., Taylor K., Thormann A., Threadgold G., Walts B., Winterbottom A., Chakiachvili M., Chaubal A., de Silva N., Flint B., Frankish A., Hunt S.E., Ilesley G.R., Langridge N., Loveland J.E., Martin F.J., Mudge J.M., Morales J., Perry E., Ruffier M., Tate J., Thybert D., Trevanion S.J., Cunningham F., Yates A.D., Zerbino D.R., Flicek P. (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**(D1), D884–D891. DOI: 10.1093/nar/gkaa942
107. Lindblad-Toh K., Garber M., Zuk O., Lin M.F., Parker B.J., Washietl S., Kheradpour P., Ernst J., Jordan G., Mauceli E., Ward L.D., Lowe C.B., Holloway A.K., Clamp M., Gnerre S., Alföldi J., Beal K., Chang J., Clawson H., Cuff J., di Palma F., Fitzgerald S., Flicek P., Guttman M., Hubisz M.J., Jaffe D.B., Jungreis I., Kent W.J., Kostka D., Lara M., Martins A.L., Massingham T., Moltke I., Raney B.J., Rasmussen M.D., Robinson J., Stark A., Vilella A.J., Wen J., Xie X., Zody M.C., Broad Institute Sequencing Platform and Whole Genome Assembly Team, Baldwin J., Bloom T., Chin C.W., Heiman D., Nicol R., Nusbaum C., Young S., Wilkinson J., Worley K.C., Kovar C.L., Muzny D.M., Gibbs R.A., Baylor College of Medicine Human Genome Sequencing Center Sequencing Team, Cree A., Dihn H.H., Fowler G., Jhangiani S., Joshi V., Lee S., Lewis L.R., Nazareth L.V., Okwuonu G., Santibanez J., Warren W.C., Mardis E.R., Weinstock G.M., Wilson R.K., Genome Institute at Washington University, Delehaunty K., Dooling D., Fronik C., Fulton L., Fulton B., Graves T., Minx P., Sodergren E., Birney E., Margulies E.H., Herrero J., Green E.D., Haussler D., Siepel A., Goldman N., Pollard K.S., Pedersen J.S., Lander E.S., Kellis M. (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478**(7370), 476–482. DOI: 10.1038/nature10530
108. Sommer M.J., Cha S., Varabyou A., Rincon N., Park S., Minkin I., Pertea M., Steinegger M., Salzberg S.L. (2022) Structure-guided isoform identification for the human transcriptome. *eLife*, **11**, e82556. DOI: 10.7554/eLife.82556
109. Palazzo A.F., Lee E.S. (2015) Non-coding RNA: What is functional and what is junk? *Front. Genetics*, **6**, 2. DOI: 10.3389/fgene.2015.00002
110. Ponting C.P., Haerty W. (2022) Genome-wide analysis of human long noncoding RNAs: A provocative review. *Annu. Rev. Genomics Hum. Genet.*, **23**, 153–172. DOI: 10.1146/annurev-genom-112921-123710
111. Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G.M., Sherlock G. (2000) Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nat. Genet.*, **25**(1), 25–29. DOI: 10.1038/75556
112. Qiu S., Yu G., Lu X., Domeniconi C., Guo M. (2022) Isoform function prediction by Gene Ontology embedding. *Bioinformatics*, **38**(19), 4581–4588. DOI: 10.1093/bioinformatics/btac576

113. Eksi R., Li H.-D., Menon R., Wen Y., Omenn G.S., Kretzler M., Guan Y. (2013) Systematically differentiating functions for alternatively spliced isoforms through integrating RNA-seq data. *PLoS Comput. Biol.*, **9**(11), e1003314. DOI: 10.1371/journal.pcbi.1003314
114. Luo T., Zhang W., Qiu S., Yang Y., Yi D., Wang G., Ye J., Wang J. (2017) Functional annotation of human protein coding isoforms via non-convex multi-instance learning. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA.* pp. 345–354. DOI: 10.1145/3097983.3097984
115. Li H.-D., Yang C., Zhang Z., Yang M., Wu F.-X., Omenn G.S., Wang J. (2021) IsoResolve: Predicting splice isoform functions by integrating gene and isoform-level features with domain adaptation. *Bioinformatics*, **37**(4), 522–530. DOI: 10.1093/bioinformatics/btaa829
116. Shaw D., Chen H., Jiang T. (2019) DeepIsoFun: A deep domain adaptation approach to predict isoform functions. *Bioinformatics*, **35**(15), 2535–2544. DOI: 10.1093/bioinformatics/bty1017
117. Chen H., Shaw D., Zeng J., Bu D., Jiang T. (2019) DIFFUSE: Predicting isoform functions from sequences and expression profiles via deep learning. *Bioinformatics*, **35**(14), i284–i294. DOI: 10.1093/bioinformatics/btz367
118. Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., Tunyasuvunakool K., Bates R., Židek A., Potapenko A., Bridgland A., Meyer C., Kohl S.A.A., Ballard A.J., Cowie A., Romera-Paredes B., Nikolov S., Jain R., Adler J., Back T., Petersen S., Reiman D., Clancy E., Zielinski M., Steinegger M., Pacholska M., Berghammer T., Bodenstein S., Silver D., Vinyals O., Senior A.W., Kavukcuoglu K., Kohli P., Hassabis D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**(7873), 583–589. DOI: 10.1038/s41586-021-03819-2
119. Tunyasuvunakool K., Adler J., Wu Z., Green T., Zielinski M., Židek A., Bridgland A., Cowie A., Meyer C., Laydon A., Velankar S., Kleywegt G.J., Bateman A., Evans R., Pritzel A., Figurnov M., Ronneberger O., Bates R., Kohl S.A.A., Potapenko A., Ballard A.J., Romera-Paredes B., Nikolov S., Jain R., Clancy E., Reiman D., Petersen S., Senior A.W., Kavukcuoglu K., Birney E., Kohli P., Jumper J., Hassabis D. (2021) Highly accurate protein structure prediction for the human proteome. *Nature*, **596**(7873), 590–596. DOI: 10.1038/s41586-021-03828-1
120. Deiana A., Forcelloni S., Porrello A., Giansanti A. (2019) Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PloS ONE*, **14**(8), e0217889. DOI: 10.1371/journal.pone.0217889
121. Mirdita M., Schütze K., Moriwaki Y., Heo L., Ovchinnikov S., Steinegger M. (2022) ColabFold: Making protein folding accessible to all. *Nat. Methods*, **19**(6), 679–682. DOI: 10.1038/s41592-022-01488-1
122. Chang E., Fu C., Coon S.L., Alon S., Bozinoski M., Breymaier M., Bustos D.M., Clokie S.J., Gothilf Y., Esnault C., Michael Iuvone P., Mason C.E., Ochocinska M.J., Tovin A., Wang C., Xu P., Zhu J., Dale R., Klein D.C. (2020) Resource: A multi-species multi-timepoint transcriptome database and webpage for the pineal gland and retina. *J. Pineal Res.*, **69**(3), e12673. DOI: 10.1111/jpi.12673

Поступила в редакцию: 27. 04. 2024.
После доработки: 21. 06. 2024.
Принята к печати: 18. 07. 2024.

IN SILICO AND IN CELLULO APPROACHES FOR FUNCTIONAL ANNOTATION OF HUMAN PROTEIN SPLICE VARIANTS

O.I. Kiseleva, V.A. Arzumanyan, I.Yu. Kurbatov, E.V. Poverennaya*

¹Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; *e-mail: k.poverennaya@gmail.com

The elegance of pre-mRNA splicing mechanisms continues to interest scientists even after over a half century, since the discovery of the fact that coding regions in genes are interrupted by non-coding sequences. The vast majority of human genes have several mRNA variants, coding structurally and functionally different protein isoforms in a tissue-specific manner and with a linkage to specific developmental stages of the organism. Alteration of splicing patterns shifts the balance of functionally distinct proteins in living systems, distorts normal molecular pathways, and may trigger the onset and progression of various pathologies. Over the past two decades, numerous studies have been conducted in various life sciences disciplines to deepen our understanding of splicing mechanisms and the extent of their impact on the functioning of living systems. This review aims to summarize experimental and computational approaches used to elucidate the functions of splice variants of a single gene based on our experience accumulated in the laboratory of interactomics of proteoforms at the Institute of Biomedical Chemistry (IBMC) and best global practices.

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

Key words: functional annotation; alternative splicing; splice form; proteoforms; proteome heterogeneity; multiomics studies

Funding. This study was supported by the Russian Science Foundation (project no. 21-74-10061).

Received: 27.04.2024; revised: 21.06.2024; accepted: 18.07.2024.