

## ПРОТЕОМНЫЙ АНАЛИЗ ПЕРЕКОДИРОВАНИЯ БЕЛКОВ *Danio rerio* ВСЛЕДСТВИЕ РЕДАКТИРОВАНИЯ МАТРИЧНОЙ РНК АДЕНОЗИНДЕЗАМИНАЗАМИ ADAR

© 2022 Ш.Ш. Насаев<sup>1,2</sup>, А.С. Копейкина<sup>3</sup>, К.Г. Кузнецова<sup>4</sup>,  
Л.И. Левицкий<sup>5</sup>, С.А. Мошковский<sup>1,3\*</sup>

<sup>1</sup> Федеральный научно-клинический центр физико-химической медицины ФМБА России,  
119435 Москва, Россия; электронная почта: moshrrft@gmail.com

<sup>2</sup> НИИ биомедицинской химии имени В.Н. Ореховича, 119121 Москва, Россия

<sup>3</sup> Российский национальный исследовательский медицинский университет  
им. Н.И. Пирогова Минздрава России, 117997 Москва, Россия

<sup>4</sup> Бергенский университет, NO-5020 Берген, Норвегия

<sup>5</sup> Федеральный исследовательский центр химической физики им. Н.Н. Семенова РАН,  
Институт энергетических проблем химической физики им. В.Л. Тальрозе, 119334 Москва, Россия

Поступила в редакцию 20.09.2022

После доработки 08.10.2022

Принята к публикации 08.10.2022

Редактирование РНК аденозиндезаминазами семейства ADAR может вести к перекодированию белков, поскольку преобразованный из аденозина инозин в составе мРНК комплементарен цитозину. При редактировании кодонов в транскрибируемые белки вносятся аминокислотные замены. Перекодирование протеомов может иметь функциональные последствия, что описано для многих животных, включая человека. С использованием базы данных перекодирования, предсказанной по полученным из литературы транскриптомным данным, в представленной работе впервые идентифицировали участки перекодирования в панорамных протеомах *Danio rerio*. В шести наборах протеомных данных примерно из ста предсказанных событий перекодирования удалось идентифицировать десять. Семь из них принадлежали к хорошо изученным и консервативным среди позвоночных участкам перекодирования в субъединицах глутаматных рецепторов АМРА-типа. Три были специфичными для белков *D. rerio*. Среди них участвующие в адгезии и сигнализации нервных клеток трансмембранные рецепторы астротактин 1 и нейрегулин 3b, а также белок пресинаптической мембраны, участвующий в высвобождении нейромедиаторов – продукт гена *rims2b*. Дальнейшие исследования требуются для выявления функции перекодирования указанных белков у *D. rerio*.

**КЛЮЧЕВЫЕ СЛОВА:** протеогеномика, панорамная протеомика, *Danio rerio*, ADAR, РНК-зависимая аденозиндезаминаза, редактирование РНК.

**DOI:** 10.31857/S0320972522110070, **EDN:** LVQEEA

### ВВЕДЕНИЕ

Аденозиндезаминазы семейства ADAR представлены в большинстве многоклеточных животных. Ферментативно активные изоформы ADAR дезаминируют остатки аденозина рибонуклеиновых кислот в положении 6 пуринового гетероцикла, в результате чего они преобразуются в неканонические остатки инозина. Эта естественная посттранскрипционная модификация РНК широко распространена в клетках животных [1].

Особенность работы указанных ферментов заключается в их зависимости от участков двухцепочечной РНК, для чего в их структуре, помимо ферментативного, присутствует связывающий соответствующие участки домен. У грызунов и человека, как и у нескольких исследованных позвоночных других классов, активные формы ADAR делятся на 2 функциональные группы. У мыши и человека каждая из них включает по одному гену. Фермент ADAR1 (ген *ADAR* человека) редактирует протяженные участки регулярной двухцепочечной РНК, находящейся обычно в не кодирующей белки РНК или в некодирующей части мРНК.

\* Адресат для корреспонденции.

Фермент ADAR2 (ген *ADARB1* человека) работает с небольшими и нерегулярными дуплексами, часто в кодирующих участках матричных РНК [2].

Каковы же функции этих двух типов активности ADAR? Двухцепочечные нуклеиновые кислоты в цитоплазме, будь то ДНК или дуплексы РНК, даже в достаточно низких концентрациях активируют неспецифический иммунитет, поскольку повышение уровня таких молекул там может означать вирусную атаку. Активность ADAR в отношении протяженных двухцепочечных структур РНК обладает способностью нейтрализовать их. Известно, что инозин, в отличие от аденозина, комплементарен в большей степени цитидину, нежели уридину. Поэтому дуплексные структуры РНК разрушаются после редактирования ADAR. Одна из основных функций ADAR1 – нейтрализация в ядре протяженных дуплексов, а при иммунном ответе интерферонов I-типа на такие дуплексы, проникшие в цитоплазму, – подавление этого ответа по принципу отрицательной обратной связи. Не случайно интерфероны у млекопитающих индуцируют синтез альтернативной, более протяженной изоформы ADAR1, которая транспортируется в цитоплазму, оказывая там противоинтерфероновое действие [3].

В отличие от ADAR1, фермент ADAR2 действует более локально, осуществляя тонкую настройку сплайсинга матричных РНК и, в ряде случаев, изменяя аминокислотную последовательность кодируемых им продуктов. Действительно, изменение комплементарности оснований в РНК может изменять смысл кодона при соответствующем положении в последовательности. Остаток инозина ведет себя как гуанозин, в результате чего в кодируемых белках возникают единичные аминокислотные замены, и, в редких случаях, пропуск стоп-кодона. Таким образом, редактирование РНК ферментами ADAR, причем в подавляющем числе случаев у позвоночных – изоформами ADAR2, производит перекодирование протеома [3].

В разных организмах масштаб этого явления различен. Например, у головоногих моллюсков описан значительный уровень перекодирования транскрипта и протеома с несколькими сотнями выявляемых на уровне протеома сайтов перекодирования у разных видов [4]. Плодовая мушка, которая имеет всего один ген, кодирующий ADAR, ортологичный генам, кодирующим ADAR2 позвоночных, также характеризуется интенсивным редактированием РНК и перекодированием протеома. На протеомном уровне удается подтвердить

около 70 аминокислотных замен [5]. Иная ситуация у млекопитающих, например, у приматов и грызунов – у них из транскрипта предсказывается значительно меньшее количество событий перекодирования по сравнению с моллюсками и насекомыми, а на уровне протеома надежно выявляется всего около двух десятков замен [6].

Несмотря на это, мутации, приводящие к потере функции продуктов гена, кодирующего ADAR2, у млекопитающих, как правило, летальны. У трансгенных мышей обратить летальный фенотип при нокауте гена *ADARB1* (ADAR2) удавалось, изменив единственный сайт редактирования – в гене субъединицы 2 ионотропного глутаматного рецептора АМРА-типа (*GRIA2*). В процессе развития эмбриона происходит редактирование с перекодированием остатка 607 субъединицы 2 с глутамин на аргинин, при 100%-ном перекодировании у взрослых особей. В итоге снижается проводимость соответствующего ионного канала, что необходимо для нормального функционирования головного мозга. Если при нокауте гена, кодирующего фермент, осуществить перекодирование (*Q607R*) за счет искусственной геномной мутации, удается «спасти» мышей от летальной мутации [7].

На клеточных и мышечных моделях охарактеризовано функциональное значение других участков перекодирования за счет редактирования РНК. Например, аминокислотная замена за счет перекодирования в филамине А (*FLNA*) влияет на тонус гладких мышц сосудов [8]. Консервативная на первый взгляд замена I→V в альфа-коамере (*COPA*), как оказалось, имеет онкосупрессорный эффект и играет роль при гепатоцеллюлярной карциноме человека [9].

Оценка масштабов перекодирования протеома за счет редактирования РНК ферментами ADAR долгое время осуществлялась по транскриптомным данным. Вопрос в том, какое количество событий редактирования, способных перекодировать последовательность аминокислот, доходит до стадии транслированных белков. Многие события редактирования, как предполагается, являются побочным эффектом действия ферментов и не имеют биологического значения [10]. Очевидно, что те случаи редактирования РНК, последствия которых могут быть обнаружены на протеомном уровне, имеют больше шансов быть функционально значимыми. Поэтому после проведения полнотранскриптомного анализа редактирования РНК ферментами ADAR данные такого анализа нередко используются для поиска об-

разующейся в результате полиморфизма аминокислотной последовательности уже в протеомных данных. Поскольку основным подходом к идентификации последовательностей белка служит поиск в масс-спектрах соответствия предсказанным из генома или транскриптома белкам, для идентификации событий перекодирования нужно дополнить теоретический протеом этими измененными последовательностями [11].

В литературе представлено несколько проектов, в которых осуществляют анализ таких участков в протеоме. В значимой для этой области работе, где был впервые показан значительный масштаб переустройства транскриптома головоногих моллюсков за счет активности ADAR, изменения также подтверждали в соответствующих протеомах [4]. Позднее нами был проведен поиск перекодированных белков в протеомах плодовой мушки [5], причем изменения в профиле перекодирования регистрировали на разных стадиях метаморфоза этого насекомого [12]. Кроме того, нами идентифицированы перекодированные участки в белках головного мозга мыши и человека [6]. В пределах этих двух видов млекопитающих они отличаются консерватизмом, практически не пересекаясь с участками перекодирования в нервной системе дрозофилы. В дополнение полномасштабному анализу были подвергнуты протеомы различных злокачественных опухолей человека, где также были выявлены перекодированные сайты [13].

В дополнение к хорошо описанному РНК-эдитому головоногих моллюсков, дрозофилы, мыши и человека, редактирование ферментами ADAR было подробно описано в транскриптом известного модельного позвоночного – рыбки *Danio rerio* или зебрафиш [14]. Эта пресноводная рыбка отряда карпообразных, происходящая из тропической Азии, с 1960-х гг. используется в лаборатории для широкого круга биомедицинских исследований и играет большую роль в биологии развития, нейронауках и в исследовании широкого круга патологий, присущих всем позвоночным [15]. Мы использовали представленные в указанной работе транскриптомные данные для запуска отработанной нами серии методов. Целью исследования была надежная идентификация участков перекодирования белков в протеомах *D. rerio*. Представленные на белковом уровне сайты перекодирования ценны для дальнейших исследований, поскольку они с большей вероятностью обладают функциональной значимостью по сравнению с участками, представленными только на уровне транскриптома.

## МАТЕРИАЛЫ И МЕТОДЫ

**Создание базы данных для протеомного поиска с предсказанными участками перекодирования белков вследствие редактирования РНК.** Для анализа использовали референсный протеом *D. rerio* (инвентарный номер базы знаний Uniprot UP000000437\_7955 [16]). Из работы Buchumenski et al. [14] получали предсказанные по материалам РНК-секвенирования несинонимичные замены в экзонах мРНК транскриптома *D. rerio*. Из указанных замен формировали vcf-файл. Последний аннотировали с использованием скрипта snpEff [17] и сборки генома *D. rerio* в версии GRCz11.99 [18]. Таким образом, перекодированные остатки синхронизировали с записями в протеомной базе данных. Всего были предсказаны замены в 116 положениях белков *D. rerio*, из них 63 положения имели по 2 альтернативных варианта замены.

С помощью написанных нами программ на языке Python последовательности белков преобразовывали в последовательности трипсиновых пептидов с одним возможным пропущенным участком расщепления трипсином. Далее, белки с перекодированием по тому же принципу преобразовывали в пептиды. В итоговой базе данных к геномным пептидам добавляли пептиды с перекодированием. Таким образом, на начальной стадии 116 сайтов с альтернативными заменами формировали 471 пептид с учетом одного пропуска расщепления. Особенности масс-спектрометрического анализа препятствуют идентификации слишком длинных и слишком коротких пептидов, поэтому в исходном списке оставляли те, длина которых составляла от 7 до 40 аминокислотных остатков включительно. После данной операции количество пептидов с вариантами сократилось до 332 с потерей 139 несовместимых с анализом пептидов.

С помощью библиотеки Pyteomics [19] в базу данных для поиска добавляли фальшивые (decoy) последовательности пептидов, а именно: обращенные пептиды с сохранением C-концевого остатка.

**Наборы протеомных данных.** Для переобработки с целью поиска участков перекодирования в репозитории Proteomexchange [20] выбрали 6 наборов панорамных протеомных данных тканей *D. rerio* [21–25]. Подробности о наборах данных представлены в табл. 1.

**Протеомный поиск.** С помощью программы ThermoRawFileParser [26] исходные RAW-файлы преобразовались в mzML-файлы. В поисковой машине IdentiPy [27] для каждого из наборов данных устанавливались следующие

**Таблица 1.** Доступные общественности данные протеомного анализа тканей *Danio rerio*, использованные для поиска участков перекодирования белков вследствие редактирования РНК ферментами ADAR

Инвентарный номер репозитория ProteomeXchange [20]	Биоматериал для анализа	Детектор для получения MS/MS
PXD005630 [21]	синапсомы, постсинаптическое уплотнение выделены из мозга взрослых рыб	LTQ (линейная ионная ловушка)
PXD009612* [22]	целый организм, мальки на 6-е сутки после оплодотворения	Orbitrap
PXD014228 [23]	головной мозг взрослых рыб	Orbitrap
PXD016847 [24]	головной мозг взрослых рыб	LTQ
PXD023967***	моторные нейроны взрослых рыб	Orbitrap
PXD030733 [25]	головной мозг взрослых рыб	LTQ

Примечание. Все данные получены авторами путем хроматомасс-спектрометрии фрагментов расщепленных трипсином белков с использованием детекторов с ловушкой Orbitrap, по крайней мере для анализа родительских ионов в тандемном анализе. Спектры MS/MS, полученные в линейной ионной ловушке, обладают примерно на порядок более низкой точностью определения  $m/z$ , чем полученные в ловушке Orbitrap. Поиск проводили с добавлением в поисковую базу данных участков перекодирования, взятых из статьи Vuchumenski et al. [14].

\* Участков перекодирования белков, прошедших отбор по принципам, сформулированным в этой работе, в отобранных наборах данных не обнаружено.

\*\* Данные доступны в репозитории, но отсутствуют в публикациях.

настройки файла конфигурации, в соответствии с условиями получения масс-спектров в оригинальных работах:

- PXD023967 – product accuracy: 0,01 Da, fixed: camC, variable: oxM, acetyl-;
- PXD030733 – product accuracy: 0,5 Da, fixed: camC, variable: oxM, acetyl-;
- PXD009612 – product accuracy: 0,01 Da, fixed: camC, variable: oxM;
- PXD005630 – product accuracy: 0,5 Da, fixed: camC, variable: oxM, camM;
- PXD016847 – product accuracy: 0,5 Da, fixed: camC, variable: oxM, acetyl-;
- PXD014228 – product accuracy: 20 ppm, fixed: camC, variable: oxM.

С помощью IdentiPy в режиме автооптимизации параметров проводили идентификацию пептидов. С помощью программы Scavenger [28] производилась валидация и группоспецифическая фильтрация перекодированных пептидов (FDR – 1%).

**Визуализация масс-спектров.** Результаты поиска визуализировали в программе просмотра масс-спектров xiSPEC [29]. Изображения подвергали экспертной оценке: в аннотированных масс-спектрах фрагментации определяли покрытие предполагаемого сайта перекодирования.

**Аннотация участков перекодирования с точки зрения пространственной структуры белков.** Информацию получали, комбинируя общедоступные ресурсы. Аннотировали перекодированные

белки по материалам баз знаний UniProt [16] и neXtProt [30]. Неохарактеризованные структурно белки *D. rerio* выравнивали на белки млекопитающих посредством Blastp [31]. Трехмерные структуры анализировали посредством программы для визуализации молекулярных структур PyMOL [32]. Влияние на стабильность структуры белков оценивали с помощью I-Mutant2.0 [33], MUpPro [34] и iStable [35].

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

**Наборы данных и фильтрация идентифицированных участков перекодирования.** В отличие от данных по человеку и мыши, в свободном доступе находится не так много протеомов *D. rerio*. Кроме того, для оптимальной работы протеоеномной серии методов, использованных в этой и других наших работах [6], подходят данные, отвечающие определенным техническим требованиям. Так, для повторной обработки применяли данные панорамного протеомного анализа без использования метки, полученные на приборах с ловушкой Orbitrap (табл. 1). Из-за небольшого количества доступных протеомных данных использовали как данные, где тандемные спектры получали в детекторе с высоким разрешением, как это делается в большинстве работ последнего времени, так и данные более ранних приборов (Orbitrap Velos или Elite) с тандемными спек-

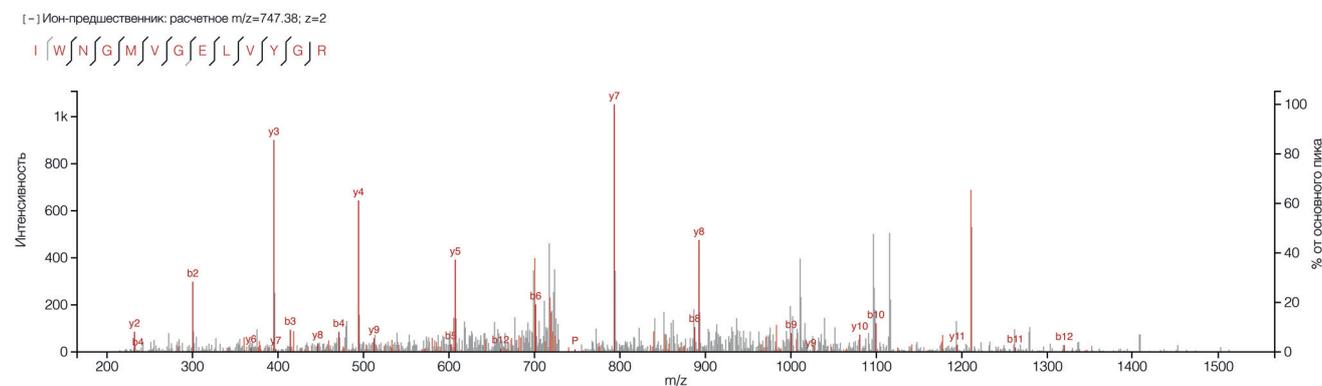
трами, полученными в линейной ионной ловушке (LTQ), совмещенной с Orbitrap. Спектры MS/MS последнего типа обладают менее высокой точностью, что отражается при задании параметров протеомного поиска.

В связи с тем, что *D. rerio* представляет собой предпочтительный объект для нейронаук, большая часть наборов данных, выбранных для исследования, представляла собой препараты головного мозга, в том числе разные фракции синапсов [21], а также культуру моторных нейронов. Кроме того, в исследование были включены данные одного исследования по протеому целых мальков рыбки в возрасте 6 суток [22] (табл. 1). С точки зрения существующих знаний о редактировании РНК, центральная нервная система – удачный объект, поскольку у позвоночных, как известно, большая часть событий редактирования с перекодированием белков происходит именно в соответствующих тканях, где высока экспрессия гена изоформы ADAR2 [3].

Наличие в составе данных масс-спектров с более низким разрешением MS/MS заставила нас особенно внимательно относиться к критериям фильтрации идентификаций перекодированных пептидов. К критериям включения отнесли идентификацию как минимум в двух наборах данных, идентификацию в двух пептидах разного состава, образовавшихся за счет неполного расщепления трипсином, или подтверждение идентификации путем экспертного просмотра масс-спектров, сформированных в программе для визуализации xiSPEC [29]. Кроме того, исключению подлежали пептиды с недорасщеплением трипсином без подтверждения полностью расщепленными пептидами. Недавно мы показали, что такие идентификации имеют повышенный уровень ложноположительных результатов (FDR) [36].

Большие интервалы, в которые могут попасть нужные значения  $m/z$  на tandemных масс-спектрах с более низким разрешением, делают сложным экспертный анализ визуализированных масс-спектров. Обычно такие масс-спектры, полученные путем хромато-масс-спектрометрии сложных смесей, содержат много пиков, нерелевантных целевому пептиду. Однако в ряде случаев интересующие нас пики выделяются по своей интенсивности. На рис. 1 проиллюстрировано убедительное доказательство идентичности масс-спектра и пептида субъединицы глутаматного рецептора *gria4b*. Из рис. 1 видно, что наиболее интенсивные пики в масс-спектре ассоциированы с молекулярными массами, предсказанными теоретически для MS/MS кодируемого в геноме пептида с заменой K→R. Остальные масс-спектры, идентичность которых с перекодированными пептидами подтверждали путем экспертной оценки, представлены на рис. S1 в Приложении.

**Участки перекодирования, обнаруженные в протеомах *D. rerio*.** В четырех из шести наборов данных, исследованных в данной работе, обнаружили перекодированные участки белков. Все протеомы с обнаруженным перекодированием относились к центральной нервной системе. Всего таких участков оказалось 10 (в одном из них обнаружены 2 варианта перекодирования, см. табл. 2) из 116 предсказанных из транскриптома [14], что даже меньше, чем было обнаружено в центральной нервной системе мыши и человека. В этих протеомах надежно идентифицировали 14 и 18 сайтов перекодирования соответственно [6]. Впрочем, база данных для поиска, состоящая из предсказанных по результатам секвенирования РНК участков перекодирования, в случае *D. rerio* была намного меньше. Кроме того, данные по млекопитающим были значительно объемнее по суммарному количеству полученных масс-спектров.



**Рис. 1.** Tandemный масс-спектр пептида субъединицы глутаматного рецептора *gria4b*, визуализированный в программе XiSpec [29]. Наиболее интенсивные пики в масс-спектре ассоциированы с молекулярными массами, предсказанными теоретически для MS/MS кодируемого в геноме пептида с заменой K→R

Таблица 2. Участки перекодирования белков *Danio rerio* в результате редактирования РНК ферментами ADAR, идентифицированные в данных панорамной протеомики

Название гена	Название белка	Аминокислотная замена при перекодировании	Идентифицированный пептид	Результат экспертного анализа масс-спектра	Данные, в которых найден участок (подробнее см. табл. 1)	Идентификация альтернативного участка, кодируемого в геноме
Участки перекодирования в субъединицах глутаматных рецепторов						
<i>gria2a</i> и/или <i>gria3b</i>	субъединица глутаматного рецептора (flor-изоформа)	R756G и/или R764G	GSAL[G]NAVNLA VLK, GYGVATPKGSAL[G]NAVNLA VLK	+	PXD014228 (мозг); PXD005630 (синапсомы, постсинаптическое уплотнение); PXD030733 (мозг)	+
<i>gria2a</i> и/или <i>gria3b</i>	субъединица глутаматного рецептора (flor-изоформа)	R756G и/или R764G	GSAL[G]TPVNLAVLK	+	PXD005630 (синапсомы, постсинаптическое уплотнение)	+
<i>gria2b</i>	субъединица глутаматного рецептора (flor-изоформа)	R760G	GSSL[G]TPVNLAVLK	+	PXD005630 (синапсомы, постсинаптическое уплотнение)	+
<i>gria2b</i>	субъединица глутаматного рецептора (flor-изоформа)	R760G	GSSL[G]NAVNLA VLK	+	PXD014228 (мозг)	+
<i>gria4a</i>	субъединица глутаматного рецептора (flor-изоформа)	R760G	GSQ[L]G]TPVNLAVLK	+	PXD014228 (мозг); PXD005630 (синапсомы, постсинаптическое уплотнение)	+
<i>gria4b</i>	субъединица глутаматного рецептора (flor-изоформа)	R766G	GSQ[L]G]SAVNLA VLK	+	PXD005630 (синапсомы, постсинаптическое уплотнение)	-
<i>gria4b</i>	субъединица глутаматного рецептора	K493R	IWNGMVGEIYV[G]R]	+	PXD005630 (постсинаптическое уплотнение)	+
Участки перекодирования в прочих белках						
<i>astn1</i>	астрогактин 1	K935E	TLTYPAYIASLLDSGA[E]RMAAGVR	-	PXD005630 (синапсомы)	+
<i>astn1</i>	астрогактин 1	K935R	TLTYPAYIASLLDSGA[R]	-	PXD005630 (синапсомы, постсинаптическое уплотнение)	+
<i>nrg3b</i>	нейрегулин 3b	T10A	TAIGA[A] METMTLEEPGGEQASPRAPGPLR, MSERTALGA[A] METMTLEEPGGEQASPR	-	PXD005630 (синапсомы, постсинаптическое уплотнение); PXD030733 (мозг)	-
<i>rim2b</i>	регулятор экзоцитоза синаптической мембраны 2b	S489G	L[G]STYSMSVQSER, SSSDSDVSDVSAMSSASRL[G] STYSMSVQSER	+	PXD016847 (мозг); PXD005630 (постсинаптическое уплотнение)	-

Как указывалось выше, классический и наиболее хорошо изученный пример перекодирования белков млекопитающих — замены в субъединицах глутаматных рецепторов NMDA-типа, среди которых замена Q→R влияет на проводимость составляемых ими ионных каналов и является необходимой для развития и существования нервной системы организма. При этом существующие подходы к получению панорамных протеомов не дают возможности надежно зарегистрировать эту замену. Однако в генах глутаматных рецепторов указанного типа есть другие участки редактирования, которые хорошо выявлялись в протеомах млекопитающих [6].

Гомологичные белкам млекопитающих, местами идентичные по последовательности субъединицы рецепторов *D. rerio* также оказались перекодированными (7 из 10 обнаруженных участков перекодирования). В списке обнаруженных перекодированных сайтов имеются участки флип- и флоп-изоформ альтернативного сплайсинга [37] продуктов генов *gria2a*, *gria2b*, *gria4a* и *gria4b* (табл. 2). Такие изоформы хорошо известны у разных позвоночных животных [37]. К обнаруженным заменам R→G, характерным для 760–766 положений парало-гов, в зависимости от конкретного гена, которые также хорошо детектировались у млекопитающих, добавилась замена K493R в продукте гена *gria4b* (масс-спектр представлен на рис. 1). Гомологичные сайты перекодирования не были выявлены в других организмах.

Что касается других участков перекодирования, наблюдаемых у млекопитающих, то у *D. rerio* в основном они не наблюдаются и на уровне транскриптома [14]. Исключением является замена Q→G в продукте гена *cadpsa*, наблюдаемая также в эдитомах и протеомах мыши и человека. Тем не менее в белках рыбки мы этой замены не нашли.

В эдитомах *D. rerio* ранее обнаружили много других кодирующих замен, отсутствовавших у млекопитающих [14]. Мы смогли надежно идентифицировать 3 таких участка в трех белках соответственно. Примечательно, что все 3 участка находились в тех частях молекул, которые не удалось разрешить с помощью систем предсказания пространственной структуры белковых молекул, таких как AlphaFold 2 [38] и Swiss-model [39], ввиду отсутствия достаточно сходства с экспериментально определенными структурами гомологов.

Астротактин 1 (*astn1*) характеризовался заменами лизина на аргинин или глутаминовую кислоту в положении 935. Этот трансмембранный рецептор представляет собой регулятор

адгезии нейронов, участвующий в миграции незрелых нейробластов во время развития головного мозга [40]. В составе внеклеточного участка рецептора имеются домены, подобные эпидермальному фактору роста, а также фибронектиновый домен. Участок, в котором содержится перекодированный остаток, располагается во внеклеточной части белка за пределами указанных доменов между двумя спиралями, предположительно, в участке, необходимом для взаимодействия с лигандами. На это косвенно указывает близкое расположение участков гликозилирования. Перекодированный остаток лизина экспонирован, не скрыт другими структурами. Интересно, что этот остаток более или менее консервативен и также находится в ортологах млекопитающих (чего тем не менее недостаточно для предсказания фолдинга этой части белка), но у них перекодированию не подлежит. Оценка влияния замены на стабильность с помощью трех использованных программ дала противоречивые, разнонаправленные результаты. Можно предположить, что замена лизина незначительно влияет на конформацию белка, однако может менять сродство взаимодействия с потенциальными лигандами.

Нейрегулин 3b (*neu3b*) — еще один нейрональный трансмембранный белок, участвующий в онтогенезе нервной ткани. Нейрегулин 3 также характеризуется некоторыми свойствами эпидермального фактора роста: у млекопитающих его ортолог связывает один из EGF-рецепторов [41]. Перекодирование T→A обнаружено в положении 10 в самом N-конце этой молекулы. Как это часто бывает, N-концевая часть молекулы нейрегулина характеризуется неструктурированностью. Она экспонирована во внеклеточное пространство. Два использованных метода предсказания влияния замены на структуру, MUPro [34] и i-Stable [35], указывали на уменьшение стабильности белка после замены. I-Mutant [33] указывал на отсутствие разницы в стабильности белка до и после замены. Наблюдаемая замена, по-видимому, как и в случае астротактина, существенно не влияет на пространственную структуру молекулы, но может воздействовать на межбелковые взаимодействия. В отличие от предыдущего случая перекодирования в астротактине, N-конец нейрегулина 3 не консервативен и отличается по последовательности ортологов млекопитающих.

Если астротактин 1 и нейрегулин 3 обладают некоторой общностью функций, обеспечивая адгезию и другие рецепторные взаимодействия развивающихся нейронов, то функции третьего перекодированного белка,

не относящегося к глутаматным рецепторам, отличаются. Продукт гена *rims2b* локализован вблизи пресинаптической мембраны, ассоциирован с пресинаптическими везикулами и, по-видимому, вовлечен в кальций-зависимый экзоцитоз, т.е. в работу синапса. Замена S489G находится в неконсервативном участке этого белка, отсутствующем у млекопитающих, поэтому информация по его трехмерной структуре отсутствует. Три использованных метода предсказания влияния замен на структуру белков указывали на высокую вероятность снижения стабильности белка после замены.

Интересно, что белки пресинаптической мембраны – участники того же каскада экзоцитоза, включая ортолог упомянутого выше продукта гена *cadpsa*, интенсивно перекодируются в нервной системе плодовой мушки [12]. Есть предположение, что редактирование РНК в этом пойкилотермном организме вносит вклад в адаптацию нервной системы к работе при изменяющихся температурах [42]. Гипотетически можно предполагать нечто подобное у пойкилотермного позвоночного *D. rerio*, хотя для проверки этой гипотезы требуются объемные эксперименты с трансгенными моделями *in vivo*.

### ЗАКЛЮЧЕНИЕ

Редактирование РНК ферментами ADAR – эволюционно древний клеточный процесс, присущий большинству эуметазоев. В процессе эволюции дезаминирование аденозинов, которое осуществляют ферменты, функционально разделилось по двум направлениям. Во-первых, это инактивация потенциально иммуногенных дуплексов РНК, а во-вторых – перекодирование некоторых белков с внесением аминокислотных замен в последовательность. Значимость этого перекодирования на первый взгляд вызывает удивление, ведь функциональные замены аминокислот могли бы сформироваться в процессе мутаций на геномном уровне, а для модулирования функции белка существует большое число посттрансляционных модификаций. Однако подобный способ модификации последовательности и функции некоторых белков закрепился в процессе эволюции и используется, например, для внесения регуляции по ходу онтогенеза.

Мы систематически исследуем перекодирование белков на уровне протеома в разных организмах. Всего 10 сайтов перекодирования

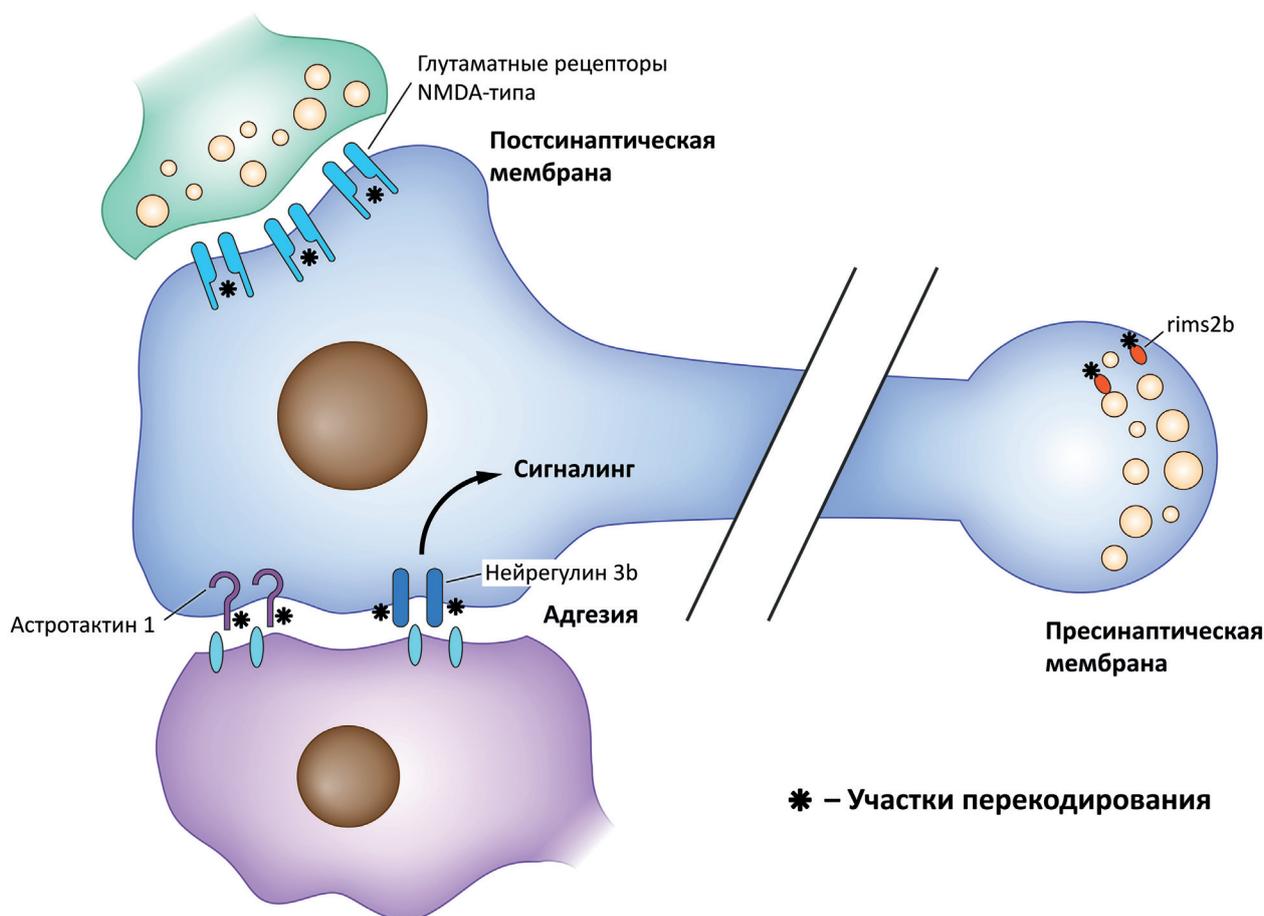


Рис. 2. Предполагаемая локализация перекодированных белков в нейроне *D. rerio*

обнаружены при анализе 6 наборов данных по модельной рыбке *D. rerio*. Отметим, что, судя по литературным данным, перекодирование мРНК в этом организме выражено несколько менее масштабно, чем у млекопитающих. Большая часть перекодирования обнаружена в субъединицах глутаматных рецепторов. Следует полагать, что перекодирование глутаматных рецепторов за счет редактирования РНК консервативно у позвоночных. В трех белках выявлены участки перекодирования, характерные, по имеющимся данным, только у *D. rerio*. Несмотря на то, что эти 3 аминокислотные замены вряд ли существенно меняют структуру белка, они могут быть функциональными, влияя на межмолекулярные взаимодействия. Свежим примером может быть незначимая на первый взгляд замена I→V в альфа-коомере человека, модулирующая злокачественный потенциал рака печени [9]. Интересно, что все 3 специфичные для рыбы аминокислотные замены вследствие перекодирования оказались в неконсервативных в отношении остальных позвоночных частях белков. Вероятно, такое специфичное для таксона (впрочем, неизвестно, насколько консервативное, например, среди костных рыб) перекодирование более совместимо с эволюционно молодыми, изменяющимися частями белков. Отметим, что обоснованные суждения об эволюции в контексте редактирования РНК и связанного с ним перекодирования, все же следует оставить для транскриптомных работ, где у авторов больше возможностей сравнивать разные организмы [43]. Предполагаемое расположение перекодированных белков в нейроне *D. rerio* проиллюстрировано на рис. 2.

Продолжением данной работы может быть экспериментальная проверка гипотез о функциональной значимости найденных в белках адгезии и синаптической мембраны участков редактирования. Кроме того, для данного мо-

дельного объекта представлено мало объемных протеомных данных по сравнению, например, с данными по мышам и человеку, и в этом направлении также можно продолжить работу. К настоящему моменту наиболее результативным, с точки зрения редактирования РНК, был набор данных от выделенных компонентов синапса, к сожалению, полученный с низким разрешением тандемных спектров [21]. Получение более качественных протеомных данных по разным клеткам и тканям *D. rerio* может расширить наше представление о перекодировании белков рыбы за счет редактирования РНК ферментами ADAR.

**Вклад авторов.** Ш.Ш. Насаев, А.С. Копейкина – составление базы данных участков перекодирования для протеомного поиска, проведение поиска, анализ пространственной структуры, оформление результатов; К.Г. Кузнецова, Л.И. Левицкий – настройка серии поисковых методов повторной обработки данных, написание статьи; С.А. Мошковский – идея работы, экспертный анализ результатов, написание статьи.

**Благодарности.** Авторы благодарят Центр коллективного пользования Федерального научно-клинического центра физико-химической медицины за предоставление вычислительных ресурсов для данного исследования.

**Финансирование.** Работа выполнена при финансовой поддержке Российского научного фонда (грант № 20-15-00072).

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

**Соблюдение этических норм.** Настоящая статья не содержит описания каких-либо исследований с участием людей или животных в качестве объектов.

**Дополнительные материалы.** Приложение к статье опубликовано на сайте журнала «Биохимия» (<https://biochemistrymoscow.com>).

## СПИСОК ЛИТЕРАТУРЫ

- Goncharov, A. O., Shender, V. O., Kuznetsova, K. G., Kliuchnikova, A. A., and Moshkovskii, S. A. (2022) Interplay between A-to-I editing and splicing of RNA: A potential point of application for cancer therapy, *Int. J. Mol. Sci.*, **23**, 5240, doi: 10.3390/ijms23095240.
- Luciano, D. J., Mirsky, H., Vendetti, N. J., and Maas, S. (2004) RNA editing of a miRNA precursor, *RNA*, **10**, 1174-1177, doi: 10.1261/rna.7350304.
- Goncharov, A. O., Kliuchnikova, A. A., Nasaev, S. S., and Moshkovskii, S. A. (2019) RNA editing by ADAR adenosine deaminases: from molecular plasticity of neural proteins to the mechanisms of human cancer, *Biochemistry (Moscow)*, **84**, 896-904, doi: 10.1134/S0006297919080054.
- Liscovitch-Brauer, N., Alon, S., Porath, H. T., Elstein, B., Unger, R., et al. (2017) Trade-off between transcriptome plasticity and genome evolution in cephalopods, *Cell*, **169**, 191-202.e111, doi: 10.1016/j.cell.2017.03.025.
- Kuznetsova, K. G., Kliuchnikova, A. A., Pina, I. U., Chernobrovkin, A. L., Novikova, S. E., et al. (2018) Proteogenomics of adenosine-to-inosine RNA editing

- in the fruit fly, *J. Proteome Res.*, **17**, 3889-3903, doi: 10.1021/acs.jproteome.8b00553.
6. Levitsky, L. I., Kliuchnikova, A. A., Kuznetsova, K. G., Karpov, D. S., Ivanov, M. V., et al. (2019) Adenosine-to-inosine RNA editing in mouse and human brain proteomes, *Proteomics*, **19**, e1900195, doi: 10.1002/pmic.201900195.
  7. Higuchi, M., Maas, S., Single, F. N., Hartner, J., Rozov, A., et al. (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2, *Nature*, **406**, 78-81, doi: 10.1038/35017558.
  8. Jain, M., Weber, A., Maly, K., Manjaly, G., Deek, J., et al. (2022) A-to-I RNA editing of Filamin A regulates cellular adhesion, migration and mechanical properties, *FEBS J.*, **289**, 4580-4601, doi: 10.1111/febs.16391.
  9. Song, Y., An, O., Ren, X., Chan, T. H. M., Tay, D. J. T., et al. (2021) RNA editing mediates the functional switch of COPA in a novel mechanism of hepatocarcinogenesis, *J. Hepatol.*, **74**, 135-147, doi: 10.1016/j.jhep.2020.07.021.
  10. Buchumenski, I., Roth, S. H., Kopel, E., Katsman, E., Feiglin, A., et al. (2021) Global quantification exposes abundant low-level off-target activity by base editors, *Genome Res.*, **31**, 2354-2361, doi: 10.1101/gr.275770.121.
  11. Moshkovskii, S. A., Ivanov, M. V., Kuznetsova, K. G., and Gorshkov, M. V. (2018) Identification of single amino acid substitutions in proteogenomics, *Biochemistry (Moscow)*, **83**, 250-258, doi: 10.1134/S0006297918030057.
  12. Kliuchnikova, A. A., Goncharov, A. O., Levitsky, L. I., Pyatnitskiy, M. A., Novikova, S. E., et al. (2020) Proteome-wide analysis of ADAR-mediated messenger RNA editing during fruit fly ontogeny, *J. Proteome Res.*, **19**, 4046-4060, doi: 10.1021/acs.jproteome.0c00347.
  13. Peng, X., Xu, X., Wang, Y., Hawke, D. H., Yu, S., et al. (2018) A-to-I RNA editing contributes to proteomic diversity in cancer, *Cancer Cell*, **33**, 817-828 e817, doi: 10.1016/j.ccell.2018.03.026.
  14. Buchumenski, I., Holler, K., Appelbaum, L., Eisenberg, E., Junker, J. P., et al. (2021) Systematic identification of A-to-I RNA editing in zebrafish development and adult organs, *Nucleic Acids Res.*, **49**, 4325-4337, doi: 10.1093/nar/gkab247.
  15. Ribas, L., and Piferrer, F. (2014) The zebrafish (*Danio rerio*) as a model organism, with emphasis on applications for finfish aquaculture research, *Rev. Aquaculture*, **6**, 209-240, doi: 10.1111/raq.12041.
  16. UniProt, C. (2021) UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Res.*, **49**, D480-D489, doi: 10.1093/nar/gkaa1100.
  17. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3, *Fly (Austin)*, **6**, 80-92, doi: 10.4161/fly.19695.
  18. Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., et al. (2013) The zebrafish reference genome sequence and its relationship to the human genome, *Nature*, **496**, 498-503, doi: 10.1038/nature12111.
  19. Goloborodko, A. A., Levitsky, L. I., Ivanov, M. V., and Gorshkov, M. V. (2013) Pyteomics – a Python framework for exploratory data analysis and rapid software prototyping in proteomics, *J. Am. Soc. Mass Spectrom.*, **24**, 301-304, doi: 10.1007/s13361-012-0516-6.
  20. Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., et al. (2020) The ProteomeXchange consortium in 2020: enabling “big data” approaches in proteomics, *Nucleic Acids Res.*, **48**, D1145-D1152, doi: 10.1093/nar/gkz984.
  21. Bayes, A., Collins, M. O., Reig-Viader, R., Gou, G., Goulding, D., et al. (2017) Evolution of complexity in the zebrafish synapse proteome, *Nat. Commun.*, **8**, 14613, doi: 10.1038/ncomms14613.
  22. Watchon, M., Luu, L., Robinson, K. J., Yuan, K. C., De Luca, A., et al. (2021) Sodium valproate increases activity of the sirtuin pathway resulting in beneficial effects for spinocerebellar ataxia-3 *in vivo*, *Mol. Brain*, **14**, 128, doi: 10.1186/s13041-021-00839-x.
  23. Faria, M., Prats, E., Gomez-Canela, C., Hsu, C. Y., Arick, M. A., 2nd, et al. (2019) Therapeutic potential of N-acetylcysteine in acrylamide acute neurotoxicity in adult zebrafish, *Sci. Rep.*, **9**, 16467, doi: 10.1038/s41598-019-53154-w.
  24. Nonnis, S., Angiulli, E., Maffioli, E., Frabetti, F., Negri, A., et al. (2021) Acute environmental temperature variation affects brain protein expression, anxiety and explorative behaviour in adult zebrafish, *Sci. Rep.*, **11**, 2521, doi: 10.1038/s41598-021-81804-5.
  25. Maffioli, E., Angiulli, E., Nonnis, S., Grassi Scalvini, F., Negri, A., et al. (2022) Brain proteome and behavioural analysis in wild type, BDNF<sup>+/-</sup> and BDNF<sup>-/-</sup> adult Zebrafish (*Danio rerio*) exposed to two different temperatures, *Int. J. Mol. Sci.*, **23**, 5606, doi: 10.3390/ijms23105606.
  26. Hulstaert, N., Shofstahl, J., Sachsenberg, T., Walzer, M., Barsnes, H., et al. (2020) ThermoRawFileParser: modular, scalable, and cross-platform RAW file conversion, *J. Proteome Res.*, **19**, 537-542, doi: 10.1021/acs.jproteome.9b00328.
  27. Levitsky, L. I., Ivanov, M. V., Lobas, A. A., Bubis, J. A., Tarasova, I. A., et al. (2018) IdentiPy: an extensible search engine for protein identification in shotgun proteomics, *J. Proteome Res.*, **17**, 2249-2255, doi: 10.1021/acs.jproteome.7b00640.
  28. Ivanov, M. V., Levitsky, L. I., Bubis, J. A., and Gorshkov, M. V. (2019) Scavenger: a versatile postsearch validation algorithm for shotgun proteomics based on gradient boosting, *Proteomics*, **19**, e1800280, doi: 10.1002/pmic.201800280.
  29. Kolbowski, L., Combe, C., and Rappsilber, J. (2018) xiSPEC: web-based visualization, analysis and

- sharing of proteomics data, *Nucleic Acids Res.*, **46**, W473-W478, doi: 10.1093/nar/gky353.
30. Lane, L., Argoud-Puy, G., Britan, A., Cusin, I., Duek, P. D., et al. (2012) neXtProt: a knowledge platform for human proteins, *Nucleic Acids Res.*, **40**, D76-83, doi: 10.1093/nar/gkr1179.
  31. Madden, T. L., Tatusov, R. L., and Zhang, J. (1996) Applications of network BLAST server, *Methods Enzymol.*, **266**, 131-141, doi: 10.1016/s0076-6879(96)66011-x.
  32. Schrödinger, L. L. C., and DeLano, W. *PyMOL*, URL: <https://www.pymol.org/pymol>.
  33. Capriotti, E., Fariselli, P., and Casadio, R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res.*, **33**, W306-310, doi: 10.1093/nar/gki375.
  34. Cheng, J., Randall, A., and Baldi, P. (2006) Prediction of protein stability changes for single-site mutations using support vector machines, *Proteins*, **62**, 1125-1132, doi: 10.1002/prot.20810.
  35. Chen, C. W., Lin, J., and Chu, Y. W. (2013) iStable: off-the-shelf predictor integration for predicting protein stability changes, *BMC Bioinformatics*, **14** Suppl 2, S5, doi: 10.1186/1471-2105-14-S2-S5.
  36. Levitsky, L. I., Kuznetsova, K. G., Kliuchnikova, A. A., Ilina, I. Y., Goncharov, A. O., et al. (2022) Validating amino acid variants in proteogenomics using sequence coverage by multiple reads, *J. Proteome Res.*, **21**, 1438-1448, doi: 10.1021/acs.jproteome.2c00033.
  37. Pei, W., Huang, Z., and Niu, L. (2007) GluR3 flip and flop: differences in channel opening kinetics, *Biochemistry*, **46**, 2027-2036, doi: 10.1021/bi062213s.
  38. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., et al. (2021) Highly accurate protein structure prediction with AlphaFold, *Nature*, **596**, 583-589, doi: 10.1038/s41586-021-03819-2.
  39. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Res.*, **46**, W296-W303, doi: 10.1093/nar/gky427.
  40. Adams, N. C., Tomoda, T., Cooper, M., Dietz, G., and Hatten, M. E. (2002) Mice that lack astrotactin have slowed neuronal migration, *Development*, **129**, 965-972, doi: 10.1242/dev.129.4.965.
  41. Miyagi, T., Wada, T., Yamaguchi, K., Hata, K., and Shiozaki, K. (2008) Plasma membrane-associated sialidase as a crucial regulator of transmembrane signalling, *J. Biochem.*, **144**, 279-285, doi: 10.1093/jb/mvn089.
  42. Rieder, L. E., Savva, Y. A., Reyna, M. A., Chang, Y. J., Dorsky, J. S., et al. (2015) Dynamic response of RNA editing to temperature in *Drosophila*, *BMC Biol.*, **13**, 1, doi: 10.1186/s12915-014-0111-3.
  43. Moldovan, M., Chervontseva, Z., Bazykin, G., and Gelfand, M. S. (2020) Adaptive evolution at mRNA editing sites in soft-bodied cephalopods, *PeerJ*, **8**, e10456, doi: 10.7717/peerj.10456.

## PROTEOMIC ANALYSIS OF ZEBRAFISH PROTEIN RECODING VIA MESSENGER RNA EDITING BY ADAR ADENOSINE DEAMINASES

S. S. Nasaev<sup>1,2</sup>, A. S. Kopeykina<sup>3</sup>, K. G. Kuznetsova<sup>4</sup>, L. I. Levitsky<sup>5</sup>, and S. A. Moshkovskii<sup>1,3\*</sup>

<sup>1</sup> Federal Research and Clinical Center of Physical-Chemical Medicine, 119435 Moscow, Russia; e-mail: moshrfi@gmail.com

<sup>2</sup> Institute of Biomedical Chemistry, 119121 Moscow, Russia

<sup>3</sup> Pirogov Russian National Research Medical University, 117997 Moscow, Russia

<sup>4</sup> University of Bergen, NO-5020 Bergen, Norway

<sup>5</sup> V. L. Talrose Institute for Energy Problems of Chemical Physics,

N.N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, 119334 Moscow, Russia

RNA editing by adenosine deaminases of ADAR family may lead to protein recoding, since inosine converted from adenosine in mRNA is complementary to cytosine. Editing of codons introduces amino acid substitutions into translated proteins. Proteome recoding may have functional consequences, which are described in many animals including humans. Using protein recoding database derived from publicly available transcriptome data, recoding sites were identified for the first time in shotgun proteomes of zebrafish. Out of more than a hundred predicted recoding events, ten substitutions were identified in six datasets. Seven of them belonged to recoding sites in AMPA-type glutamate receptor subunits whose recoding is well-described and conservative among vertebrates. Three sites were specific for zebrafish proteins. Among them, there are transmembrane receptors astrotactin 1 and neuregulin 3b involved in neuronal adhesion and signalling, as well as *rim52b* gene product, a presynaptic membrane protein, which participated in neurotransmitter release. Further studies are needed to elucidate function of recoding said three proteins in zebrafish.

**Keywords:** proteogenomics, shotgun proteomics, zebrafish, ADAR, RNA-dependent adenosine deaminase, RNA editing