

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ
«НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ БИОМЕДИЦИНСКОЙ ХИМИИ ИМЕНИ
В.Н. ОРЕХОВИЧА»

На правах рукописи

КЛЮЧНИКОВА АННА АЛЕКСЕЕВНА

ПЕРЕКОДИРОВАНИЕ БЕЛКОВ В ЦЕНТРАЛЬНОЙ НЕРВНОЙ
СИСТЕМЕ МОДЕЛЬНЫХ ОРГАНИЗМОВ И ЧЕЛОВЕКА
ВСЛЕДСТВИЕ РЕДАКТИРОВАНИЯ МАТРИЧНОЙ РНК
АДЕНОЗИНДЕЗАМИНАЗАМИ

1.5.4 – биохимия

Диссертация

на соискание ученой степени кандидата биологических наук

Научный руководитель:

д.б.н., проф. РАН

Мошковский Сергей Александрович

Москва 2021

ОГЛАВЛЕНИЕ

СЛОВАРЬ ТЕРМИНОВ И СОКРАЩЕНИЙ	4
1. ВВЕДЕНИЕ, ЦЕЛЬ И ЗАДАЧИ	6
2. ОБЗОР ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ	9
2.1. Ферменты ADAR – древний механизм инактивации двухцепочечной РНК.....	9
2.2. Связь активности ферментов ADAR с иммунитетом.....	13
2.2. Редактирование РНК при злокачественных опухолях.....	17
2.3. Ферменты ADAR и процессы в центральной нервной системе	18
2.4. Способы идентификации событий редактирования РНК	21
2.5. Протеогеномный подход для поиска последствий редактирования РНК на белковом уровне.....	22
3. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ	24
3.1. Исходные данные, полученные из открытого доступа.....	24
3.2. Характеристика образцов.....	25
3.2.1. Культура <i>Drosophila melanogaster</i> . Возрастная модель	25
3.2.2. Характеристика образцов млекопитающих	26
3.3. Выделение мозга <i>Drosophila melanogaster</i>	26
3.4. Пробоподготовка к протеомному анализу	27
3.4.1. Измерение концентрации белков с использованием ВСА	27
3.4.2. Гидролитическое расщепление белков в растворе	28
3.5. Панорамный протеомный анализ	29
3.6. Синтез пептидных стандартов	29
3.7. Мониторинг множественных реакций.....	31
3.8. Секвенирование ДНК.....	33
3.9. Метод химического удаления инозина.....	34
3.10. Составление видоспецифических баз данных редактирования	36
3.11. Протеогеномный поиск	37
3.12. Белок-белковые взаимодействия и визуализация данных.....	40
4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ	41
4.1. Поиск сайтов перекодирования в протеоме плодовой мушки	41

4.1.1. Открытый поиск для дополнительной фильтрации ложно-положительных результатов	45
4.1.2. Функциональная аннотация отредактированных белков, обнаруженных в данных панорамной протеомики.....	45
4.1.3. Генотипирование сайтов геномной ДНК, соответствующих найденным событиям редактирования РНК.....	47
4.1.4. Идентификация событий редактирования на уровне мРНК	47
4.1.5. Проверка и количественная оценка отредактированных сайтов на уровне белка с помощью мониторинга множественных реакций (MRM) с использованием меченых стабильными изотопами стандартов пептидов.....	48
4.1.6. Анализ протеома плодовой мушки на разных стадиях онтогенеза для выявления участков белка, перекодированных при редактировании ферментами ADAR	54
4.1.7. Распределение перекодированных белков по стадиям жизненного цикла.....	61
4.1.8. Перекодированные с помощью ADAR участки в транскриптомных и протеомных данных жизненного цикла плодовой мушки	64
4.1.9. Количественное определение специфичных для мозга белков, перекодированных редактированием РНК, на разных стадиях жизненного цикла плодовой мушки	65
4.2. Поиск сайтов редактирования РНК в протеоме мыши и человека.....	69
4.2.1. Анализ находящихся в открытом доступе результатов транскриптомных и протеомных данных для биоинформатического поиска редактирования РНК.....	69
4.2.2. ADAR-опосредованное редактирование мРНК в протеоме мышинного мозга.....	71
4.2.3. ADAR-опосредованное редактирование мРНК в протеоме мозга человека	79
4.2.4. Анализ редактирования РНК на уровне протеома головного мозга трансгенных мышей - моделей заболеваний ЦНС	82
4.2.5. Анализ редактирования РНК на уровне протеома цереброспинальной жидкости	85
5. ЗАКЛЮЧЕНИЕ	85
6. ВЫВОДЫ	88
7. СПИСОК ЛИТЕРАТУРЫ	89
8. БЛАГОДАРНОСТИ	105
9. ФИНАНСИРОВАНИЕ	105
10. ПРИЛОЖЕНИЕ.....	106

СЛОВАРЬ ТЕРМИНОВ И СОКРАЩЕНИЙ

4M-pip - 4-метилпиперидин (от англ. 4-methylpiperidine).

ADAR - РНК-зависимая аденозиндезаминаза, код классификации ферментов EC: 3.5.4.37 (от англ. AdenosineDeAminase, RNA-dependent).

BCA - бицинхониновая кислота (от англ. bicinhoninic acid).

BSA - бычий сывороточный альбумин (от англ. bovine serum albumin).

DMF - N, N-диметилформамид (от англ. dimethylformamide)

DTT - дитиотреитол (от англ. dithiothreitol).

Fmoc – флуоренилметилоксикарбонил (от англ. fluorenylmethyloxycarbonyl).

Gene ontology, или GO – крупный биоинформатический проект по представлению генов и генных продуктов для всех видов живых организмов, цели которого состоят в поддержании и развитии словаря генов и генных продуктов; аннотации генов и генных продуктов, ассимиляции и диссимиляции аннотированных данных, а также в обеспечении инструментами для облегчения восприятия всех аспектов представленных данных.

HCTU - 1H-Бензотриазолиум 1-[бис(диметиламино)метилен]-5хлоро-гексафлуорофосфат (1-),3-оксид (от англ. 1H-Benzotriazolium 1-[bis(dimethyl-amino)methylene]-5-chloro-hexafluorophosphate (1-),3-oxide).

LC-MS/MS - метод жидкостной хроматографии с tandemной масс-спектрометрией.

LFQint. - интенсивность сигнала при количественном анализе без использования изотопной метки (от англ. intensity-based label-free quantification) – метод количественного анализа в масс-спектрометрии, который позволяет определить относительное количество белков при сравнении двух или более образцов. Метод основывается на измерении и сравнении интенсивностей пиков прекурсорных ионов. Его особенность заключается в количественном анализе белков без использования стабильных изотопов в качестве меток.

MRM - мониторинг множественных реакций (от англ. multiple reaction monitoring).

NSAF - индекс нормированных спектральных коэффициентов обилия (от англ. normalized spectral abundance factors).

PSM – спектральная идентификация - соответствие пары теоретический пептид — настоящий спектр (от англ. peptide-spectrum match).

Tris - трис-(гидроксиметил)-аминометан (от англ. tris-(hydroxymethyl)-aminomethane, THAM).

SNP - однонуклеотидный полиморфизм (от англ. single nucleotide polymorphism).

TMP - 2,4,6-триметилпиридин (от англ. trimethylpyridine)

дцРНК – молекула двухцепочечной рибонуклеиновой кислоты.

Перекодированный пептид – то же, что и отредактированный пептид – пептид, возникший в результате редактирования РНК аденозиндезаминазами ADAR.

Эдитом – совокупность событий перекодирования, в данной работе – ADAR-опосредованных.

1. ВВЕДЕНИЕ, ЦЕЛЬ И ЗАДАЧИ

1.1. Актуальность проблемы, цель и задачи

Протеогеномика является новой областью молекулярного анализа. Она основана на интеграции результатов геномных, транскриптомных и протеомных данных [1]. В частности, она направлена на выявление и количественную оценку вариации первичной структуры белка, кодируемой геномом, и увеличение диапазона результатов панорамной протеомики для более точной идентификации белковых последовательностей. В настоящее время изучение белковых вариантов методами протеогеномики в основном используется при исследовании злокачественных опухолей, что позволяет, вслед за геномом злокачественной опухоли, точнее охарактеризовать и ее протеом [2]. В дополнение к изменениям в геноме, одним из источников вариантов аминокислотной последовательности является редактирование РНК - не до конца изученное явление посттранскрипционной модификации РНК. Редактирование матричной РНК может приводить к изменению структуры белков, в результате чего синтезируется укороченный белок или белок с другой аминокислотной последовательностью. Этот процесс подразумевает химическую модификацию оснований, точнее, сайт-специфическое дезаминирование с преобразованием аденозина в инозин или цитозина в урацил [3]. Наиболее распространенный вид такого явления – преобразование остатка аденозина с получением инозина, возникающая посредством активности ферментов семейства РНК-зависимых аденозиндезаминаз (ADAR, код классификации ферментов EC: 3.5.4.37). Некоторые события редактирования РНК, приводящие к аминокислотным заменам в белках, как установлено, имеют функциональную значимость у многоклеточных организмов с точки зрения функционирования молекулярных механизмов в их клетках [4].

Представленная работа была направлена на изучение последствий редактирования РНК ферментами ADAR на протеомном уровне с использованием протеогеномного подхода. Методами хромато-масс-спектрометрии выявлены белки центральной нервной системы плодовой мушки, мыши и человека, последовательность которых перекодирована указанным путем. Исследование включало применение находящихся в открытом доступе результатов геномных и протеомных данных для биоинформатического поиска редактирования РНК и получение протеомов выбранных клеток и тканей модельных организмов для подтверждения результатов этого поиска.

Цель данной работы - идентификация перекодирования белковых последовательностей, возникших в результате редактирования РНК с дезаминированием аденозина, в масштабе протеомов клеток и тканей центральной нервной системы плодовой мушки, мыши и человека.

Для достижения указанной цели были поставлены следующие **задачи**:

1. Поиск перекодированных редактированием РНК участков в протеомах плодовой мушки с использованием протеомных и транскриптомных данных, как находящихся в открытом доступе, так и собственных.
2. Валидация и количественный анализ событий перекодирования с помощью таргетного протеомного анализа головного мозга плодовой мушки на разных стадиях онтогенеза.
3. Поиск и сравнительный анализ участков перекодирования в масштабах протеомов клеток и тканей центральной нервной системы мыши и человека с использованием находящихся в открытом доступе данных.
4. Валидация и количественный анализ событий перекодирования с помощью таргетного протеомного анализа на уровне протеома головного мозга трансгенных мышей – моделей заболеваний ЦНС.

Научная новизна работы

В этом проекте впервые были использованы панорамные протеомы головного мозга дрозофилы, мыши и человека для выявления в этих организмах перекодирования белков вследствие дезаминирования остатков аденозина в кодонах мРНК ферментами семейства ADAR. Для *Drosophila melanogaster* были проверены и количественно оценены отобранные перекодированные сайты с помощью таргетной протеомики в тканях мозга на разных стадиях метаморфоза. На дрозофиле показано, что некоторые белки с известной функцией, например, эндофилин А, в мозге насекомого перекодированы в значительной степени. Также в работе впервые выявлены последствия ADAR-опосредованного редактирования мРНК в масштабе протеомов мыши и человека. Хотя исследуемый тип редактирования РНК реализуется почти во всех тканях организма млекопитающего, только десятки перекодированных участков обнаруживаются на уровне протеома. В этом исследовании найдено тридцать восемь потенциальных перекодированных сайтов в центральной нервной системе человека и двадцать у мыши в протеомах с наибольшим количеством идентификаций, доступных на сегодняшний день. Часть из них валидированы в головном мозге экспериментальных мышей методами таргетной протеомики.

Теоретическая и практическая значимость работы

Показано, что редактирование РНК потенциально образует тысячи аминокислотных замен, но до уровня детектирования самого чувствительного протеомного анализа доходят немногие - 1% у дрозофилы примерно из 8 тысяч сайтов, обнаруженных путем анализа РНК, 2-3% у млекопитающих из 1-2 тыс. таких сайтов. Мы показали, что консерватизм в участках

редактирования белков между насекомым и млекопитающими отсутствует. У дрозофилы оно затрагивает белки пресинаптической мембраны, в частности, участвующие в созревании и высвобождении мембранных везикул с нейромедиатором. Кроме того, редактируются некоторые белки цитоскелета и ионные каналы. У мыши и человека, наоборот, воспроизводимо редактируются одни и те же сайты, часть из которых уже хорошо известна из биохимических исследований, не связанных с протеомикой. Это глутаматные рецепторы АМРА-типа, белки цитоскелета и корецепторы (продукты генов *FLNA* и *CYFIP2*), белок везикул комплекса Гольджи альфа-коатомер, а также единственный продукт, который редактируется и у дрозофилы - участник экзоцитоза секреторных везикул *CADPS*. В данной работе мы демонстрируем подход и инструменты, которые могут использоваться для анализа роли редактирования РНК в функциональных моделях патологических процессов, а также для использования перекодированных протеоформ в качестве биомаркеров различных патологий нервной системы.

Личный вклад автора

Работа была выполнена в лаборатории медицинской протеомики ИБМХ в период с 2016 по 2020 год. Соискатель проводил поиск и анализ литературы, планирование экспериментов, пробоподготовку к анализу нуклеиновых кислот и протеомный анализ данных, интерпретацию результатов и подготовку публикаций. Автор благодарит коллег за проведение масс-спектрометрического анализа, синтез изотопно меченных пептидных стандартов, подготовку программного обеспечения для обработки данных и работу с трансгенными мышами.

Основные положения, выносимые на защиту

1. Протеогеномный подход позволяет идентифицировать десятки участков перекодирования белков вследствие редактирования РНК аденозиндезаминазами, в том числе и ранее описанные случаи перекодирования белковой последовательности, например, в субъединицах ионотропных глутаматных рецепторов млекопитающих.
2. Впервые идентифицированы события перекодирования вследствие редактирования РНК в центральной нервной системе плодовой мушки, мыши и человека. Часть из обнаруженных участков перекодирования подтверждены таргетной протеомикой.
3. Перекодирование за счет редактирования РНК у насекомых и млекопитающих затрагивает в основном компоненты цитоскелета и белки, участвующие в синаптической передаче.
4. Редактирование РНК потенциально образует тысячи аминокислотных замен, но до уровня детектирования с помощью протеомного анализа доходят немногие, например, у млекопитающих 2-3% из 800-1800 таких сайтов.

Степень достоверности и апробация результатов

Достоверность полученных результатов обеспечена использованием методов обработки и оценки данных, соответствующих современным научным критериям. Обсуждение результатов проведено с учетом современных исследований, опубликованных в области белковой биохимии, протеомики и протеогеномики.

Положения и выводы, изложенные в диссертации, подтверждены публикациями и следуют из результатов исследований, проведенных диссертантом. Основные положения работы были опубликованы в рецензируемых научных изданиях, а также представлены в виде постерных докладов на конгрессе международной организации «Протеом человека» (HUPO 2017, Дублин), конференции «Клиническая протеомика. Постгеномная медицина» (ClinProt 2017, Москва), конгрессе Федерации Европейских биохимических обществ (FEBS 2018, Прага) и конгрессе Европейской протеомной ассоциации (EuPA 2019, Потсдам). Устные сообщения с соавторством диссертанта представлялись на 7-м и 9-м съезде Всероссийского масс-спектрометрического общества (Москва, 2015 и 2019) и международном конгрессе «Биотехнология: состояние и перспективы развития. Науки о жизни» (Москва, 2019).

По теме диссертационной работы опубликовано 13 работ, из которых 6 статей представлены в рецензируемых научных журналах и 7 публикаций находятся в трудах конференций.

2. ОБЗОР ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ

2.1. Ферменты ADAR – древний механизм инактивации двухцепочечной РНК

В мире клеток, особенно эукариотических, большую роль играют вирусы за счет постоянного проникновения своими геномами в цитоплазму. Они одновременно могут как уничтожить клетку, так и действовать как эволюционный фактор, так как их встраивание в геном может иметь функциональное значение. Некоторые из таких встроенных элементов кодируют конкретные белки, например, обратную транскриптазу, так как являются прообразом гена обратной транскриптазы, некогда вирусной. Участки, появившиеся в геноме благодаря вирусам, в некоторых условиях характеризуются экспрессией, а иногда образуют функциональные, в том числе кодирующие белки гены. По некоторым предположениям, подобные вирусные вторжения являются инструментом эволюции многоклеточных организмов [5].

Исходная роль активности ферментов класса ADAR, согласно теории, принятой во многих исследованиях, заключается в их участии в механизмах неспецифического иммунитета [6,7]. В

ответ на вирусные инфекции в цитоплазме выработались инструменты, детектирующие свободную нуклеиновую кислоту. При возникновении в цитоплазме свободных двухцепочечных рибонуклеиновых кислот в высокой концентрации, эукариотическая клетка сразу же на это реагирует. Например, интерфероновый ответ у млекопитающих, помимо иммунного ответа, сопровождается подавлением трансляции. Это препятствует размножению вирусов. Последние при этом стараются ускользнуть от систем защиты. Например, в клетках млекопитающих избыток дцРНК любого происхождения вызывает специфический ответ, инициированный сенсорами дцРНК, за которым следует активация каскада интерферона типа I. К таким сенсорам относятся, например RIG I-подобные (в частности, MDA5) и Толл-подобные (например, TLR 3). Когда вирусная РНК появляется в клетке, ферменты ADAR могут смягчать эти ответы, действуя как отрицательная обратная связь и избегая гиперактивации неспецифического иммунитета против собственной дцРНК [8,9]. У дрозофилы действуют специальные методы РНК-интерференции, так как у нее нет интерфероновой реакции [10]. Считается, что для устранения избыточного ответа на двухцепочечные структуры РНК, которые должны в каком-то количестве присутствовать в норме, и образовался механизм отрицательной обратной связи, в который вовлечены аденозиндезаминазы – ферменты редактирования РНК [8,9].

Редактирование РНК является одним из природных посттранскрипционных механизмов изменения химической структуры нуклеотидов, как правило, единичных. Наиболее распространенный вид такого процесса – замена аденозина на инозин, возникающая посредством ферментов семейства РНК-зависимых аденозиндезаминаз (ADAR, код классификации ферментов EC: 3.5.4.37 [3]). Отличительной особенностью ферментативных реакций, катализируемых белками этого семейства, является то, что они дезаминируют остатки аденозина во вторичных структурах двухцепочечной РНК (дцРНК) [11]. В результате самой замены происходит замещение аминогруппы на кетонную. Полученные остатки инозина, в отличие от исходных аденозинов, менее аффинны к остаткам уридина и предпочитают связывать цитидины посредством водородной связи. Таким образом, нарушается вторичная структура участков дцРНК. Концентрация дцРНК снижается, ее клеточные сенсоры снижают сигнал. Повышение концентрации такой структуры РНК в цитоплазме в итоге индуцирует синтез интерферонов I типа – альфа и бета, которые, действуя аутокринно и паракринно, вызывают противовирусный ответ [12]. Описано, что у млекопитающих интерфероновый ответ запускает альтернативный сплайсинг РНК гена *ADAR*, запуская синтез цитоплазматической изоформы фермента ADAR1 [13]. Предполагается, что эта индукция предназначена для отрицательной обратной связи. ADAR1 снижает концентрацию дцРНК и снижает действие ее сенсоров, тем самым угнетая интерфероновую сигнализацию. Таким образом, работа фермента может оказывать провирусное

действие. С другой стороны у аденозиндезаминаз описана также противовирусная активность, поскольку при расплетании двухцепочечных участков РНК в вирусе их геномы могут инактивироваться [14]. К известным вирусам, патогенным для человека, у которых геном представлен дцРНК, относятся, например, ротавирусы [15].

Как известно, в мРНК тоже есть двухцепочечные участки, которые могут подвергаться действию аденозиндезаминаз. Было показано, что нацеливание ферментов ADAR на дцРНК в основном зависит от ее вторичной структуры и от контекста последовательности, по крайней мере, для ADAR1 человека [16]. Если фермент нацелен на экзоны, он дезаминирует аденозины в кодирующей последовательности. Аденозиндезаминазы находятся в ядре, модифицируя там созревающую РНК, а в цитоплазму выходит преимущественно индуцированная изоформа ADAR. Поскольку образующиеся остатки инозина комплементарны цитидину, а не уридину, в кодируемых белках могут происходить одиночные аминокислотные замены. Соответственно, из генетического кода предсказан конкретный набор аминокислот, которые могут перекодироваться. Иногда ADAR влияет на сплайсинг. Если модифицировать аденозин в 3'-сайте сплайсинга, то он может произойти другим путем. Сайты сплайсинга, подвергнутые редактированию РНК, которое, как показано, происходит до сплайсинга мРНК, способны обеспечить образование новых сплайс-изоформ [17].

Гипотетически можно предположить, что эти связанные с протеомом последствия изначально представляли собой просто неблагоприятные эффекты дезактивации дцРНК, часто неадаптивные и вредные. Напротив, те несинонимичные сайты редактирования, которые, как обнаруживается, редактируются с высокой частотой, существуют в рамках четко определенного положительного отбора. Таким образом, естественный отбор по-разному действует на кодирующие и не кодирующие участки РНК [18]. Стоит также отметить, что редактирование кодирующих сайтов мРНК во многих случаях имеет важное значение, поскольку полученные протеоформы приобретают особую роль в течение жизни животного [19].

Ферменты ADAR имеют общую модульную организацию доменов: один или несколько двухцепочечных РНК-связывающих доменов и каталитический дезаминазный домен [20]. Помимо этого, ADAR1 также имеет два Z-ДНК-связывающих домена - $Z\alpha$ и $Z\beta$. Домен $Z\alpha$ направляет ADAR1 к транскрибируемой ДНК для обеспечения редактирования перед событием сплайсинга. Домен $Z\beta$ не способен связываться с Z-ДНК и его функция пока неясна.

У насекомых, например, плодовой мушки, присутствует только одна каталитически активная форма фермента редактирования, в то время как для млекопитающих характерны две. У человека известно два гена ADAR (ген *adar*/белок ADAR1, ген *adarb1*/белок ADAR2),

обладающих ферментативной активностью в отношении двухцепочечной РНК, а также по крайней мере два ADAR-подобных гена (ген *ADARB2*/белок ADAR3, ген *ADAD*/белок TENR). В масштабной работе по тканям мыши и человека [21] установлено, что ADAR1 в большей степени продуцируется конститутивно и в основном отвечает за редактирование некодирующих участков генома, в том числе, повторов. ADAR2 в большей степени отвечает за редактирование уникальных последовательностей, включая кодирующие. У плодовых мушек единственный ген, *Adar*, кодирует дезаминазу, аналогичную ADAR2 млекопитающих [22]. Однако недавнее исследование, сфокусированное на функциональных характеристиках мутировавших штаммов *Drosophila melanogaster*, показало, что фермент плодовой мушки разделяет функциональность обеих изоформ человека, регулируя иммунные ответы дцРНК и редактируя транскрипты мРНК [23]. Как правило, у плодовых мушек события редактирования встречаются чаще, чем у людей и мышей, что было показано как на уровне транскриптома [24,25], так и на уровне протеома [26,27]. Частично этот факт можно объяснить необходимостью глубоких морфологических и физиологических изменений насекомого при его метаморфозе. У *Drosophila* редактирование РНК, как сообщается, способствует поддержанию компактности генома, что важно из-за необходимости метаморфоза. Варианты белков, перекодированные посредством редактирования, обеспечивают измененные функции, необходимые на новых этапах жизни насекомого [28].

Функция продуктов ADAR-подобных генов в большей степени связана с ингибированием редактирования. ADAR3 служит ингибитором реакций ферментов редактирования и представлен в большей степени в центральной нервной системе. Так, на больных глиобластомой показано, что ADAR3 конкурирует с ADAR2 за связывание с транскриптом гена глутаматного рецептора GRIA2, ингибируя при этом редактирование мРНК [29]. Второй ADAR-подобный ген экспрессируется в клетках мужской зародышевой линии. В литературных источниках описывается участие TENR в специфичных для тестикул посттранскрипционных процессах, таких как альтернативный сплайсинг, упаковка гетерогенной ядерной РНК или транспорт мРНК [30]. При связывании с субстратами TENR выступает в процессе редактирования в качестве ингибитора [31]. Кроме того, обнаружен белок, который, по всей видимости, представляет собой ключевой отрицательный регулятор редактирования РНК ферментами ADAR в целом – AIMP2 [21].

Функциональная роль перекодирования в белках вследствие редактирования выяснена для небольшой группы хорошо охарактеризованных сайтов. Например, только один кодирующий сайт, возникший в результате редактирования, с заменой глутамина на аргинин в субъединице GRIA2 глутаматного AMPA-рецептора необходим для выживания мышей. Нокаут

гена *adarb1*, кодирующего мышиный ADAR2, приводит к возникновению неотредактированных форм глутаматных рецепторов и вызывает повышенное возбуждение в головном мозге в виде судорожных припадков, во взрослом возрасте несовместимых с жизнью. Внесение этого полиморфизма путем геномных мутаций в сочетании с инактивацией ADAR приводит к нормальному выживанию трансгенных мышей без очевидной патологии [32].

Некоторые важные функции приписываются несинонимичному редактированию РНК филамина А в сосудистой сети мышей [33] и коамера альфа при опухолевых заболеваниях человека [34]. Многие из известных несинонимичных сайтов редактирования РНК, например, у плодовой мушки, до сих пор функционально не охарактеризованы.

События редактирования, приводящие к перекодированию, в основном изучались в транскриптах [35], и лишь в нескольких работах использовались данные масс-спектрометрии для идентификации аминокислотных замен, вызванных активностью ADAR на транскриптомном уровне, в частности, при раковых заболеваниях человека [34].

Ранее сообщалось, что редактирование транскриптов РНК является дифференцированным в онтогенезе различных организмов, от насекомых до млекопитающих. Какие факторы могут регулировать это дифференциальное редактирование? Обычно ферментативно управляемые процессы зависят от экспрессии фермента. Однако, это не было подтверждено для ADAR у людей [36]. Более того, было обнаружено, что некоторые из мишень-специфичных РНК-связывающих белков модулируют ферментативную активность одной изоформы млекопитающих, ADAR2, которая, как полагают, отвечает за дифференциальное редактирование мишеней мРНК [37]. Например, SRSF9 представляет собой фактор сплайсинга РНК, который, как показано, ингибирует редактирование мишени мРНК CYFIP2 в клетках человека [37].

2.2. Связь активности ферментов ADAR с иммунитетом

Как было сказано ранее, ADAR дезаминирует двухцепочечную РНК, содержащую аденозиновые остатки. Накоплены разрозненные сведения о том, что ADAR1, в отличие от ADAR2, играет роль в иммунных реакциях в ответ на заражение РНК-содержащими вирусами.

ADAR-опосредованное редактирование чаще всего встречается в транскриптах генов приматов. У человека выявлено большое количество сайтов редактирования приблизительно в 1600 различных генах [38,39]. Такой уровень редактирования обусловлен тем, что большинство сайтов находятся в расположенных рядом *Alu*-повторах. Эти повторы, встроенные в геном в обратной ориентации, после транскрипции приводят к образованию двухцепочечных структур

РНК [39–41]. У человека около 10% генома представлено данными структурами - это свыше миллиона *Alu*-повторов длиной 300 пар нуклеотидов [42]. Как известно, такие короткие диспергированные повторы (SINE) являются отпечатками древних ретровирусов, закрепившихся в геноме. Инсерция транспозонов в новые геномные сайты может быть одной из движущих сил появления новых мишеней ADAR. Необходимы дальнейшие исследования для выяснения регулирующей функции этого явления.

Протяженные двухцепочечные участки РНК (дцРНК) вызывают настороженность у клеточных механизмов, поскольку именно такая РНК содержится в геноме многих вирусов. Выброс интерферона I типа и противовирусный ответ возникает как раз при накоплении таких участков и их действии, к примеру, на RIG-I и Toll-подобные рецепторы. Посредством активации протеинкиназы R (PKR), в том числе за счет взаимодействия с двухцепочечной РНК, ингибируется трансляция, что может препятствовать размножению вируса [43]. Показано, что интерферон 1 типа индуцирует выработку длинного сплайс-варианта ADAR1 p150 [13]. Такая изоформа фермента активно дезаминирует двухцепочечную РНК, способствуя разворачиванию молекулы. Этот механизм противодействует интерфероновому ответу по принципу отрицательной обратной связи [20]. Активность фермента подавляет образование стрессовых гранул против вируса кори и препятствует ответу клеток на данный вирус. В эту картину вписывается недавнее исследование [44], в котором найдено, что ADAR1 в виде сплайс-варианта p150 как раз ингибирует образование стресс-гранулы и стрессовый ответ клетки на собственную двухцепочечную РНК. Это было показано на мышинных фибробластах, где индуцированный ADAR1 редактировал не только повторы, но и экзонные области. Как предполагается, ADAR1 уравнивает потенциально повреждающий эффект интерферонов в отсутствие инфекции. Подтверждает данную теорию и то, что мыши с нокаутом гена ADAR1 умирают в процессе эмбрионального развития из-за обширного апоптоза и увеличенного уровня интерферонов [20]. Кроме того, ADAR1 может взаимодействовать с PKR неферментативным путем, ингибируя таким образом активность этой протеинкиназы [45].

Количество мРНК с участками редактирования увеличивается в Т-лимфоцитах и макрофагах, стимулированных такими медиаторами воспаления, как фактор некроза опухоли- α и γ -интерферон [46]. Важно также отметить, что ADAR1 может взаимодействовать с семейством белков ядерного фактора 90 (NF90), влияя тем самым на опосредованный NF90 механизм регуляции генов [47]. Подавление пути активации интерферона через взаимодействие ADAR1 с NF90 важно для защиты гемопоэтических клеток-предшественников от апоптоза [48].

Таким образом, активность ADAR1 способствует защите клеток от избыточного воспаления, опосредованного интерферонами I типа (табл.1). Этот фермент не дает развиваться

вышеуказанным реакциям при отсутствии вирусных РНК, когда RIG-I-подобные рецепторы или PKR могут реагировать на эндогенные молекулы [9,49,50].

Миссенс-мутации, приводящие к аминокислотным заменам в ADAR1, у человека уменьшают активность редактирования РНК и вызывают синдром Айкарди-Гутьерес, представляющий собой прогрессирующую энцефало- и интерферопатию, которые имитируют вирусную инфекцию [51]. Так, на примере данного синдрома, можно понять, что интерфероновый ответ усиливается из-за отсутствия аденозиндезаминазы, его уравнивающей.

Таблица 1. Функции ферментов дезаминирования ADAR и их связь с заболеваниями

Характеристика	ADAR1	ADAR2	Ссылка
Субстратная специфичность	Протяженные участки дцРНК	Участки дцРНК, принципы избирательности неясны	[20]
Иммунная функция	Подавление клеточного иммунитета за счет снижения концентрации дцРНК Индукция интерферонами I типа за счет экспрессии длинного сплайс-варианта p150 с альтернативного промотора	Нет воздействия	[3,9]
Редактирование, приводящее к изменению белковой последовательности	Участвует в жизненном цикле вируса гепатита D. В неотредактированной форме способствует репликации вируса. В отредактированной амбер-кодон заменяется на кодон триптофана, репликацию вируса ингибируется.	Основная функция – редактирование кодирующих участков. Например, редактирование одной из субъединиц глутаматного рецептора <i>gluR2</i> с изменением проницаемости канала	[52,53]
Функция при злокачественных новообразованиях	За счет подавления клеточного иммунитета способствует опухолевой прогрессии. Участвует в	Предположительно участвует в образовании опухолевых неоантигенов за счет редактирования кодирующих белки участков	[54,55]

	образовании опухолевых неоантигенов		
Роль в клеточном старении	Препятствует путем ингибирования активности LINE1-транспозонов	Не выявлена	[56]
Мутации в гене	Синдром Айкарди-Гутьереса: аутоиммунные и воспалительные реакции. Симметричный врожденный дисхроматоз, или акропигментация – нарушение пигментации кожи.	Приводят к гибели эмбриона. Фермент редактирует транскрипты белков, необходимых для нормального нейронального развития.	[57,58]

Дезаминирование аденозина может влиять и на судьбу самого вируса. В зависимости от природы вируса, а также активации того или иного молекулярного каскада в клетке, ADAR может обладать как про-, так и противовирусным действием.

Изменение вирусной РНК в процессе редактирования увеличивает разнообразие ее последовательностей. Такая отредактированная молекула способна в одних случаях оказывать провирусный эффект, а в других – наоборот, однако механизмы данного явления до конца не изучены. Редактирование вируса гриппа А может способствовать его деградации с помощью рибонуклеаз, распознающих инозиновые сигналы [59], а редактирование последовательности РНК вирусов лимфоцитарного хориоменингита и гепатита С препятствует их репликации [60,61]. В тоже время, за счет редактирования вирусных микроРНК, ADAR1 может способствовать усилению репликации вирусов Эпштейна-Барр и герпеса 8 типа [62,63]. Для вируса кори имеются данные о зависимости эффекта от количества событий редактирования. Так, в одном случае, провирусный эффект достигается в ответ на двухцепочечную РНК путем подавления активации протеинкиназы R [64]. Если же событий редактирования становится больше и возникает так называемое «гиперредактирование», то функция может стать противовирусной [65].

Функция ADAR1, связанная с иммунитетом, в основном, не приводит к перекодированию белковых последовательностей. Данный фермент для работы требует протяженных двухцепочечных участков РНК, в частности, для взаимодействия с транскриптами повторов геномов млекопитающих. У человека ADAR1 может ограничивать активность LINE1-ретротранспозонов по независимому от дезаминирования механизму [56]. С этими данными соотносятся новые исследования процессов старения. Показано, что активация транскрипции

упомянутого ретротранспозона вовлечена в процессы старения и развития интерферон-зависимого воспаления в стареющих клетках.

2.2. Редактирование РНК при злокачественных опухолях

В 2017 году была аннотирована полномасштабная база данных событий ADAR-опосредованного редактирования РНК человека, названная REDportal (a comprehensive database of A-to-I RNA editing events in humans). По результатам проекта найдено свыше 4 миллионов событий редактирования, представленных в 38% генов человека. Из них более четырнадцати тысяч принадлежали белок-кодирующим генам и примерно семь тысяч являлись генами для некодирующих РНК [66].

Большинство отредактированных сайтов, характерных для онкологических заболеваний, располагаются в 3'-нетранслируемых областях, интронах и межгенных областях. Однако редактирование кодирующих областей, составляющее примерно 72% от общих аннотированных белок-кодирующих локусов, может оказывать большое влияние на развитие рака [66]. В зависимости от расположения в геноме, эти события могут создавать разнообразные функциональные последствия, включая введение аминокислотных замен, воздействие на биосинтез микроРНК и распознавание мишеней, изменение сплайсинга и влияние на функции длинных некодирующих РНК. Важно также отметить, что изменение экспрессии самих генов *adar* влияет на изменения уровня редактирования, и, как следствие, может приводить к онкологическим заболеваниям [67]. Далее более детально будет рассмотрено, как отдельные события редактирования РНК приводят к развитию рака.

Изменение экспрессии ADAR1 и увеличение событий редактирования обнаружено при различных видах злокачественных опухолей, что описано в недавнем обзоре [68]. Показано, что из 504 типов злокачественных новообразований только в 7% повышается редактирование [68]. К злокачественным опухолям, наиболее подверженным дезаминированию, относятся рак яичника, молочной железы и меланома [55]. Такой феномен объясним тем, что ADAR1 помогает опухолям бороться с иммунной системой. С другой же стороны, редактирование способствует формированию антигенов, распознающихся молекулами главного комплекса гистосовместимости I класса. Иммунная система в процессе обучения вырабатывает резистентность к собственным антигенам на ранних стадиях развития организма, однако в этот процесс на тот момент могут быть не включены отредактированные фрагменты. Позднее, когда, например, активность ADAR с возрастом станет выше, такие отредактированные белки могут стать чужеродными.

В недавнем исследовании обнаружено, что неоантиген циклина I, белка клеточного цикла, образуется за счет активности ADAR1. В опухолях человека найдены эффекторные CD8+ Т-клетки с представленным эпитопом, специфичные по отношению к отредактированным пептидам вышеупомянутого белка [55].

В некоторых случаях связь ферментов редактирования и злокачественный рост может объясняться не только иммунными функциями. Блокировка ADAR1 в некоторых клеточных линиях замедляла рост опухолей другим путем [54]. Этот фермент способен подавлять протеинкиназу R (PKR), индуцируемую не только двухцепочечной РНК, но и другими стрессовыми факторами, характерными для раковых клеток. PKR способна ингибировать трансляцию белка, тем самым подавляя рост некоторых раковых клеточных линий.

На уровне целого организма ADAR1 может препятствовать ответу на лечение ингибиторами иммунных контрольных точек [69]. Снижение ферментативной активности ADAR может служить прогностическим фактором ответа на терапию лекарственным антителом PD-L1. Показано, что Т-клеточный ответ, пробуждаемый противоопухолевой терапией, идет вместе с интерфероновым ответом. ADAR может препятствовать этой реакции, делая опухолевые клетки устойчивыми к терапии. Для того, чтобы сделать опухоль чувствительной к терапии, предлагаются адресные методы доставки ингибиторов экспрессии ADAR1 [68]. Таким образом ADAR1 способен не только помогать опухолям уклоняться от иммунитета, но и формировать опасные для них неоантигены.

2.3. Ферменты ADAR и процессы в центральной нервной системе

Редактирование молекул РНК особенно выражено в центральной нервной системе (ЦНС) и в большей степени ассоциировано с активностью ADAR2 у млекопитающих (табл. 2). У плодовой мушки ген *Adar* также в большей степени экспрессируется в центральной нервной системе, что похоже на ADAR2 / *adarb1* млекопитающих [70]. Как правило, мутации в этом гене приводят к недостаточности двигательной способности и ускоренной нейродегенерации с возрастом [70]. Недавно было показано, что эти эффекты могут быть отменены активацией аутофагии, указывающих на то, что нарушение протеостаза и усиление передачи сигналов TOR являются важными механистическими ключами к объяснению эффектов дефицита *Adar* [71]. Более того, экспрессия гена *Adar* усиливается во время метаморфоза насекомых [72]. Этот сложный процесс регуляции транскрипции описан множеством альтернативно запускаемых сплайсированных транскриптов, различных для эмбрионов и взрослых насекомых [72]. Кратко говоря, конститутивный промотор 4A генерирует эмбриональные транскрипты, несколько из которых также характеризуются включением дополнительного экзона 3a, который кодирует

фрагмент белка из 38-ми аминокислот в конкретной изоформе 3a ADAR. Считается, что промотор 4B обеспечивает исключение экзона 3a с образованием специфичных для взрослых 3/4 изоформ белка, которые преобладают у насекомого после метаморфоза [73]. В транскриптах, специфичных для взрослых, также был описан процесс ауторедктирования, который приводит к преобразованию остатка серина в глицин в соответствующих белках. У взрослых насекомых встречаются как отредактированные, так и неотредактированные изоформы ферментов [73]. Из этих данных можно предположить, что разнообразие транскриптов во время онтогенеза плодовой мушки может указывать на дифференциальное редактирование выбранных мишеней. Однако подробный механизм этого процесса еще предстоит выяснить.

На белковом уровне ADAR-опосредованное редактирование РНК у млекопитающих вносит точечные замены в аминокислотную последовательность потенциал-управляемых ионных каналов и рецепторов G-белков. Также ADAR2 воздействует в ЦНС на малые интерферирующие РНК [74]. Наиболее известными примерами природного редактирования является дезаминирование аденозина в мРНК глутаматных рецепторов групп GluK и GluA. При нормальном развитии нервной системы после действия фермента происходит замена глатумина на аргинин [75]. Если бы редактирования не произошло, дефектная субъединица пропускала бы ионы кальция, вызывая эксайтотоксичность из-за повышения глутамата в синапсе. Такое явление привело бы к гибели нейронов, как, например, в случае спинальной дегенерации при боковом амиотрофическом склерозе [76]. Нокаут гена, кодирующего ADAR2 у мышей, приводит к судорожным припадкам, несовместимых с жизнью, из-за повышенного возбуждения в головном мозге. У человека мутации в *adarb1* связаны с эпилепсией и такими дефектами нейрональной пластичности, как аутизм и синдром Мартина-Белл [77].

Таблица 2. Редактирование ферментами ADAR субъединиц глутаматных рецепторов, вовлеченное в развитие патологических состояний нервной системы человека. Адаптировано из [78,79].

Субъединица глутаматного рецептора	Сайт редактирования	Патологическое состояние	Функция рецептора	ADAR, участвующий в замене	Ссылка
GLUA2	Q/R	ишемия переднего мозга	изменение проницаемости ионов Ca^{2+}	ADAR2	[53,80]
		боковой амиотрофический склероз			[81–83]
		шизофрения			[84]
		глиобластома			[85]
	R/G	травма спинного мозга	снижение чувствительности и рецепторов	ADAR1, ADAR2	[86,87]
		эксцитотоксичность			[88]
		эпилепсия			[89,90]
		лечение антидепрессантами			[91,92]
	Q/R, R/G	болезнь Альцгеймера		ADAR1, ADAR2	[93,94]
GLUA3, GLUA4	R/G	травма спинного мозга	снижение чувствительности и рецепторов	ADAR1, ADAR2	[86,87]
		эксцитотоксичность			[88]
		болезнь Альцгеймера			[93]
GLUK1	Q/R	травма спинного мозга	изменение проницаемости ионов Ca^{2+}	ADAR1, ADAR2	[95]
		эпилепсия			[96]
		условно-рефлекторное замирание			[97]
GLUK2	Q/R, I/V, Y/C,	травма спинного мозга	изменение проницаемости ионов Ca^{2+}	Y/C - ADAR2	[95]
	Q/R	эпилепсия		ADAR1, ADAR2	[96,98,99]
	I/V	биполярное расстройство			[100]
	Q/R, I/V	лечение антидепрессантами			[91,101]

Редактирование РНК в головном мозге может усиливаться с возрастом и для управления возбуждающими сигналами через ионные каналы. Также оно изменяется при спинальной травме и нейропатической боли [102]. Хотя значение редактирования РНК, за исключением этого явления в некоторых значимых мРНК, в ЦНС изучено недостаточно, оно, похоже, имеет функциональное значение. Например, как стало недавно известно, у крыс профиль участков редактирования РНК, вызванного ADAR, менялся при хроническом стрессе [103]. В работе Учида с соавторами показано, что ADAR2 является медиатором тактильной аллодинии, индуцированной травмой и, следовательно, потенциальной терапевтической мишенью для лечения нейропатической боли [102].

2.4. Способы идентификации событий редактирования РНК

В настоящее время описано несколько способов идентификации событий редактирования РНК. Большинство известных отредактированных участков обнаружены с использованием платформ высокопроизводительного секвенирования и вычислительных технологий, сравнивающих геномные и транскриптомные последовательности одного образца [104,105]. Инозин при анализе распознается как гуанозин, а выявленные отличия соответствующих остатков свидетельствуют о наличии редактирования РНК. В 2010 г. Сакураи и соавторами предложен метод химической идентификации инозинов, названный ICE (Inosine Chemical Erasing – химическое удаление инозина). В данном методе события редактирования обнаруживаются посредством цианоэтилирования неклассического азотистого основания, обратной транскрипции, и дальнейшего секвенирования коротких продуктов [106]. Чжан и соавт. разработали метод mmPCR-seq, который объединяет мультиплексную микрофлюидную ПЦР с высокопроизводительным секвенированием. Данная техника позволяет одновременную амплификацию и детектирование событий редактирования в 960 локусах 48 образцов [107].

Один из наиболее оптимальных методов поиска событий ADAR-опосредованного редактирования РНК, основан на химической модификации инозина и называется ICE-seq [108]. Суть его заключается в комбинации цианоэтилирования инозина с высокопроизводительным секвенированием. Первый этап данного метода – реакция цианоэтилирования (ЦЭ) инозина, в которой акрилонитрил присоединяется к N¹-атому инозина с образованием N¹-цианоэтилинозина [106]. Выделенную поли(А)-РНК обрабатывают раствором, содержащим акрилонитрил [109]. В качестве контроля используют такой же раствор, не содержащий акрилонитрил. Последующие этапы анализа следуют стандартным протоколам высокопроизводительного секвенирования, в котором сравнивают прочтения кДНК, полученной из образцов после обработки акрилонитрилом и без нее.

Накопленные обозначенными выше методами сведения вошли в специализированные базы данных редактирования мРНК. Наиболее известными из таких баз считаются RADAR (A collection of rigorously annotated A-to-I events) [110] и DARNED (DAtabase of RNa EDiting) [111]. В таких базах содержатся аннотированные участки редактирования для человека и таких модельных организмов, как плодовая мушка и мышь.

2.5. Протеогеномный подход для поиска последствий редактирования РНК на белковом уровне

Использование панорамной (shotgun) протеомики с масс-спектрометрией высокого разрешения позволило резко увеличить количество идентифицируемых в образцах белков [112]. Именно с появлением возможности регистрации тысяч пептидов и стала достижима идея анализа тонких изменений структуры белка. Таким образом появилась возможность регистрировать протеоформы – изменения белковой структуры по сравнению с вариантом, кодируемым в консенсусном геноме [113]. В качестве примера протеоформ можно назвать результаты генного полиморфизма, ферментативного расщепления, а также различные посттранскрипционные модификации, к которым относится как альтернативный сплайсинг, так и события редактирования.

Во времена становления протеомики геномные базы использовались в одностороннем порядке. С повышением надежности масс-спектрометрии, начали предприниматься попытки изменения направления потока информации от протеома к геному. Например, панорамная масс-спектрометрия стала использоваться для уточнения расположения кодирующих рамок считывания в геномах плохо изученных организмов [114]. Такой подход называли протеогеномикой. По мере развития данного подхода, в состав протеогеномики вошло также такое направление, как поиск белков с отклонениями от консенсусного генома с использованием модифицированных геномных баз данных. Такие базы учитывают различные события изменения генома, например, полиморфизм или альтернативный сплайсинг мРНК [115,116]. Это направление нашло свое применение для поиска протеоформ при злокачественных опухолях [117].

Как указывалась выше, последние исследования ADAR-опосредованного редактирования в основном осуществляются посредством секвенирования нуклеиновых кислот нового поколения. Однако, присутствие в исследуемом органе или ткани отредактированных последовательностей белка предполагается по умолчанию. Количественные же отношения отредактированных и неотредактированных форм белка являются малоизученными и

представляют большой интерес. Важно понимать, стабильны ли отредактированные формы белка или после трансляции быстро элиминируется системой протеостаза.

Какие же функциональные последствия на белковом уровне связаны с заменой аденозина на инозин в молекуле мРНК? Впервые события редактирования в масштабах протеома были зарегистрированы в работе по исследованию протеоформ в печени крыс [118]. В статье идентифицированы 20 участков редактирования на белковом уровне без дополнительной валидации. На настоящий момент известно, что процесс перекодирования белкового уровня особенно широко представлен у беспозвоночных, в частности, головоногих моллюсков. Методом панорамной протеомики обнаружено несколько сотен участков последствий редактирования у калифорнийского осьминога. Около 50 из таких событий найдены в одном из калиевых каналов, что меняет кинетику его инактивации [119]. Такой высокий уровень редактирования считается адаптацией к меняющимся температурам воды, которые по термодинамическим принципам могут воздействовать на работу каналов. Таким образом, для пойкилотермных животных редактирование РНК важно для быстрого контроля множества разнообразных физиологических функций клетки, постольку геномные адаптации в этой ситуации невозможны [120].

Гомойотермным млекопитающим нет нужды интенсивно отвечать на изменения температуры окружающей среды, однако их протеом тоже перекодируется в процессе редактирования РНК ферментами ADAR. Наиболее выражен данный процесс в центральной нервной системе, за что отвечает ADAR2 [41]. Можно предположить, что такая система реагирования важна для правильного функционирования нервных клеток под воздействием различных внешних факторов.

Что касается отредактированного протеома млекопитающих, важное исследование в этой области выявило различия между последовательностями ДНК и РНК с помощью полногеномного анализа В-клеток человека. Белок-кодирующие замены были подтверждены панорамной протеомикой, многие из которых были вызваны активностью ферментов ADAR [121]. Совсем недавно белки, измененные путем редактирования РНК были идентифицированы путем повторного анализа протеогеномных данных рака груди, яичника и колоректального рака проекта TCGA (атлас ракового генома, The Cancer Genome Atlas) [34]. Только тринадцать отредактированных сайтов были подтверждены в протеомах по данным TCGA. Предполагается, что, по крайней мере, один из этих сайтов, расположенный в белке транспортных везикул COPA, участвует в прогрессировании рака молочной железы [34]. Напротив, отредактированный сайт в белке циклина I, как сообщалось, индуцирует неопитоп, который вызывает цитотоксический Т-клеточный ответ против опухоли [55]. В этом исследовании пептидом, связанный с человеческим лейкоцитарным антигеном, главным комплексом гистосовместимости I класса (major

histocompatibility complex I, MHC-I), был изучен протеоеномикой, при этом в подмножестве из 1369 сайтов только пять отредактированных антигенных пептидов были достоверно идентифицированы [55].

Таким образом, на настоящий момент известно, что ферменты ADAR функционируют в центральной нервной системе и некоторых других тканях большинства многоклеточных организмов, включая человека. Однако само явление перекодирования вследствие редактирования аденозиндезаминазами на белковом уровне при развитии и нормальном функционировании центральной нервной системы изучено недостаточно. Поэтому исследованию этого явления в масштабах протеомов и посвящена данная работа.

3. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

3.1. Исходные данные, полученные из открытого доступа

В этой работе использовали результаты панорамного протеомного анализа *D. melanogaster* линии Canton-S, выполненного для протеомов целых тел, голов, а также протеомы изолированных мозгов плодовой мушки. Для поиска сайтов редактирования РНК применяли данные нескольких протеомных поисков. Один протеом получали экспериментально, а два других брали из открытого доступа [122,123]. Для изучения процесса перекодирования белков в процессе жизненного цикла, выбирали данные, представляющие собой протеомы эмбрионального развития (PXD005713) и всего жизненного цикла плодовой мушки (PXD005691) [29]. В этих наборах данных исследовали протеомы целых тел *D. melanogaster* линии Oregon R, выращенных при 25 ° C. Данные эмбриогенеза получали из яиц, собираемых каждый час в диапазоне от 0 до 6 часов, а затем каждые 2 часа до 20 часов. Весь жизненный цикл анализировали в следующих стадиях: активация зиготических генов (0-2 часа), гаструляция (4-6 часов), органогенез (10-12 часов) и поздние стадии эмбриогенеза (18-20 часов). Для личинки выбирали четыре стадии: L1, L2, ранняя L3, поздняя L3 (ползающая личинка). Куколки анализировали в следующие временные точки: P1 (0-14 часов после окукливания), P2-P5 через каждые 24 часа после окукливания. Из данных о взрослых особях для анализа выбирали неоплодотворенных мух (взрослые молодые самец и самка - aumf) и недельных животных (взрослые самцы и самки - amf).

Протеомы мозга мыши и человека, по три для каждого вида, отбирали из репозитория открытых данных Proteomexchange [124]. Наборы данных представляли собой протеомы, полученные масс-спектрометрией высокого разрешения. Для мозга мыши и человека выбирали хромато-масс-спектрометрические данные высокого разрешения мозга, доступные по

идентификаторам PXD001250 [125] и PXD005445 [126], соответственно. Эти данные, взятые в качестве базовых эталонных, выбирали из проектов по инвентаризации протеомов головного мозга, так как они имели наибольшее количество идентифицированных белков. В качестве дополнительных, менее полных данных для проверки результатов, для мыши выбирали протеомы моделей болезни Гентингтона [127] и нейропатической боли [128] с номерами доступа PXD003442 и PXD004087 соответственно. Дополнительные данные протеомов человека получали из работы по изучению спинномозговой жидкости [129], PXD009646, и из протеоеномного исследования мозга, предназначенного для идентификации белковых вариантов [130], PXD004143. Исследование спинномозговой жидкости выбиралось для проверки, могут ли перекодированные белки быть идентифицированы в образце, который обычно лучше доступен в клинических условиях, чем ткани мозга.

3.2. Характеристика образцов

3.2.1. Культура *Drosophila melanogaster*. Возрастная модель

Живые особи *D. melanogaster* линии Canton S были предоставлены Натальей Романовой с биологического факультета Московского государственного университета имени М. В. Ломоносова. Питательная среда для мух готовилась с использованием порошка Formula 5-24 Instant *Drosophila* Medium (Carolina Biological Supply Company, США), разведенного дистиллированной водой в соотношении 1:1. Для получения гормонов, необходимых для окукливания, в среду добавлялись несколько гранул сухих пекарских дрожжей – *Saccharomyces cerevisiae*. Мушки выращивались в одноразовых пробирках объемом 50 мл, закрытых ватными тампонами, при температуре 25°C. Каждые 14 дней взрослые особи пересаживались в новую пробирку со свежей питательной средой.

Для изучения явления редактирования мРНК на разных стадиях онтогенеза плодовых мушек с помощью таргетного протеомного анализа использовали мозг личинок третьей стадии (L3), мозг поздних куколок (P12-14 согласно [131]) и молодых взрослых мух (возраст примерно 7 дней) обоих полов. Личики третьей стадии появлялись в пробирке примерно через 5 суток после подсаживания взрослых особей. Куколки использовались на стадии, когда под покровной кутикулой идентифицировались структуры тела, придающие сходство со строением взрослой особи – зачатки крыльев, туловище, ноги, антенны, ротовые придатки. Такие особи появлялись через 9 суток после подсаживания взрослых особей.

3.2.2. Характеристика образцов млекопитающих

Образцы мозжечка и коры трех линий трансгенных мышей - моделей заболеваний ЦНС, и соответствующих им контрольных групп предоставлены лабораторией генетического моделирования нейродегенеративных процессов ИФАВ РАН. Трансгенные мыши характеризовались такими состояниями, как FUS-опатия (отчасти моделирует нейродегенерацию, например, при боковом амиотрофическом склерозе), таупатия (гиперфосфорилирование белка тау, имитирующее в итоге признаки болезней Паркинсона, Пика и Альцгеймера), а также моделирующий психическое расстройство нокаут гена некодирующей РНК NEAT1. Каждый образец проанализировали в трех биологических повторах.

Образцы цереброспинальной жидкости от 25 пациентов с нейродегенеративными и сердечно-сосудистыми заболеваниями предоставлены ФНКЦ ФХМ ФМБА России.

3.3. Выделение мозга *Drosophila melanogaster*

В течение всей процедуры замороженные насекомые находились на льду в чашке Петри. При изолировании мозга *Drosophila melanogaster* в стадии имаго использовались только особи мужского пола. С помощью препаративных пинцетов туловище особи захватывалось в области грудного отдела чуть ниже головы и полностью погружалось вентральной поверхностью тела вверх в 0,01 М натрий-фосфатного буфера (англ. phosphate buffered saline, PBS) pH 7,4 (приготовленный из таблетированной формы PBS, Sigma-Aldrich, США). Голова отделялась от тела за хоботок, туловище удалялось. Все дальнейшие манипуляции осуществлялись под визуальным контролем стереомикроскопа SMZ645 (Nikon, Япония) с увеличением 10x1. Диссекция начиналась с разрывания кутикулы между хоботком и глазом. На протяжении всей процедуры пинцеты находились только под сетчаткой или кутикулой, чтобы не повредить ткань мозга. Головная капсула раскрывалась последовательно от ротовой к затылочной части вплоть до полного обнажения мозга. Выделенный мозг отбирали с помощью автоматической пипетки вместе с PBS.

Работа с куколкой начиналась с освобождения тела от покровной кутикулы с помощью разрыва её в головной части. Дальнейший алгоритм изолирования мозга из куколок повторял те же стадии, что и у взрослых особей. Процедура извлечения головного мозга из личинок третьей стадии начиналась с погружения живой личинки в буфер. Далее с помощью пинцетов личинка разрывалась чуть ниже области проекции головного мозга на дорсальной стороне. Чтобы найти головной мозг, из верхней трети туловища за среднюю кишку вытягивалось всё содержимое

вплоть до визуализации мозга. Слюнные железы, имагинальные диски и пищевод, соединенные с мозгом, отделяли для полного его обнажения.

Для получения каждого образца в зависимости от типа анализа собирали определенное количество материала, которое затем центрифугировали при 6000 g в течение 15 минут (центрифуга 5415R, Eppendorf, Германия). В конце процедуры буферный раствор удаляли, а осадок мозгов замораживали при -80°C для последующей пробоподготовки.

3.4. Пробоподготовка к протеомному анализу

Для получения клеточного осадка образцов мозга мышей клетки лизировали в буфере, состоящем из 0,1 М трис-HCl, pH 8,0, 0,1 М DTT и 2% SDS (мас./об.) при 99°C в течение 5 мин. Лизаты обрабатывали с помощью ультразвукового гомогенизатора Bandelin Sonopuls HD2070 (Bandelin Electronic, Германия). После измерения концентрации белка, описанной ниже, 100 мкг белка подвергали метанол-хлороформной экстракции. Для этого добавляли 400 мкл метанола, тщательно перемешивали и добавляли 100 мкл хлороформа. После перемешивания смеси добавляли 300 мкл ddH₂O и центрифугировали 1 минуту при 14 000 g на центрифуге 5415R (Eppendorf, Германия). После отбора верхнего водного слоя, добавляли 400 мкл метанола, перемешивали образцы и центрифугировали 5 минут при 20 000 g. После удаления метанола супернатант отбирали и хранили образцы при -80°C для последующей пробоподготовки.

Замороженные осадки образцов мозга мух, мыши и ликвор ресуспендировали в 100 мкл лизирующего буфера, содержащего 0.1% (мас./об.) Protease MAX Surfactant (Promega, США), 50 mM бикарбонат аммония и 10% (об./об.) ацетонитрил (ACN). Полученный раствор инкубировали в термошейкере TS-100 (BioSan, Латвия) при комнатной температуре в течение 60 мин при 550 об/мин. Затем смесь подвергали обработке ультразвуком с помощью ультразвукового гомогенизатора Bandelin Sonopuls HD2070 (Bandelin Electronic, Германия) с помощью коротких импульсов, амплитудой 30% в течение 5 минут. Супернатант отбирали после центрифугирования при 15700 g в течение 10 минут при 20°C на центрифуге 5415R (Eppendorf, Германия). Измерение концентрации белков проводили с применением бицинхониновой кислоты [132].

3.4.1. Измерение концентрации белков с использованием ВСА

Реагент А получали путем растворения в 80 мл деионизованной воды (mQ) 0,1 г 2,2-бицинхониновой кислоты, 2 г натрия карбоната безводного (Acros Organics, США), 0,95 г натрия гидрокарбоната, 0,1 г сегнетовой соли и 0,4 г. натрия гидроксида (AcrosOrganics, США).

Доводили раствором гидроксида натрия до pH 11,25. Добавляли деионизованную воду до 100мл. Реагент Б получали путем растворения в 100 мл деионизованной воды 4 г сульфата меди.

Для построения калибровочной кривой выбирали следующие точки, соответствующие количеству белка-стандарта (BSA), разведенному в 30 мкл ddH₂O: 2 мкг, 5 мкг, 10 мкг, 25 мкг, 50 мкг. Каждую измеряемую пробу белка объемом 2 мкл разводили в 28 мкл деионизованной воды (ddH₂O). К каждой пробе стандарта, измеряемым пробам и контрольной пробе (30 мкл ddH₂O) добавляли по 1 мл реагента А и 20 мкл реагента В, после чего перемешивали и инкубировали в течение 20 мин при температуре 56°C со скоростью перемешивания 300 об/мин в термо-шейкере TS-100 (BioSan, Латвия). Затем пробы остужали до комнатной температуры и проводили измерение оптической плотности при длине волны 562 нм с помощью спектрофотометра NanoDrop-1000 (Thermo Fisher Scientific, США) с программным обеспечением ND-1000 v.3.5.1.

3.4.2. Гидролитическое расщепление белков в растворе

Восстановление дисульфидных мостиков: к образцам белков, полученных в результате экстракции, добавляли раствор для восстановления до конечной концентрации DTT 10 mM и инкубировали при температуре 56°C в течение 20 минут в термошейкере TS-100 (BioSan, Латвия) со скоростью перемешивания 300 об/мин.

Алкилирование сульфгидрильных групп цистеинов: по окончании процедуры восстановления к каждой пробе добавляли алкилирующий раствор до конечной концентрации хлорацетимида 10 mM и инкубировали в темноте при комнатной температуре в течение 30 минут.

После восстановления и алкилирования в каждый образец добавляли Trypsin Gold (Promega, USA) в соотношении массы трипсина к массе тотального белка 1/100 и затем инкубировали в течение 8 часов при температуре 37°C. Гидролиз останавливали добавлением 100% муравьиной кислоты в каждый образец до конечной концентрации 5% (мас./об.). Затем центрифугировали при 15700 g в течение 10 минут при 20°C (центрифуга 5415R, Eppendorf, Германия). Супернатант отбирали и наносили на фильтр с размером пор 10 кДа (Millipore, США), затем центрифугировали при 13400 g в течение 20 минут при 20°C. После этого, добавляли на фильтр 100 мкл 50% муравьиной кислоты и повторно центрифугировали при 13400 g в течение 20 минут при 20°C. Конечную концентрацию пептидов измеряли с использованием стандартного протокола набора Peptide Assay (Thermo Fisher Scientific, США). Измерение оптической плотности проводили при длине волны 562 нм с помощью спектрофотометра

NanoDrop-1000 (Thermo Fisher Scientific, США) с программным обеспечением ND-1000 v.3.5.1.

После измерения концентрации, пептиды высушивали в вакуумном концентраторе Concentrator 5301 (Eppendorf, Германия) при 45°C. Высушенные пептиды хранили при –80 до масс-спектрометрического анализа.

3.5. Панорамный протеомный анализ

Хроматографическое разделение пептидов производили с использованием колонки C18, 25 см (внешний диаметр 360 мкм, внутренний диаметр 75 мкм, New Objective, США), подключенной к системе нанохроматографии Ultimate TM 3000 RSLC (Thermo Fisher Scientific, США). Пептиды элюировали при скорости потока 300 нл/мин в течение 240 мин при линейном градиенте от 2% до 26% ацетонитрила в 0,1% муравьиной кислоте. Элюированные пептиды подвергали электроспрейной ионизации и анализировали на масс-спектрометре Orbitrap QExactive Plus (Thermo Fisher Scientific, США) с разрешением 60000 в диапазоне m/z 200-2000.

Данные MS/MS для 20 наиболее интенсивных ионов-предшественников, имеющих заряд 2 и выше, получали с использованием метода высокоэнергетической столкновительной диссоциации (HCD) с разрешением 15000. Динамическое исключение до 500 предшественников в течение 60 с использовалось для избегания повторного анализа тех же пептидов.

Протеомные данные, полученные в этой работе, поместили в общедоступный репозиторий ProteomeXchange ([http:// www. Proteomexchange.org/](http://www.Proteomexchange.org/)) [124], идентификационный номер PXD009590.

3.6. Синтез пептидных стандартов

Синтез пептидов осуществляли твердофазным методом с использованием 9-флуоренилметилоксикарбонил (Fmoc)-защищенных по α -аминогруппам производных аминокислот (Novabiochem, Merck, США) со следующими защитами боковых функциональных групп: аспарагиновая кислота, глутаминовая кислота – трет-бутоксигруппой (OtBu¹); глутамин, аспарагин, гистидин – тритильной (Trt); треонин – трет-бутильной (tBu). Для включения изотопномеченого аминокислотного остатка в пептиды белка комплексина (ген *srpX*) (NQMETQVNELhK и NQIETQVNELhK)¹ использовали изотопномеченый лейцин (Fmoc-Leu-¹³C₆,¹⁵N, Cambridge Isotope Laboratories). Для этих пептидов использовали носители типа

¹ Здесь и далее в скобках указаны, соответственно последовательности перекодированных и кодируемых в геноме вариантов пептидов

Lanterns D-серии фирмы Mimotopes с присоединённой первой аминокислотой L-Lys (Boc). В качестве твердой фазы для пептидов белков синтаксин-1A (Syx1A) (IEYHVEHAMDYVQTATQDTKh и IEYHVEHAVDYVQTATQDTKh), эндофилин-A (EndoA) (YSLDDNIKh и YSLDDNIEQNFLEPLHHMQTKh), Ca^{2+} -зависимый активатор секреции (Cadps) (LMSVLESTLSKh и LVSVLESTLSKh) и неотредактированного варианта пептида белка атаксин-2 (Atx2) (GVGPAPSANASADSSSKh) использовали смолу с присоединённым изотопсодержащим лизином (L-Lys (Boc) ($^{13}\text{C}_6$, 99%; $^{15}\text{N}_2$, 99%) – 2-Cl-Trt, Cambridge Isotope Laboratories). Смолу с присоединённым изотопсодержащим аргинином (L-Arg (Pbf) ($^{13}\text{C}_6$, 99%; $^{15}\text{N}_4$, 99%) – 2-Cl-Trt, Cambridge Isotope Laboratories) использовали для синтеза пептидов субъединицы кальциевого канала (stol) (LVTTVSTPVFDRh и LVTTVSTPVFDGRh), ГТФаза-активирующего белка (RhoGAP100F) (YLLQIWPQPQAQHRh и YLLQIWPQPQAQHQRh) и отредактированного варианта пептида белка атаксин-2 (ген Atx2) (GVGPAPSANASADSSSRh).

Емкость твердофазных носителей составляла 2 мкМ. В качестве реакционного сосуда использовали 96-луночный фильтр-планшет (ArcticWhite, USA). Для конденсации Fmoc-аминокислот использовали 1Н-Бензотриазолиум 1-[бис(диметиламино)метиле́н]-5хлорогексафлуорофосфат (1-),3-оксид (HCTU), в присутствии 2,4,6-триметилпиридина (TMP). В качестве растворителя в пептидном синтезе использовали N, N-диметилформамид (DMF) (ос.ч.), который перед использованием выдерживали не менее 3 дней в темноте над свежeproкаленными при 300°C молекулярными ситами 4А (80 г на 1 л DMF). Удаление защитных Fmoc-групп с первой аминокислоты на носителе и с растущей пептидной цепи проводили 20%-ным раствором 4-метилпиперидина (4M-pip) в N, N-диметилформамиде. Операции по наращиванию пептидной цепи проводили по следующему протоколу для каждого синтетического цикла:

1. Для снятия Fmoc защиты с α -аминогруппы в лунки фильтр-планшета добавляли 0,2 мл 20% 4M-pip, инкубировали при перемешивании в течение 10 минут при комнатной температуре, после чего раствор отфильтровывали. Операцию проводили дважды.
2. Для промывки твердофазного носителя в лунки планшета добавляли 0,5 мл DMF, инкубировали при перемешивании в течение 1 минуты при комнатной температуре, после чего раствор отфильтровывали. Операцию проводили 6 раз.
3. Для присоединения аминокислоты в лунки планшета добавляли 0,05 мл 0.2M раствора аминокислоты, 0,1 мл 0,1 М раствора HCTU и 0,1 мл 0,2 М раствора TMP, инкубировали при перемешивании в течение 60 минут при комнатной температуре, после чего раствор отфильтровывали.

4. Для промывки твердофазного носителя в лунки планшета добавляли 0,5 мл DMF, инкубировали при перемешивании в течение 1 минуты при комнатной температуре, после чего раствор отфильтровывали. Операцию проводили 4 раза.

По окончании наращивания пептидной цепи проводили снятие Fmoc защиты с концевой α -аминогруппы пептида (см. шаг 1 и 2), затем носители промывали в 0,5 мл дихлорметана и высушивали. После чего проводили снятие пептида с твердой фазы и одновременное деблокирование боковых групп в 0,2 мл смеси трифторуксусная кислота: 3,6-диокса-1,8-октандитиол: триизопропилсилан:анизол:вода в объемном соотношении 183:5:2:5:5 в течение 2 часов. По окончании инкубации смесь отфильтровывалась от полимера через фильтр планшета в пробирки. В полученный раствор затем добавляли 1,8 мл охлажденной смеси метил-трет-бутиловый (или диэтиловый) эфир (далее – эфир):петролейный эфир (1:2 об./об.). Полученную суспензию охлаждали 30 мин при -20°C и затем центрифугировали 6 мин. при скорости 6000 об/мин. Надосадочную жидкость удаляли, полученный осадок высушивали при 30°C в вакуумном концентраторе Concentrator 5301 (Eppendorf, Германия). Для дальнейшей работы пептиды растворяли в 0,2 мл 50% ацетонитрила в воде.

Для установления наличия целевого пептида в образце после синтеза проводили хромато-масс-спектрометрический анализ. Для этого использовали хроматографическую станцию Agilent ChemStation 1200 series сопряженную с масс-спектрометром Agilent 1100 series LC/MSD Trap XCT Ultra (Agilent Technologies, США). Agilent ChemStation 1200 series включала: 4-канальный вакуумный блок дегазации растворов (G1322A), 4-канальный насос (G1311A), автосамплер (1329A), термостат для автосамплера (G1330B), термостатируемое колоночное отделение (G1316A), диодно-матричный спектрофотометрический детектор (G1315B). Аналитическую хроматографию проводили в следующих условиях: раствор А - 0,1% трифторуксусной и муравьиной кислоты в воде, раствор В - 0,1% трифторуксусной 0,1% муравьиной кислоты 80% ацетонитрила в воде. Колонка Zorbax 300 Extend-C18 4,6x150 мм 5 μ (Agilent Technologies, США). Скорость потока 1 мл/мин. Объем наносимой пробы 2 мкл. Настройки масс-спектрометрического детектора: напряжение капилляра 3500В, скорость потока осушающего газа 11 л/мин (температура 350°C); давление газа в небулайзере 23 psi.

3.7. Мониторинг множественных реакций

Пробы для первого биологического повтора анализировались с помощью хроматографической системы Dionex UltiMate 3000 RSLC nano System Series (Thermo Fisher Scientific, США), соединенной с масс-спектрометром с тремя квадрупольями TSQ Vantage (Thermo Fisher Scientific, США) в пяти технических повторах.

Каждый образец объемом 1 мкл, содержащий 2 мг нативных пептидов и 100 фмоль изотопномеченных пептидных стандартов, загружали на предколонку Zorbax 300SB-C18 (диаметр частиц 5 мкм, 5 мм × 0.3 мм) (Agilent Technologies, США) и промывали подвижной фазой, содержащей 5% ацетонитрил, в течение 5 мин при скорости потока 10 мкл/мин.

Пептиды разделяли на аналитической колонке Zorbax 300SB-C18 (диаметр частиц 3,5 мкм, 150 мм × 75 мкм) (Agilent Technologies, США), используя линейный градиент от 95% фазы А (0,1% муравьиная кислота) и 5% фазы Б (80% ацетонитрил, 0,1% муравьиная кислота) до 60% фазы А и 40% фазы Б, в течение 25 минут при скорости потока 0,4 мкл/мин.

MRM анализ проводили на масс-спектрометре с тремя квадрупольями (QqQ) TSQ Vantage (Thermo Fisher Scientific, США), оборудованном наноэлектроспрейным источником. Капиллярное напряжение устанавливали на уровне 2100 В, с окном изоляции 0,7 Да для первого и третьего квадрупольей, время цикла составляло 3 с. Фрагментация ионов-предшественников проводилась при 1 мТорр с использованием энергии столкновения, рассчитанной программным обеспечением Skyline 3.1 (MacCoss Lab Software, США).

Измерения для второго биологического повтора проводили следующим образом. Каждый образец анализировали с помощью микропоточной системы ВЭЖХ Agilent 1200 (Agilent Technologies, США), подключенной к тройному квадрупольному масс-спектрометру TSQ Quantiva (Thermo Fisher Scientific, США) в трех технических повторах. 3 мкл образца, содержащего приблизительно 10 мкг нативных пептидов и 500 фмоль каждого стабильного изотопно-меченого стандарта, загружали в аналитическую колонку Zorbax 300SB-C18 (5 мкм, 150 × 0,3 мм) (Agilent Technologies, США) и промывали 5% ацетонитрилом в течение 5 мин при скорости потока 20 мкл / мин. Пептиды разделяли с использованием линейного градиента от 95% растворителя А (0,1% муравьиной кислоты) и 5% растворителя В (80% ацетонитрила, 0,1% муравьиной кислоты) до 50% растворителя А и 50% растворителя В в течение 30 минут при скорости потока 20 мкл/мин. MRM-анализ проводили с использованием тройного квадруполья TSQ Quantiva (Thermo Scientific, США), снабженного электроспрейной ионизацией. Капиллярное напряжение было установлено на 4000 В, температура трубки для переноса ионов составляла 325 ° С, температура испарителя составляла 40 ° С. Окно изоляции было установлено на 0,7 Да для первого и третьего квадруполья, а время цикла составляло 1,2 с. Эксперимент MRM проводили по расписанию с временным окном удерживания 2 мин для каждого иона-предшественника. Фрагментацию ионов-предшественников проводили при 1,2 мТорр с использованием энергии столкновения, рассчитанной с помощью программы Skyline 4.1 (MacCoss Lab Software, США) [48]. Необходимость пересчета параметров для анализа

продиктовывалась использованием другого масс-спектрометра для второго биологического повтора.

Количественный анализ данных MRM также выполнялся с использованием программного обеспечения Skyline 3.1. Данные количественной оценки получали из соотношения площадей под кривой MRM-переходов нативных пептидов к изотопномеченным стандартам. Для каждого пептида использовали 5 переходов. Результаты были проверены с использованием программного обеспечения Skyline для сравнения хроматографических профилей нативных и изотопномеченных пептидов, типовые результаты с сигналами от природного пептида и синтетического стандарта представлены на рис. П1 Приложения.

Все данные масс-спектрометрии MRM депонированы в PASSEL (<http://www.peptideatlas.org/passel/>) [49] под номером доступа PASS01553.

3.8. Секвенирование ДНК

ДНК для генотипирования выделялась из 100 голов *D. melanogaster* с использованием стандартного метода фенол-хлороформной экстракции [133]. Семь полиморфных сайтов в генах плодовой мушки (*Syx1A*, *cpx*, *Atpalpha*, *stol*, *alpha-Spec*, *Cadps*, *EndoA*, *Atx2*, *RhoGAP100F*) секвенировались с использованием метода Сэнгера на генетическом анализаторе Applied Biosystems 3500xL с программным обеспечением SeqScape (Thermo Fisher Scientific, США).

Первоначальные ПЦР проводили в объеме 25 мкл, содержащем 50 мкг геномной ДНК-матрицы, 10-кратный ПЦР-буфер, 0,5 МЕ ДНК-полимеразы HS Taq, 0,2 мМ dNTP (Evrogen, Россия) и 80 пмоль каждого праймера. Условия ПЦР были одинаковыми для всех сайтов: 95°C в течение 5 минут, затем 35 циклов, включающих 94°C в течение 15 секунд, 59°C в течение 20 секунд, 72°C в течение 20 секунд и конечной элонгации при 72°C в течение 6 минут. Праймеры подбирали с использованием бесплатного программного обеспечения PerlPrimer (<http://perlprimer.sourceforge.net/>). Для секвенирования использовали те же праймеры. ПЦР-продукты затем очищали путем инкубации с смесью 1 МЕ ExoI и 1 МЕ SAP (Thermo Fisher Scientific, США) при 37 ° C в течение 30 минут с последующим повышением температуры до 80°C на 15 минут. Праймеры, используемые в работе, синтезировались компанией Евrogen, Россия и представлены в таблице 3.

Секвенирование с последующей очисткой при помощи ЭДТА и этанола проводили с использованием набора BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, США) в соответствии с инструкциями производителя.

3.9. Метод химического удаления инозина

Для идентификации инозинов в составе молекулы РНК использовали метод химического удаления инозина (англ. inosine chemical erasing, ICE), предложенный Сакураи и соавт. с небольшими модификациями [106]. Реакция заключалась в том, что акрилонитрил присоединялся к инозину в положении N¹ с образованием N¹-цианоэтилинозина. Цианоэтилированные остатки блокировали синтез кДНК с этой молекулы РНК в качестве матрицы. Для идентификации инозинов в образце РНК обрабатывали акрилонитрилом. Параллельно, необработанная РНК готовилась в качестве контроля для последующих этапов процедуры. Затем с этих молекул РНК синтезировали цепи кДНК с помощью реакции обратной транскрипции и ПЦР (англ. reverse transcription PCR, RT-PCR). Если остатки аденозина (А) в части транскриптов отредактированы до инозина (I), неотредактированный А и отредактированный I превращались в тимидин (Т) и цитозин (С) соответственно. В цианоэтилированной РНК синтез кДНК заблокирован, что не приводило к амплификации молекул, содержащих С, соответствующие сайту редактирования. Затем эти кДНК амплифицировали с помощью ПЦР и секвенировали. В необработанных образцах участки А→I редактирования обнаруживались как смешанные сигналы А/Г на хроматограмме. В то время как в цианоэтилированной РНК сигнал от гуанозина (G), исчезал и сайт главным образом представлялся как А.

Для идентификации инозинов в этой работе тотальная РНК была выделена из 100 голов плодовых мушек с использованием набора RNeasy (Qiagen, Германия). Два образца, содержащих по 10 мкг РНК, параллельно, обработали путем инкубации в 38 мкл раствора для цианоэтилирования (CE +) и контрольного раствора (CE-) при 70°C в течение 30 мин. После обработки РНК очищали с использованием набора RNeasy MinElute (Qiagen, Германия) и проводили обратную транскрипцию с использованием набора Low RNA Input Linear Amp Kit (Agilent Technologies, США). Участки из девяти генов *D. melanogaster* (*Syx1A*, *Atx2*, *Atpalpha*, *stol*, *Cpx*, *EndoA*, *Alpha-Spec*, *Cadps*, *RhoGAP100F*) были проанализированы методом секвенирования по Сэнгеру, описанным в предыдущем разделе. Согласно [106] сайт редактирования в состоянии CE- обнаруживали как объединенный сигнал от А и Г на хроматограмме. Для состояния CE+ сигнал обнаруживали как А. Для тех участков, транскрипты которых редактируются почти на 100%, в состоянии CE+ амплификации кДНК не происходило совсем.

Таблица 3. Характеристика праймеров, использованных в работе для секвенирования ДНК и РНК.

№	ген	координаты замены аденозина на инозин	прямой праймер 5'- 3'	обратный праймер 5'- 3'
1	<i>Syx1A</i>	chr3R:24103714:A/ G:M244V	DNA:	DNA:
			ATATAGATCGGGTCTGGATGAG	GGATACAGCGTCAACTGGA
			cDNA:	cDNA:
			ATATAGATCGGGTCTGGATGAG	AACTCGTCTGTGTTCACGC
2	<i>Atx2</i>	chr3R:15413305:A/ G:K398R	DNA:	DNA:
			GGACGCGATCGTGACA	GTAGGAGTATTGACTCGGCAT
			cDNA:	cDNA:
			GGACGCGATCGTGACA	GTAGGAGTATTGACTCGGCAT
3	<i>Atpalpha</i>	chr3R:20965039:A/ G:Y390C	DNA:	DNA:
			AGAACTGTCTGGTGAAGAATCT	CAGAGCCAGTTCCATGCA
			cDNA:	cDNA:
			AGAACTGTCTGGTGAAGAATCT	CAGAGCCAGTTCCATGCA
4	<i>stol</i>	chr2L:15651359:A/ G:R489G	DNA:	DNA:
			GTCGATGTACTGGTTGGCA	TCGTTCAAGTCAACGATTACGA
			cDNA:	cDNA:
			GTCGATGTACTGGTTGGCA	TCGTTCAAGTCAACGATTACGA
5	<i>cpx</i>	chr3R:4297504:A/G :I125M	DNA:	DNA:
			ACATAACAGTTACAGCTACAGTAGA	GCTATGTTATCAGTATTACACGTGT
			cDNA:	cDNA:
			GAAGGAGGAGATCGTGGAGG	GCTATGTTATCAGTATTACACGTGT
6	<i>EndoA</i>	chr3R:18907675:A/ G:K137E	DNA:	DNA:
			GCGGTCAAGGGCATCT	GGAGCGATTACCGAACT
			cDNA:	cDNA:
			GCGGTCAAGGGCATCT	GGAGCGATTACCGAACT
7	<i>alpha-Spec</i>	chr3L:1782670:A/G :Q1700R	DNA:	DNA:
			AGATCGCGACCATAGTCGT	CACCTATATCGCTGCTGTCA
			cDNA:	cDNA:
			AGATCGCGACCATAGTCGT	CACCTATATCGCTGCTGTCA
8	<i>Cadps</i>	chr4:1265107:A/G; M1234V	DNA:	DNA:
			AATCGACAAAGCGAATGTAGC	CTGCTGCGAATACCACTGTTC
			cDNA:	cDNA:
			AATCGACAAAGCGAATGTAGC	AATCGACAAAGCGAATGTAGC
9	<i>RhoGAP 100F</i>	chr3R:31837086:A/ G:Q1142R	DNA:	DNA:
			CGGAGCGGAACAAGATGTCA	ATATTGCTCATGCTCCCGGC
			cDNA:	cDNA:
			CGGAGCGGAACAAGATGTCA	CGGAGCGGAACAAGATGTCA

3.10. Составление видоспецифических баз данных редактирования

Геномные координаты сайтов редактирования РНК плодовой мушки, сопоставленные с экзонами, были получены из базы данных RADAR (1328 сайтов) [110] и транскриптомных исследований, выполненных Сен-Лораном и соавторами [134]. Два списка сайтов редактирования РНК были использованы из предыдущих работ [134]. Первый, названный TIER1 содержал 645 экзонных несинонимичных сайтов редактирования. Второй список из той же работы, названный TIER2, содержал 7986 сайтов. Геномные координаты, полученные из этих источников, были преобразованы в координаты недавней сборки генома *Drosophila* Dm6 с использованием FlyBase (<http://flybase.org/>), [135]. Изменения в последовательностях белков, вызванные редактированием РНК, были аннотированы для всех трех списков (RADAR, TIER1 и TIER2) с использованием интегратора аннотаций вариантов (Variant Annotation Integrator; VAI; <http://genome.ucsc.edu/cgi-bin/hgVai>) [136].

Входные данные были подготовлены с использованием скрипта, написанного на Python. Выходные данные VAI использовались для создания баз данных белков, содержащих кодируемых в геноме и перекодированные белки плодовой мушки, с использованием другого скрипта, написанного в программе Python. Эти белковые базы данных были использованы для поиска в программах MSGF + и X! Tandem.

Геномные координаты мыши для редактирования РНК были получены из баз данных RADAR [110] и DARNED [111] и из дополнительных данных недавней работы [21], где 11103 сайтов редактирования РНК в экзонах были идентифицированы во взрослых и эмбриональных тканях мыши. Изменения в белковых последовательностях, вызванные редактированием РНК, были аннотированы по сборке генома мыши версии mm9 с использованием VAI.

Геномные координаты сайтов редактирования РНК человека были получены из тех же трех источников. В крупномасштабном исследовании были идентифицированы 12871 сайтов в экзонах в 8551 образцов тканей взрослого человека и плода проекта Genotype-Tissue Expression [21]. Изменения в белковых последовательностях, вызванные редактированием РНК, были аннотированы для сборки человеческого генома версии hg19 с использованием VAI. Файлы FASTA для протеогеномного поиска содержали 609 и 1878 перекодированных сайтов для мыши и человека соответственно.

3.11. Протеогеномный поиск

Идентификация пептидов в наборах данных взрослых особей *Drosophila melanogaster* была выполнена с использованием MSGF + версии 2017.01.1333, а также с помощью программы X!Tandem версии 2012.10.01.134 по трем настроенным базам данных, объединенным с базой данных UniProt для *D. melanogaster*, загруженной в апреле 2017 года и содержащей 42519 записей.

Поиски в tandemном режиме проводили с использованием допусков по массе 10 ppm и 0,01 Да, установленных для ионов-предшественников и фрагментов соответственно. Для наборов данных, полученных с использованием линейной ионной ловушки и взятых из работ Ксинга и соавторов [137] и Арадски и соавторов [138], допуск по массе фрагмента составил 0,3 Да. Поиски производились с допущением до 1 пропущенного сайта расщепления трипсином. Те же параметры использовались для поиска MSGF +, за исключением специфичных для прибора параметров, указанных для конкретного набора данных. Карбамидометилирование цистеина использовали в качестве фиксированной модификации. Окисление метионина и ацетилирование N-конца белка использовали в качестве переменных модификаций. Устанавливали 1% уровень ложноположительных результатов (FDR, false discovery rate). Использовали метод группоспецифичного FDR, который обеспечивал отдельный расчет FDR для вариантов пептидов, описанный в [138]. Расчет FDR с использованием базы данных, содержащей теоретически ожидаемые и заведомо ложные последовательности (target-decoy method) производился согласно следующему уравнению [139]:

$$\text{FDR} = \frac{\text{количество PSM с заведомо ложными последовательностями} + 1}{\text{количество PSM теоретически ожидаемых последовательностей}}$$

Все перекодированные пептиды отфильтровывали до группоспецифичного FDR 1%. Фильтрация выполнялась с использованием скрипта Python со специальной библиотекой Pyteomics, созданной для анализа протеомных данных [140]. Группоспецифичная фильтрация проводилась отдельно для пептидов из баз данных RADAR, TIER1 и TIER2.

Для открытого поиска (open search) использовали X!Tandem с теми же настройками, за исключением допуска массы предшественника, который устанавливали на 500 Да. Соответствия пары теоретический пептид — настоящий спектр (peptide-spectrum match, PSM), включая ложные идентификации, были сгруппированы по сдвигам масс с точностью 0,01 Да. Открытый поиск проводился только по данным протеома из нашего собственного эксперимента, так как два других набора данных содержали tandemные спектры, полученные с использованием линейной ионной ловушки с низким разрешением.

Идентификация пептидов в наборах данных стадий жизненного цикла *Drosophila melanogaster* выполнялась следующим образом. Файлы формата «RAW» преобразовывали в mzML с помощью MSConvert [141]. Затем во всех файлах выполняли поиск с помощью IdentiPy [142] с использованием следующих параметров: карбамидометилирование цистеина в качестве фиксированной модификации и окисление метионина в качестве вариабельной. Включалась автонастройка параметров с 10 ppm и 0,01 Да в качестве начальных значений допусков по массе ионов-прекурсоров и фрагментов; для деконволюции химерных спектров и извлечения информации об интенсивности на уровне MS1 [143] использовали программу Dinosaur. Результаты поиска подвергались постобработке с помощью Scavager [144] в двух режимах: первый - для получения списков проверенных PSM, пептидов и белков, отфильтрованных до 1% FDR, и второй - для группоспецифичной фильтрации идентификаций, содержащих сайты редактирования РНК. Для группоспецифичной фильтрации коррекция FDR [139] была отключена, что позволяло получить более высокие истинные значения FDR. Идентификационные данные из всех файлов, соответствующих каждой стадии развития, объединялись на уровне PSM для постобработки.

После этого для каждого набора данных строились таблицы пептидов, содержащих сайты перекодирования и идентифицированные по крайней мере в одном образце. В качестве количественной меры для идентифицированных пептидов использовали интенсивность MS1, извлеченную с помощью Dinosaur. Для количественной оценки белков дикого типа, соответствующих этим пептидам, были использованы значения NSAF (normalized spectral abundance factors, индекса нормированных спектральных коэффициентов обилия), полученные с помощью Scavager. Далее значения были преобразованы в процентные ранги, чтобы исключить систематические отклонения между образцами. Таблица пептидов содержала следующую информацию о каждом пептиде: геномные координаты сайта редактирования РНК, количество PSM и максимальную интенсивность MS1 в каждом образце.

Первичные идентификаторы участков с событиями перекодирования отфильтровывали. Сначала результаты вручную проверялись на возможные ошибки в базе данных с перекодированными белками. Затем триптические пептиды, содержащие замену аспарагина на аспарагиновую кислоту (N> D), исключали, поскольку это изменение неотлично от естественного дезаминирования аспарагина. Идентифицированные одиночные пептиды с недорезами без их полностью расщепленных трипсином аналогов, содержали больше ложноположительных результатов. Таким образом, пептиды этого типа также были исключены из результатов. Идентификации пептидов с одним PSM во всех данных были сочтены сомнительными.

Кроме того, ко всем пептидам применялся алгоритм DeepRT [145]. Прогнозируемое время удерживания линейно преобразовывали для согласования с экспериментальным временем в каждом экспериментальном запуске отдельно, с использованием лучших 200 PSM (по количеству баллов) из каждого файла для поиска оптимального линейного преобразования. Разница между прогнозируемым и экспериментальным временами удерживания (RT) была преобразована в Z-балл для каждого PSM путем его нормализации со стандартной ошибкой линейной регрессии, сравнивающей времена удерживания, предсказанные с помощью DeepRT, с экспериментальными. Затем были отфильтрованы совпадения с Z-оценкой > 2,5. Высокие значения Z-показателя соответствовали выбросам экспериментальных и прогнозируемых времен удерживания, построенных в виде линейной корреляции.

Для каждого идентифицированного перекодированного пептида и соответствующего канонического пептида экспериментальный MS/MS спектр с наивысшей оценкой был извлечен и аннотирован с использованием Spectrum-utils [146]. Кроме того, спектры фрагментации для каждого пептида (не более 30 аминокислот) предсказывались с использованием программы Prosit [147] «Prosit_2020_intensity_preview» (<https://www.proteomicsdb.org/prosit/>). Для пептидов, содержащих более 30 аминокислот, теоретические спектры моделировались путем подсчета всех b- и y-ионов с зарядами от одного до наблюдаемого состояния заряда пептида - 1 с использованием Pyteomics [140,148]. Типовые результаты представлены на рис. П2 и П3 Приложения.

Базы для поиска событий перекодирования мыши и человека составлялись с учетом всех триптических пептидов, содержащих два пропущенных сайта расщепления (недореза) и один сайт, перекодированный в результате редактирования РНК (35878 мышинных пептидов и 14234 человеческих пептидов). Данные аннотировались исходным идентификатором белка, геномными и протеомными координатами сайта редактирования. Ложные версии этих пептидов (decoy) конструировались путем переворачивания пептидных последовательностей при сохранении C-концевого остатка на месте. Канонические белковые последовательности из баз данных Mouse UniProt (май 2014 г., 57260 записей) и Human SwissProt (октябрь 2018 г., 20411 записей), соответственно, добавлены вместе с их ложными последовательностями, полученными путем изменения последовательности белка при сохранении N-концевого остатка на месте.

Поиск проводился по построенным базам данных с использованием программы X!Tandem Cyclone [149] со следующими основными параметрами поиска: допустимая ошибка массы для прекурсорных и фрагментных ионов была 10 ppm и 0,01 Да соответственно; один разрешенный пропущенный сайт расщепления; гидролиз трипсином с исключением расщепления перед пролином; карбамидометилирование цистеина в качестве фиксированной модификации и

окисление метионина в качестве вариабельной модификации. Кроме того, вариабельная модификация 229,162932 Да была применена к лизинам и N-концам в поиске данных с идентификатором PXD004143, которые помечались меткой tandemной массы (tandem mass tag, TMT). Затем с использованием Pyteomics [148] PSM, соответствующие пептидам с сайтами редактирования РНК, извлекали и отфильтровывали до группоспецифичного FDR 1% с использованием target-decoy метода.

Результаты поиска объединяли для биологических и технических повторов. Фракции, соответствующие ткани одного и того же типа, были объединены перед фильтрацией.

Отредактированные пептиды, идентифицированные автоматически, проверяли вручную на предмет подозрительных совпадений. Следующие элементы были исключены:

- замены Асн-Асп, поскольку события редактирования могут быть имитированы химическим дезамидированием;
- недорасщепленные пептиды, не имеющие своих расщепленных аналогов;
- для человеческих образцов события перекодирования, соответствующие известным геномным SNP, за исключением раковых мутаций (по базе данных NextProt [150]);
- химические модификации, имитирующие замены, которые были обнаружены при ручной проверке масс-спектров.

Для идентифицированных перекодированных пептидов канонические белковые формы определяли количественно с помощью индекса нормированных спектральных коэффициентов обилия (normalized spectral abundance factors, NSAF [151], обрабатывая выдачу X!Tandem с помощью Scavenger [144] и затем используя скрипт scav2nsaf для расчёта значения NSAF по результатам объединенной фракции. Для сравнения разных образцов все значения NSAF преобразовывали в процентиля.

3.12. Белок-белковые взаимодействия и визуализация данных

Белок-белковые взаимодействия получали с помощью ресурса STRING [152]. Минимальную необходимую оценку взаимодействия устанавливали на уровне средней достоверности (0,4), в качестве статистического фона использовался весь геном.

Строки тепловой карты располагались в соответствии с иерархической кластеризацией. Корреляция Пирсона между значениями интенсивности MS1 использовалась в качестве метрики расстояния (определяемой как 1-R). Строки дендрограммы переупорядочивались по значениям,

пропорциональным MS1 и стадии развития, когда белок был обнаружен. Тепловые карты были построены с использованием пакета ComplexHeatmap [153].

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

4.1. Поиск сайтов перекодирования в протеоме плодовой мушки

Тандемные масс-спектрометрические данные были взяты из недавних протеомных исследований Ксинга и соавторов [137] и Арадски и соавторов [138], доступных на ProteomeXchange. Эти результаты содержали данные для протеома всего тела дрозофилы [137] и целых голов [138]. Другой набор данных протеома мозга плодовой мушки был получен с помощью масс-спектрометрии высокого разрешения. Отметим также, что протеом, характеризуемый Арадской и соавторами [123], содержал только мембранные белки, так как они были преднамеренно извлечены во время пробоподготовки. В общей сложности для анализа в этом исследовании были доступны три набора данных, представляющих протеомы всего тела, головы и мозга дрозофилы.

Далее был произведен поиск сайтов редактирования РНК с использованием протеогеномного подхода [154]. Следуя этому подходу, стандартную базу данных FASTA протеома плодовой мушки расширяли за счет добавления отредактированных вариантов белков, найденных по транскриптомам. Три файла FASTA, содержащие базы данных белков с отредактированными последовательностями, полученными из результатов секвенирования транскриптома [134] и использования FlyBase (<http://flybase.org/>) [135], были сгенерированы, как описано в разделе 3.10, и названы TIER1, TIER2 и RADAR. Эти файлы FASTA были объединены с базой данных UniProt для *D. melanogaster* (версия от апреля 2017 г., 42519 записей), содержащей неотредактированные пептиды плодовой мушки, и использованы для поисков. После этого была проведена группоспецифичная фильтрация для пептидов из баз данных RADAR, TIER1 и TIER2. На рисунке 1 схематически показан рабочий процесс, используемый в этом проекте. Идентификаторы перекодированных пептидов перечислены в таблице 4. Геномные координаты и идентификаторы пептидов в базе UniProt перечислены в таблице П4 Приложения.

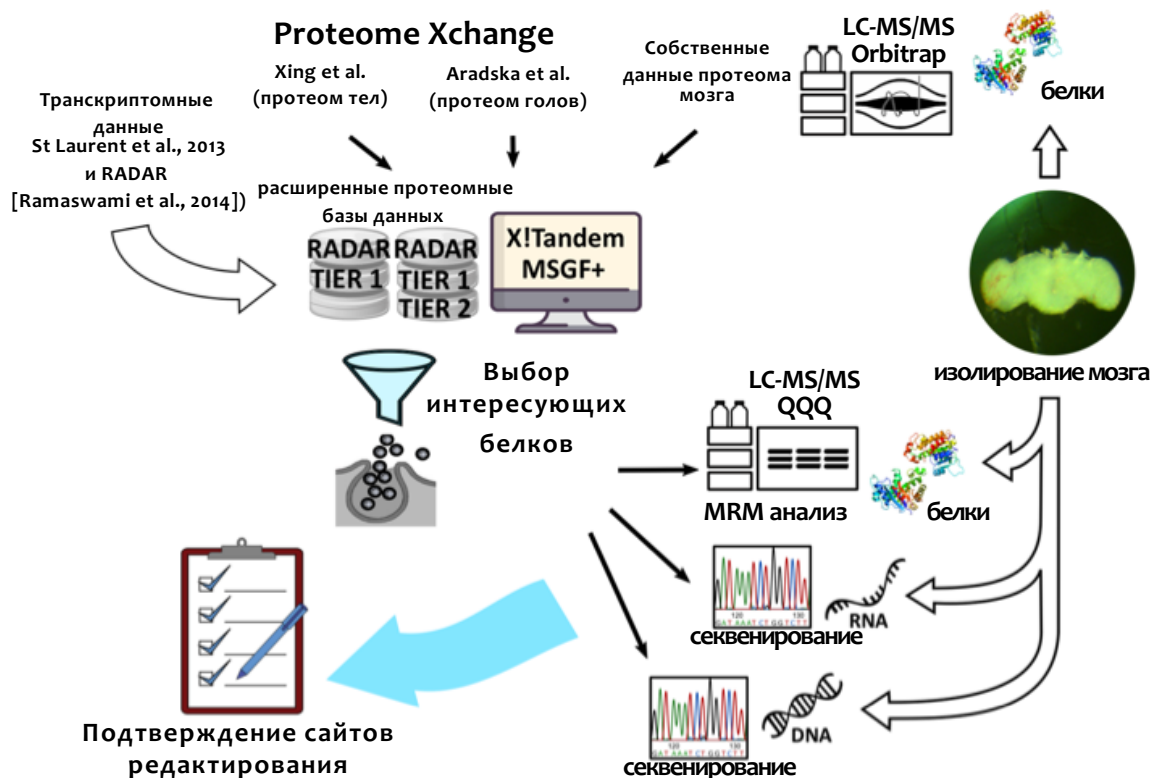


Рисунок 1. Схема протеогеномного подхода. Три набора данных, включающие протеом целых тел, голов и изолированных мозгов плодовых мушек, полученные панорамным MS-анализом были обработаны с использованием расширенных протеомных баз данных. Эти FASTA файлы были дополнены последовательностями, содержащими сайты редактирования. Поиск проводился двумя поисковиками X!Tandem и MSGF+. Всего было найдено 54 белка, содержащих сайты перекодирования. Из них были выбраны кандидаты, найденные во всех трёх наборах данных и объединенные функционально. Из [26], с изменениями.

Во всех исследованных данных было идентифицировано 62 пептида, соответствующих событиям редактирования РНК (таблица 4). Эти пептиды представляли собой 54 белка, поскольку некоторые из них, согласно транскриптомным данным, могли содержать несколько сайтов перекодирования [134]. Однако, поскольку масс-спектрометрия обеспечивает только частичное покрытие последовательности белка, некоторые из событий перекодирования белка могут отсутствовать в результатах поиска панорамного протеомного анализа.

Среди перекодированных сайтов, перечисленных в таблице 4, только 8 идентифицированы во всех наборах данных, использованных в исследовании. В эту группу, как и ожидалось, входили белки, специфичные для центральной нервной системы, поскольку собственные данные протеома получены из изолированного мозга. Другие 25 сайтов найдены как минимум в двух наборах данных. Оставшиеся 35 сайтов были идентифицированы только в одном исследовании.

Таблица 4. Пептиды, соответствующие событиям редактирования РНК, найденные в протеоме плодовой мушки.

Последовательность пептида	Ген	Белок	A/к замена	Арадска и соавт. RADAR	Арадска и соавт. TIER1	Арадска и соавт. TIER2	Ксинг и соавт. RADAR	Ксинг и соавт. TIER1	Ксинг и соавт. TIER2	Собственные данные, RADAR	Собственные данные, TIER1	Собственные данные, TIER2	Подтверждение с помощью MRM
ADSLVESGQFDTAGIQEK	alpha-Spec	M9PGV6	Q>R		✓			✓			✓		
LVTTVSTPVFDGR	stol	A8DZ06	R>G		✓			✓			✓		Да
IEYHVEHAVDYVQTATQD TK	Syx1A	Q24547	M>V		✓			✓			✓		Да
GVGPAPSANASADSSSR	Atx2	Q8SWR8-2	K>R		✓		✓			✓			Нет
LVSVLESTLSK	Cadps	Q9NHE5-2	M>V		✓		✓			✓			Да
NQMETQVNELK	cpx	E1JJ33	I>M		✓		✓			✓			Да
YSLDDNIEQNFLEPLHHM QTK	EndoA	B5RIU6	K>E		✓		✓			✓			Да
MTVAHMWFDNQIEADTT EDQSGVQCQR	Atpalph a	P13607-7	Y>C	✓			✓			✓			
YLLQIWPQQAQHQQR	RhoGA P100F	Q9V9S7	Q>R	✓						✓			Нет
AIAGVLTPLNPLWALR	CG1882	Q5U191	R>G	✓	✓		✓						
EDIQANQLVMGEFEVSDV PGQIIDYVAR	bai	Q8SXY6	I>V		✓		✓						
FIYVSEEEELLAVAK	CG5862	Q9VDD1	K>E		✓		✓						
GALSSAGYHLNNR	CalpB	M9NE73	N>S		✓		✓						
LHAGDDEEADDPIR	Calx	Q9VDG5	Q>R		✓		✓						
TAGPLPGGR	CG1050 8	M9PFU5	S>G		✓		✓						
AIHALAAGASSSAEQDGA YNPWPHFR	CG1363 0	Q9VC48	N>S			✓			✓				
DLVATVLDMK	eag	Q0KHS8	I>V		✓		✓						
ALQELEHIDDEADQLGIGF VK	l(2)0128 9	Q0E9N2	N>H		✓		✓						
LIILCVEALGQDR	Mp	E1JIE1	S>G	✓						✓			
LAIVEYMPALAGQLGQEF FDQK	Pp2A-29B	H5V8B8	I>V		✓		✓						
DPWPVPR	Sec23	Q9VNF8	Q>R		✓		✓						
CTLNPYYNESFSFEVPFEQ MQK	Syt1	P21521-4	I>M		✓			✓					
KKQQGGGGGGSGGGGGS GGGGGS	shf	Q9W3W5	N>S			✓		✓					
LAVHEIVDANGAGDAFV GGFLSQFVQGK	AdenoK	Q9VU39	T>A					✓					
IIHLLNDQHMGVVTAAS LIDALVK	AP-2alpha	P91926	T>A					✓					
LQSSVSGPGDPVR	brp	A1Z7V1	S>G							✓			
TVGETETFLTGTNEITIK	CG1007 7	Q7KU78	S>G								✓		
NQIETQVSELK	cpx	E1JJ33	N>S							✓			
LTEPLAPLK	hig	Q09101-2	Q>R								✓		

SLLESSLHR	Prm	A4V1N8	Q>R					√					
LAASQQVAVVHQHPSDQ AGESGHK	sif	Q8IQ62	R>G							√			
HDGGVIQVDDR	Unc-89	A8DYP0	E>G					√					
RKPPQTPAEGEGDPEFIK	up	M9NH07	K>R									√	
AEKPEPTENVLEPLHR	CG2267	Q9VA51	K>E									√	
EAPILVVAPHSSYVDSILV VASGPPSMVAK	CG3269	Q0KHU5	I>M		√								
LPVNLGLSSILVPHGVDLD EPDVK	stol	A8DZ06	S>G		√								
ALELIAER	CG8783	Q9VUB4	T>A					√					
PQLVNNYSYVDDK	dj-1beta	Q9VA37	K>R						√				
IGHSLR	Dyb	Q0E9B2	Q>R							√			
VDLLFR	GluClal pha	E1JIQ2	K>R							√			
GLTNSIAVVATSTCPR	hig	Q09101	Q>R								√		
MDITWFPFDDQHCEMK	nAChRa lpha6	Q8IPE2	I>M		√								
KTAGGSLEMFMTEDQK	para	M9MS79	K>T							√			
LSFRASGGSNGR	Pde1c	B7YZV2	S>R							√			
VEIELDSVR	Prm	A4V1N8	Q>R					√					
GFQPGPSVVG	pyd	A8JQV8	S>G		√								
NTCVFGGAPR	Rm62	P19109-2	K>R								√		
REGTTS GDKR	rnhl	A1Z768	K>E			√							
VTKGAGATGAAGK	RpL34b	Q9VHE5	K>T							√			
SSNQAMSLAMLR	Sh	P08510-3	I>M		√								
MDPMIYMFPR	shakB	P33085	I>M		√								
MALTSYDQIPSELAPYAF VEFTMPATQVSHATVR	stnB	X2JGF5	T>A					√					
EGLLQR	Syn	E2QCY9	S>G							√			
QQLGSILMSAPR	Tfb4	Q9VPX4	N>S									√	
DHCIAMVQCR	VhaSFD	Q9V3J1	K>R					√					
DHCIAMVQCTVLKQLSIL EQR	VhaSFD	Q9V3J1-3	K>T							√			
EQLATLMDTLR	zip	Q99323-1	K>T				√						
AHSDLTGVK	Calx	Q9VDG5	E>G				√						
SHHSSSRGKYER	CG3121	Q86B99	S>G	√									
QDSPSQPTIVVKDSSNAK LNHTK	Cp190	Q24478	I>V			√							
LKNQMETQVNELK	cpx	E1JJ33	I>M				√						
APNPLVYDPK	Cyp6a8	Q27593	T>A						√				
QQQQLQHSQQLPR	InR	P09208	N>S						√				
FLAGEADKK	Tm1	P49455-2	E>G				√						
ASTLSSSTSPSMSSSSASN HPGHSQSQR	Rab3- GEF	Q9VXY2	Q>R				√						
IMTFVGGPCSR	Sec23	Q9VNF8	Q>R				√						
RQTDGQWDLNGYHLVTL GDR	syd	Q9GQF1- 2	S>G		√								
IAVMQNGKRLK	Syt1	X2J4C1	I>V		√								

4.1.1. Открытый поиск для дополнительной фильтрации ложно-положительных результатов

Открытый поиск был выполнен для того, чтобы присвоить некоторый уровень достоверности конкретной аминокислотной замене. Такие замены могут происходить не только из-за редактирования РНК, но и могут быть вызваны модификацией аминокислоты, что в данном случае будет являться ложным результатом [155]. Полученные данные открытого поиска представлены в таблице П4 Приложения. Каждая аминокислотная замена имеет свой ранг открытого поиска, который рассчитывался на основе процента идентификаций, попадающего в соответствующий интервал сдвига массы. Если бы этот сдвиг массы был избыточно представлен, скорее всего, он соответствовал химической модификации, а не аминокислотной замене. Таким образом, ранг открытого поиска косвенно представляет собой вероятность изменения массы, вызванного модификацией. Основываясь на подходе открытого поиска, пептиды, содержащие замены аспарагина на аспарагиновую кислоту, были удалены из результатов, поскольку данное изменение не может быть объяснено исключительно дезаминированием ферментами ADAR. Открытый поиск показал, что некоторые замены имеют более высокую вероятность повторения за счет модификаций, чем другие.

4.1.2. Функциональная аннотация отредактированных белков, обнаруженных в данных панорамной протеомики

Для понимания функциональной значимости редактирования РНК список пептидов, показанный в таблице 5, анализировали с помощью системы функциональных белковых взаимодействий (STRING, версия 10.0 <http://string-db.org/>), [156]. На рисунке 2 показаны результаты анализа STRING. Найдены две группы белков, аннотированные на основе анализа биологических процессов элементами Gene Ontology (GO). В этой части не было использовано программного обеспечения, которое автоматически рассчитывало обогащение по определенным элементам GO, поскольку в нашем случае было трудно определить контрольную группу генов для этого анализа.

В первую группу входили следующие участники: синтаксин (*Syx1A*), синаптотагмин (*Syt1*), комплексин (*cpx*), синапсин (*Syn*), адаптин (AP-2 альфа), эндофилин А (*endoA*), stoned protein В (*StnB*), белок bruchpilot (*brp*), кальций-зависимый активатор секреции (*Cadps*) и субъединица кальциевых ионных каналов *stol* (CG4587). Эти белки играют роль в синаптической передаче сигнала. В частности, синтаксин и синаптотагмин являются компонентами комплекса SNARE, который обеспечивает слияние синаптических везикул с пресинаптической мембраной,

а комплексин - важный дополнительный связующий компонент этого комплекса [157]. Белок StnB также связывает этот комплекс и играет роль в эндоцитозе синаптических пузырьков [158]. Адаптины участвуют в процессе рециклинга синаптических пузырьков [159]. Эндофилин (endoA) действует в процессе эндоцитоза везикул в нервно-мышечной системе [160–162]. Кальций-зависимый активатор секреции (Cadps) является Ca^{2+} -зависимым фактором эндоцитоза везикул [163,164]. Группа синаптической передачи сигнала содержала пять белков, перекодированных во всех исследуемых масс-спектрометрических данных и представляла наиболее часто редактируемые белковые продукты.

Для второй группы показана связь с цитоскелетом. В нее включены белки немышечного миозина (zip), альфа-спектрина (alpha-Spec), титина (sls) и других генных продуктов (рис. 2). Все белки этой группы либо являлись компонентами цитоскелета, либо взаимодействовали с ними и участвовали в процессах клеточного транспорта. Эта группа включала только один перекодированный сайт, идентифицированный во всех трех исследованных наборах данных, который принадлежал белку альфа-спектрину.

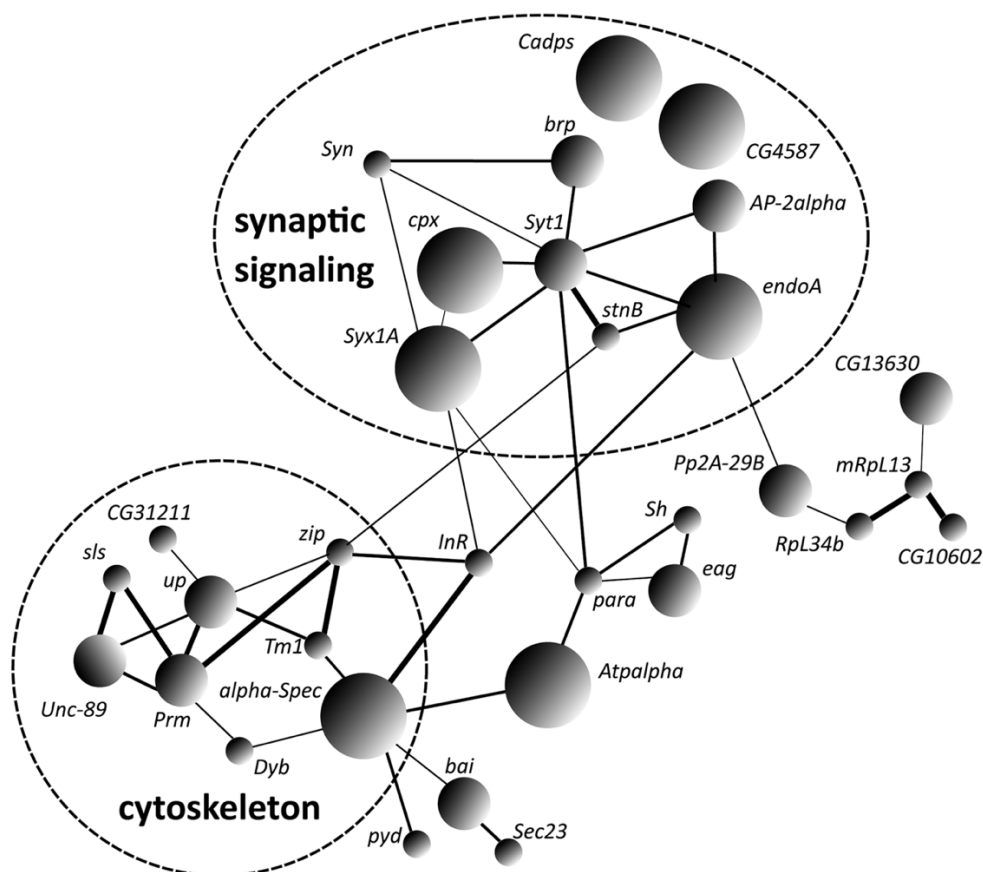


Рисунок 2. Белок-белковые взаимодействия перекодированных вследствие редактирования РНК продуктов, полученные из базы знаний STRING. Первая группа – белки, участвующие в синаптической передаче сигнала (synaptic signaling). Вторая группа – белки цитоскелета (cytoskeleton). Толщина линий показывает количество методов, подтверждающих каждое взаимодействие согласно STRING. Из [26], с изменениями.

4.1.3. Генотипирование сайтов геномной ДНК, соответствующих найденным событиям редактирования РНК

Восемь сайтов, идентифицированных панорамной протеомикой во всех исследованных наборах данных, были дополнительно проверены на ADAR-опосредованное редактирование на уровне нуклеиновых кислот. Сайт гена *RhoGAP100F*, обнаруженный в двух наборах данных, был добавлен в список из-за тесной функциональной связи с другими редактируемыми белками. Полиморфные сайты девяти отобранных генов *D. melanogaster* (M244V в *Syx1A*, K398R в *Atx2*, Y390C в *Atpalpha*, R489G в *stol*, I125 M в *cpx*, K137E в *EndoA*, Q1700R в *AlphaSpec*, M1234 V в *Cadps* и Q1142R в *RhoGAP100F* генотипировали. Анализ показал отсутствие в геноме полиморфных сайтов A/G. Этот факт подтверждает, что замены произошли в ходе посттранскрипционной модификации, а обнаруженные аминокислотные замены в пептидах являются следствием редактирования РНК. Результаты представлены в таблице 5.

4.1.4. Идентификация событий редактирования на уровне мРНК

Сайты девяти генов, выбранных, как указано выше, анализировали методом химического удаления инозина (ICE) [106] в двух биологических повторях. После сравнения секвенирования кДНК из химически удаленного (CE +) акрилонитрилом инозина и интактной (CE-) РНК, только в четырех из девяти сайтов-кандидатов, инозин-содержащие сайты были подтверждены химическим методом, включая *Atpalpha* (chr3R: 20965039: A/G: Y390C), *Cadps* (chr4: 1265107: A/G: M> V), *Cpx* (chr3R: 4297504: A/G: I125M) и *Syx1A* (chr3R: 24103714: A/G: M244 V) (таблица 5).

Таблица 5. Результаты секвенирования геномной ДНК в участках, соответствующих сайтам редактирования РНК и результаты идентификации замен аденозина на инозин методами химического удаления инозина и таргетной протеомики

Ген	Сиквент ДНК	Сиквент амплифицированной кДНК		Событие редактирования, подтвержденное ICE	Аминокислотная замена, подтвержденная MRM
		CE–	CE+		
<i>Syx1A</i>	T	A/G	A	+	+
<i>cpx</i>	T	A/G	–	+	+
<i>Atpalpha</i>	T	A/G	A	+	н/д
<i>stol</i>	T	A	A	–	+
<i>alpha-Spec</i>	T	A	A	–	н/д
<i>Cadps</i>	T	A/G	A	+	+
<i>EndoA</i>	T	A	A	–	+
<i>Atx2</i>	T	A	A	–	–
<i>RhoGAP100F</i>	T	A	A	–	–

4.1.5. Проверка и количественная оценка отредактированных сайтов на уровне белка с помощью мониторинга множественных реакций (MRM) с использованием меченых стабильными изотопами стандартов пептидов

После панорамного протеомного анализа данных *D. melanogaster* в качестве продолжения исследования был проведен таргетный анализ наиболее интересных сайтов редактирования. Также требовалось проверить расхождение между результатами протеомики и исследованием РНК, обнаружившим только четыре из девяти сайтов, представляющих интерес, как инозин-содержащие.

Для анализа выбраны те сайты, которые были представлены в нашем наборе протеомных данных мозга, идентифицированы, по крайней мере, в еще одном наборе данных панорамного протеомного анализа для лучшей достоверности, а также функционально связаны с синаптической передачей сигнала, точнее, с путями, обеспечивающими слияние синаптического пузырька с мембраной пресинаптической клетки при потенциале действия [165].

Эти предположительно отредактированные белки включали синтаксин 1A (*Syx1A*) и комплексин (*Cpx*), известные как участник и связующий белок комплекса SNARE соответственно [166]. Другие представляющие интерес белки включали эндофилин А (*EndoA*), обеспечивающий формирование мембран в синапсах [160], *Cadps* и *stol*², участвующие в кальций-зависимой передаче сигнала в пресинаптической зоне [163,167,168] и активатор ГТФазы *RhoGAP100F*, участвующий в организации пресинаптического уплотнения [169]. Как упоминалось выше, последний белок был добавлен в список для таргетного анализа несмотря на то, что он был идентифицирован только в двух наборах данных. Кроме того, мы добавили к анализу сайт атаксина-2 (*Atx2*), хорошо изученного белка, участвующего в различных процессах, таких как развитие глаз, образование актиновых филаментов и регуляции циркадной активности [170,171].

Как было упомянуто выше, не все исследованные сайты-кандидаты удалось подтвердить на уровне РНК. Тем не менее, мы включили четыре таких неподтвержденных сайта в группу для проверки путем MRM. Как это показано в таблице 6, таргетный анализ в двух случаях подтвердил данные ICE-seq (*Atx2* и *RhoGAP100F*), где в гидролизатах мозга не было обнаружено никаких следов отредактированных или неотредактированных пептидов. Также результаты MRM предоставили хорошие доказательства для сайтов *EndoA* и *stol* (*CG4587*) как в кодированной в геноме, так и в отредактированной форме, несмотря на отсутствие подтверждения этих сайтов анализом РНК. Меченые стабильными изотопами стандарты

² В процессе подготовки работы ген *CG4587* получил название *stol* [168]

пептидов для каждого сайта вели себя аналогично природным гидролизованным продуктам протеома мозга в тестах MRM согласно спектрам отредактированных и неотредактированных вариантов пептидов *stol* и EndoA (рис. П1 Приложения). Это дало убедительные доказательства их химической идентичности. Отсутствие обнаружения сайтов редактирования путем химического удаления инозина может быть объяснено низким выходом желаемых продуктов обработки для этих определенных сайтов. Другими словами, содержание мРНК и белковых продуктов одного и того же гена не всегда коррелирует, по крайней мере, если их измерять в одномоментном состоянии без временной зависимости [172]. Гипотетически, продукты РНК представляющих интерес генов могут подвергаться быстрой деградации во время процедур с животными. Наконец, в трех сайтах, принадлежащих генам *Cadps*, *Cpx* и *Syx1A* редактирование было подтверждено как в транскриптоме, так и в протеоме.

С мечеными стандартами были измерены абсолютные концентрации целевых пептидов в образцах мозга. Тем не менее, на эти уровни могли повлиять многие факторы во время нескольких этапов подготовки образца. По нашим оценкам, для пяти представляющих интерес белков эти уровни варьировались в пределах одного порядка, приблизительно от 3,5 до 60 нмоль/г общего белка в *Cadps* и EndoA, соответственно (рисунок 3).

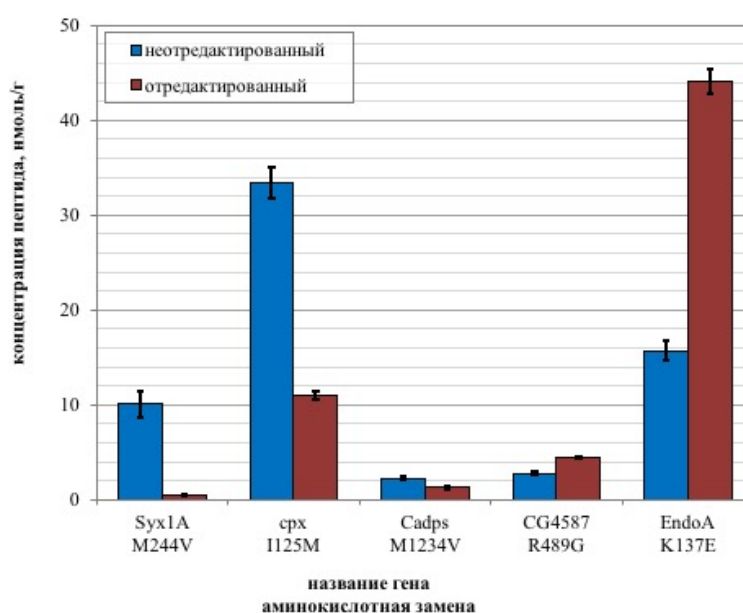


Рисунок 3. Соотношение концентраций отредактированных и неотредактированных вариантов пептидов синаптических белков в головном мозге плодовых мушек, выращенных при 25°C. Из [26], с изменениями.

Для целей этой работы было более уместно рассмотреть соотношение между неотредактированными и отредактированными формами белков, которое в меньшей степени зависело от пробоподготовки. Это соотношение варьирует между исследованными сайтами. Таким образом, в синтаксине 1A менее 5% белка оказались перекодированными. Другой

участник комплекса SNARE, комплексин, был перекодирован примерно на четверть. Уровень перекодирования продукта гена *Cadps* находился на третьем месте. Также было показано, что сайты субъединицы кальциевого канала *stol* (*CG4587*) и эндофилин А перекодируются на 62% и 74% соответственно (таблица 6). Последний белок привлек особое внимание, поскольку он был одновременно наиболее распространенным и наиболее отредактированным из изученных продуктов.

Таблица 6. Количественный анализ выбранных перекодированных участков и их геномных вариантов в гидролизате мозга плодовой мушки с помощью мониторинга множественных реакций (MRM) со стандартами, меченными стабильными изотопами. Жирным шрифтом выделены позиции аминокислотных замен, возникшие в результате редактирования РНК аденозиндезаминазами.

Название гена	Участок аминокислотной замены	Последовательность триптического пептида для анализа MRM	Подтверждение методом ICE	Содержание белка, нмоль/г (среднее±станд.отклон.)	CV, %	Процент отредактированных сайтов, %
<i>Syx1A</i>	M244V	Кодируемая в геноме	IEYHVEHAMDYVQ TATQDTK	Да	10,1±1,4	13,7
		Перекодированная	IEYHVEHA V DYVQ TATQDTK		0,5±0,04	7,7
<i>cpx</i>	I125M	Кодируемая в геноме	NQ I ETQVNELK	Да	33,4±1,6	4,9
		Перекодированная	NQ M ETQVNELK		11,1±0,5	4,1
<i>EndoA</i>	K137E	Кодируемая в геноме	YSLDD N IK	Нет	15,6±1,1	6,9
		Перекодированная	YSLDDN I EQNFLEP LHHMQTK		44,1±1,4	3,1
<i>stol</i>	R489G	Кодируемая в геноме	LVTTVSTP V FD R	Нет	2,8±0,2	6,7
		Перекодированная	LVTTVSTP V FD G R		4,5±0,1	1,5
<i>Cadps</i>	M1234V	Кодируемая в геноме	LMSVLESTLSK	Да	2,3±0,1	5,5
		Перекодированная	LVS V LESTLSK		1,3±0,2	12,5
<i>Atx2</i>	K337R	Кодируемая в геноме	GVGPAPSANASADS SSK	Нет	Не обнаружен	
		Перекодированная	GVGPAPSANASADS SS R			
<i>RhoGAP1 00F</i>	Q1142R	Кодируемая в геноме	YLLQIWQPQAQH R	Нет	Не обнаружен	
		Перекодированная	YLLQIWQPQAQH Q R			

Омиксные технологии - это мощные инструменты, предоставляющие исследователям большие массивы качественной и количественной информации. Таким образом, взяв изменения в кодирующей части транскриптома, вызванные его редактированием с помощью ADAR, мы

проверили их соответствие с протеомным уровнем. Однако мы можем только догадываться, являются ли выявленные изменения функциональными или представляют собой молекулярный шум как побочный эффект такой ферментативной активности. Хотя результаты омиксных методов интактных объектов носят описательный характер и сами по себе не могут дать биологического истолкования результатов, они представляют собой хороший инструмент для создания гипотез. Далее мы попытаемся обсудить гипотезы, которые могут возникнуть из подтвержденных протеомных данных ADAR-опосредованного редактирования.

Как упоминалось выше, синтаксин 1a и комплексин являются компонентами хорошо изученного комплекса SNARE, который обеспечивает слияние пресинаптической мембраны и синаптических пузырьков в большинстве организмов с нервной системой. Кальций-зависимый активатор секреции (CADPS или CAPS), кодируемый геном *Cadps*, также участвует во взаимодействии с этим комплексом [163]. SNARE является хорошо охарактеризованным комплексом, особенно для модельных млекопитающих [173]. Отредактированные сайты этих трех белков рассматривались с учетом их структурной организации и не были сопоставлены ни с одним из известных функциональных сайтов. В частности, известно, что человеческие и мышинные ортологи *Cadps* в значительной степени редактируются ферментом ADAR. Согласно базе данных RADAR, каждый из этих генов редактируется в десятках сайтов, по крайней мере два из которых кодируют аминокислотные замены [110].

Белок *stol* (CG4587), найденный в протеоме плодовой мушки, представляет собой продукт, участвующий в Ca^{2+} -зависимом ноцицептивном ответе [167]. Мало что известно о его пространственной структуре, так что трудно предложить роль замены R489G. Существует некоторая параллель этого типа замещения с эдитомом млекопитающих, где субъединицы глутаматных рецепторов (*Gria2* и *Gria4*) также несут аналогичные аминокислотные замены. Кроме того, как и в случае белка CAPS, человеческий ортолог *stol*, называемый потенциал-зависимой субъединицей кальциевых каналов альфа-2/дельта-4 (CACNA2D4), также был отредактирован, но в интронной области [105]. Примечательно, что, основываясь на наших данных, все же может быть определена пока неизвестная значимость этого генного продукта. *Stol* на 30% идентичен важному паралогу, белку Straitjacket (*stj*), который является еще одной субъединицей Ca^{2+} -зависимого канала, также участвующей в ноцицепции [174]. Белок *stj* не входит в наш список соответствующих событию редактирования РНК, но он был обнаружен в протеоме плодовой мушки и идентифицирован поисковыми системами в наших данных. Теоретически, *stj* может соответствовать событию редактирования, потому что его транскрипт подлежит редактированию ADAR, а белок находился в нашей видоспецифической базе данных редактирования. Тот факт, что *stj* не указан среди идентифицированных отредактированных

белков, можно объяснить его недостаточной представленностью в спектрах по сравнению со *stol*. Протеомы головы и головного мозга представляют наибольший интерес с точки зрения экспрессии Ca^{2+} -зависимых каналов. В протеоме мозга относительная интенсивность продукта *stj*, рассчитанная поисковой системой MaxQuant, в 34,7 раза ниже по сравнению с *stol*. Количественный анализ без использования изотопной метки (intensity-based label-free quantification, LFQ), полученный для результатов поиска X!Tandem, показал следующие результаты для соотношений LFQ: *stol/stj*: 12,4, 4,7 и 6,6 для алгоритмов расчета LFQ SIN [175], NSAF [176], и emPAI [177] соответственно. Для протеома целых голов соотношения *stol/stj* составляли 5,4, 4,4 и 8,5 для этих алгоритмов LFQ соответственно. Поэтому, хотя *stj* является важным и хорошо охарактеризованным белком нервной ткани, его паралог, продукт гена *stol*, имел значительно более высокое содержание в протеомах изолированного мозга, а также всей головы. Наряду с белком *stj*, *stol* заслуживает дальнейших функциональных исследований.

В отличие от всех других сайтов, представляющих интерес, о которых отсутствовала структурная информация, был найден сайт редактирования эндофилина А, где лизин-137 был заменен на глутамат с высоким содержанием отредактированной формы 74% (таблица 6). Этот остаток находится внутри $\alpha 2$ -спирали консервативного и хорошо изученного BAR-домена этого белка [178]. Домены этого типа образуют димеры и содержат много белков, связанных с динамикой межклеточной мембраны [179]. Было показано, что BAR-домен в эндофиле А обеспечивает изгиб мембраны, связывающей гидрофильную поверхность липидного бислоя вогнутой поверхностью его димера [180]. Этот остаток расположен вблизи поверхности или на ней, как это видно из пространственной структуры человеческого белка, которая в этой части почти идентична дрозофиле (рис. 4). Сродство между мембраной и доменом BAR эндофилина А обеспечивается, среди прочего, электростатическими взаимодействиями между отрицательно заряженными фосфолипидными головками и остатками лизина в вогнутой части. Таким образом, можно предположить, что смена заряда при замене лизина на глутаминовую кислоту в эндофиле А может существенно повлиять на аффинность связывания белка с мембраной, посредством чего регулируется динамика мембраны в нервных клетках. Эта предположение может быть подтверждено экспериментами с рекомбинантными белками и модельными мембранами. Влияние редактирования эндофилина А на организм насекомых может быть подтверждено путем генерации соответствующих мутантных штаммов.

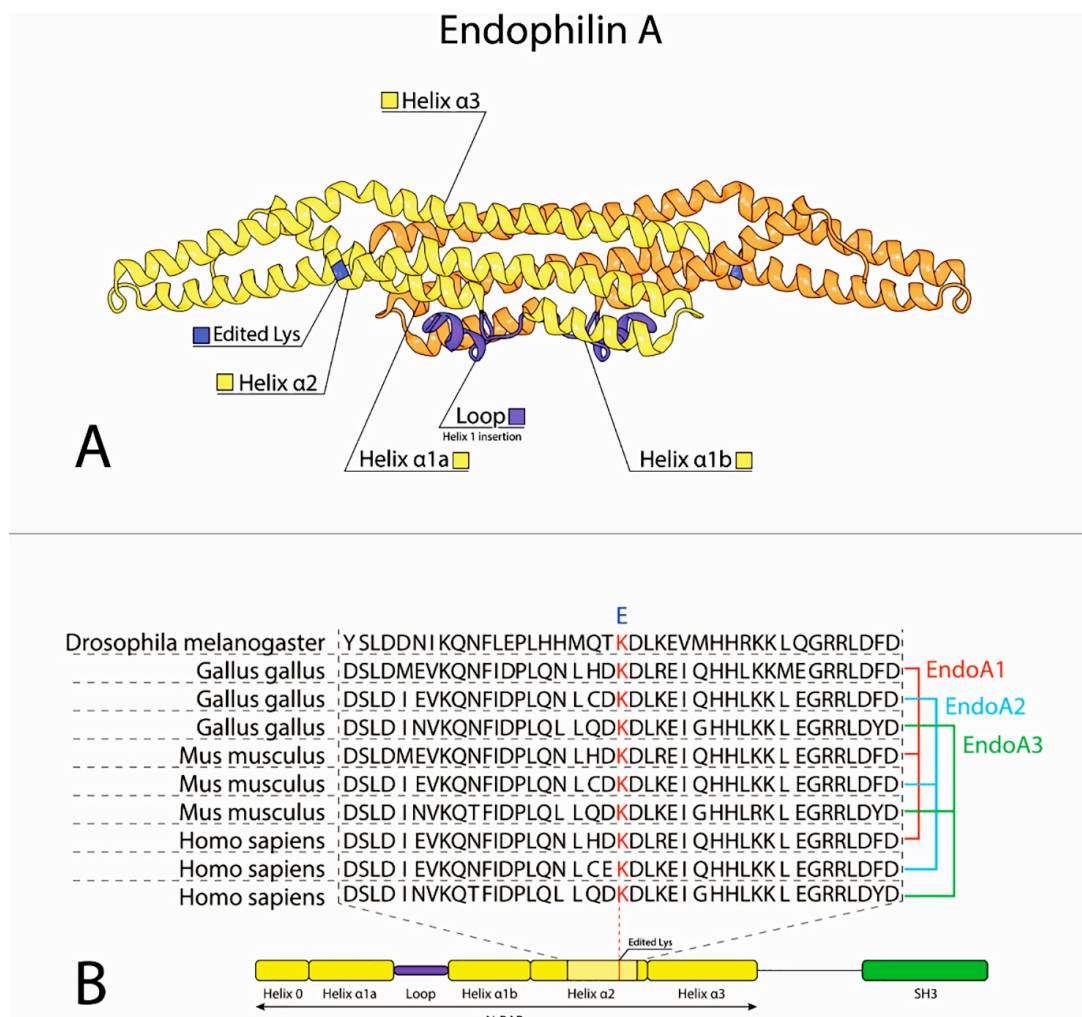


Рисунок 4. Редактирование мРНК с заменой лизина на глутаминовую кислоту. Пространственная структура домена BAR в эндофилине А плодовой мушки. **А.** Гомологичная позиция консервативного остатка лизина, соответствующего Lys-137 EndoA плодовой мушки показана в димере BAR-доменов ортолога человека (PDB 1 × 03) [180]. Структурные элементы домена BAR обозначены, как описано в [178] **В.** Локальные выравнивания между продуктом EndoA плодовой мушки и изоформами эндофилина А курицы, мыши и человека иллюстрируют высокий уровень консервативности в α2-спирали между организмами. Рисунок приводится по [26].

Примечательно, что транскрипты эндофилина А у млекопитающих, как показано, редактируются ферментами ADAR, но в их интронных частях. Напротив, для плодовой мушки наблюдалась обширная перекодировка белковой последовательности, которая, очевидно, должна оказывать функциональное влияние. Недавно было высказано предположение, что редактирование с заменой аденозина на инозин у первичноротых могло служить для настройки функции белков вместо геномных мутаций [119]. Возможно, редактирование *endoA* у плодовой мушки было функционально компенсировано у позвоночных путем эволюционного создания трех генных паралогов для эндофилина А.

4.1.6. Анализ протеома плодовой мушки на разных стадиях онтогенеза для выявления участков белка, перекодированных при редактировании ферментами ADAR

Первая часть данной работы, посвященная поиску следов редактирования РНК в протеоме мозга плодовой мушки [26,181], была сделана в то время, как опубликовали комплексную карту протеома развития *D. melanogaster* [182]. Эта карта была основана на панорамном протеомном подходе, совместимом с протеогеномным анализом, предложенным ранее [26]. Таким образом, мы повторно проанализировали новые протеомные данные, добавив аминокислотные последовательности, перекодированные редактированием, в обычную базу данных. В результате мы смогли дополнить динамику развития самого протеома, описанную в вышеупомянутой работе [182], с помощью знаний о динамике перекодированных белков [35].

Поиск полиморфизмов отдельных аминокислот, в том числе тех, которые возникают вследствие редактирования РНК, основан на совпадении пептидных спектров (PSM) в панорамных протеомных данных [155]. Для этого требуется надлежащая проверка полученных идентификаций, чтобы избежать ложноположительных результатов. Это особенно сложно, когда данные панорамной протеомики повторно анализируют без доступа к исходным образцам, как в этой части работы. Мы применили консервативную процедуру фильтрации, описанную в разделе 3.1.1, к результатам поиска, чтобы избавиться от сомнительных идентификаций. Кроме того, все идентификации с хроматографическим временем удерживания, сильно отклоняющимся от теоретически предсказанных значений, также были исключены, чтобы избежать дальнейших сомнений из-за возможности возникновения ложноположительных результатов [183].

Набор данных для повторного анализа состоял из двух подмножеств. Первое включало пятнадцать стадий развития плодовой мушки от яйца до имаго, а второе охватывало четырнадцать временных точек эмбриогенеза, то есть развития яйца [182]. Идентифицированные сайты редактирования ADAR перечислены в таблице 7. Всего было найдено 40 участков перекодирования, каждый из которых принадлежал уникальному белку. Из них только десять участков были идентифицированы в наборе данных по эмбриогенезу. Шесть из них совпадали с набором данных жизненного цикла, в котором также были четыре временных точки развития яйца. Список белковых сайтов, кодируемых редактированием, сравнивали с полученными в первой части работы протеомами целых тел, голов и мозга взрослых насекомых [26]. Около трети участков, наблюдаемых в настоящем исследовании, также были идентифицированы нами в первой части работы (таблица 7). Наблюдаемое различие можно объяснить наличием участков с

перекодировкой, специфичных для фаз цикла, которые не были включены в исследование взрослых насекомых.

Относительно короткий список идентифицированных белков, перекодированных редактированием, не показал значительного обогащения какого-либо молекулярного каскада после коррекции на множественное сравнение. Однако анализ белок-белкового взаимодействия с использованием базы знаний STRING [152] показал значительное количество взаимодействий между белками. Программа STRING рассчитала уровень значимости обогащения описанных в литературе белок-белковых взаимодействий в списке белков, перекодированных редактированием, равный 0,00002. Это показывает, что белки в списке имеют гораздо больше взаимодействий между собой, чем ожидалось для списка случайных белков. Таким образом, несинонимичное редактирование РНК не выглядит спонтанным и может модулировать работу ансамблей интеракторов, нуждающихся в такой регуляции.

Взаимодействующие белки, классифицированные по функциональным группам (таблица 7), представлены на рисунке 5. Одна группа участвует в организации структуры актомиозина, широко известными представителями которой являются титин и парамиозин. Другая группа взаимодействующих белков представляет участников путей транспорта везикул, функционально связанных с механизмом работы пресинаптической мембраны. Комплексин был добавлен к этой группе вручную, поскольку ранее было показано, что он является важным редактируемым компонентом комплекса SNARE. Как указано выше, этот комплекс отвечает за слияние пресинаптической мембраны и синаптических пузырьков [26]. Остальные перекодированные белки с известными характеристиками были разделены на более мелкие группы компонентов цитоскелета, участников механизма трансляции и протеостаза, белков, участвующих в процессинге и связывании РНК, белков яйца и белков хроматина, а также одного ионного канала (таблица 7).

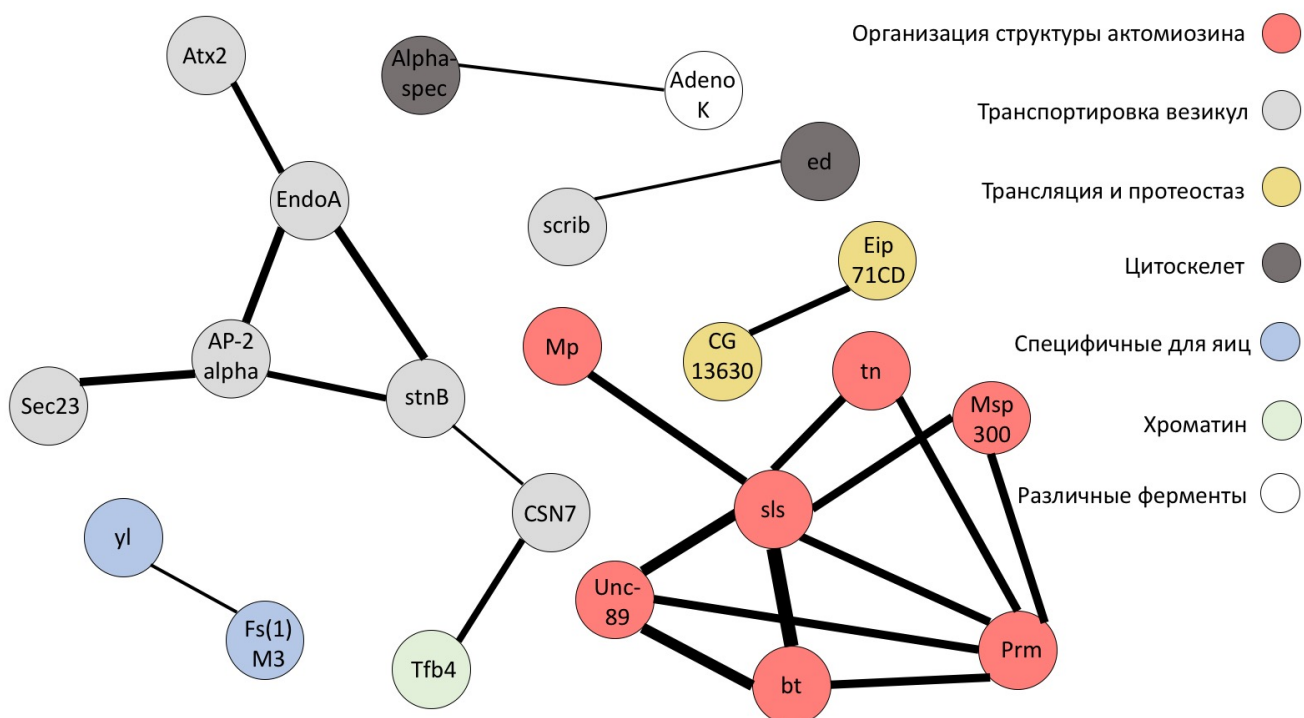


Рисунок 5. Белок-белковые взаимодействия перекодированных вследствие редактирования РНК продуктов, полученные из базы знаний STRING [152]. Толщина линий отображает количество методов, подтверждающих каждое взаимодействие согласно STRING. Цвета указывают на группы белков, указанных в таблице 1.

Таблица 7. Перекодированные редактированием мРНК сайты, идентифицированные в протеомах дрозофилы [26,182]

Белок	Перекодированный пептид	Пептид дикого типа	Идентифика- тор Uniprot	Аминокис- лотная замена	Ген	Набор данных		
						Полный жизненный цикл (PXD005691)	Эмбриогенез (PXD005713)	Наши данные
Организация структуры актомиозина								
Проджектин	VNA A ICAA NQ INCIN L IEGR	VNATICAANQINCIN L IEGR ^{a)}	O76281	T5323A	<i>bt</i>	✓		
Субъединица связывания миозина (F118194p1)	ELGTNSSET ASD VILR	ELGTNSSET AND VILR ^{b)}	Q8IQN4	N390S	<i>Mbs</i>	✓		
Мультиплексин (GH14382p)	GGFAS N LDG H K	GSFAS N LDG H K	Q8MT89	S721G	<i>Mp</i>	✓		
Мышечно-специфический белок 300 кДа	L N EL V A R	L N EL V A K	M9PC84	K6705R	<i>Msp300</i>	✓	✓	
Парамиозин, длинная форма	SLLESS LHR , r VEIELDSVR	SLLESS LHQ VEIELDSVR ^{a), b)} , r VEIELDSV Q VEIELDSVR	P35415-1	Q264R	<i>Prm</i>	✓		✓
Титин (D-титин) (Кеттин)	VTILDATD VPR	VTILDATD VPK ^{a)}	Q9I7U4	K14863R	<i>sls</i>	✓	✓	
Тин (GH06739p)	DTHAGSSTSS AA ASS SR	DTHAGSSTSS AA ASS SH AGYGS GT S GSSGR ^{a)}	Q7JUV6	H994R	<i>tn</i>	✓		
Обскурин (сборочный белок М-линии мышц Unc-89)	HDGG V IQVDD R	HDGE V IQVDD R ^{a)}	A8DYP0	E1351G	<i>Unc-89</i>	✓		✓
Транспортировка везикул								
Субъединица комплекса AP- 2 альфа (альфа-адаптин)	IIHLLNDQHMGVV TAA AS L IDAL V K	IIHLLNDQHMGVV TAA TS L IDAL V K ^{a), b)}	P91926	T207A	<i>AP- 2alpha</i>	✓		✓
Гомолог атаксина-2 (Dtx2)	GVGPAPSANASADSS SR	GVGPAPSANASADSS S K ^{a), b)}	Q8SWR8	K398R	<i>Atx2</i>	✓		✓
Комплексин	LKNQ M ETQV N EL K	LKNQ I ETQV N EL K ^{a)}	Q8IPM8-1	I124M	<i>cpx</i>	✓		✓

Субъединица 7 сигнального комплекса COP9	LQHLLIVSLAIR*, KLQHLLIVSLAIR	LQHLLIVSLAIK* ^{a), b)} , KLQHLLIVSLAIK ^{a), b)}	Q9V4S8	K115R	CSN7	✓	✓
Эндофиллин А	YSLDDNIEQNFLEPLHHMQTK	YSLDDNIK ^{a), b)} , QNFLEPLHHMQTK	B5RIU6	K137E	EndoA	✓	✓
Белок Iap4	LTIQHDPLPPGFQEVLLSR	LTIQHDPLPPGFQEVLLSK ^{a), b)}	Q7KRY7	K1340R	scrib	✓	
Белковый транспортный белок SEC23	rGPGQVVDELKHPHPIR	IMTFVGGPCSQGPGQVVDELKHPHPIR ^{a), b)}	Q9VNF8	Q304R	Sec23	✓	
Белок stoned-B	MALTSYDQIPSELAPYAFVEFTMPATQ VSHATVR	MALTSYDQIPSELAPYAFVEFTMPATQ VSHTTVR ^{a)}	Q24212	T1186A	stnB	✓	✓
Цитоскелет							
Альфа-цепь спектрина	rADSLVESGQFDTAGIQEK	DMNNQADSLVESGQFDTAGIQEK	P13395	Q1668R	alpha-Spec	✓	✓
Ехиноид	DQPTSATLSIIPR	DQPTNATLSIIPR	Q9VQW7	N210S	ed	✓	
Гельзолин	FYTGDSFIVLNTVENK	FYTGDSFIVLNTIENK ^{a)}	Q07171	I105V	Gel	✓	
Трансляция и протеолиз							
Метионин аминопептидаза	AIHALAAGASSSAEQDGAYNPWPHFR	AIHALAAGASNSAEQDGAYNPWPHFR ^{a), b)}	Q9VC48	N56S	CG1363 0 (MAP1)	✓	✓
Пептидметионинсульфокси дредуктаза	QYCYHVER	QYCYHVEQNEGQGLYC	P08761	Q238R	Eip71CD	✓	
Сплайсинг, связывание РНК							
АТФ-зависимая РНК- хеликаза p62	NTCVFGGAPR	NTCVFGGAPK ^{a), b)}	P19109	K394R	Rm62	✓	✓
Специфичные для яиц							

Предполагаемый рецептор вителлогенина (белок yolkless)	VLCSPSQFACHSGEQCVDKER*, VLCSPSQFACHSGEQCVDK	VLCSPNQFACHSGEQCVDKER*, VLCSPNQFACHSGEQCVDK	P98163	N1077S	yl	✓	✓
Белок Female sterile (1) M3	HERPLFVLR	HEKPLFVLR	Q9W451	K72R	fs(1)M3		✓
Хроматин							
Е3 убиквитин- протеинлигаза Bre1	LLLDVYK	LLLDMYK ^{a), b)}	Q9VRP9	M694V	Bre1	✓	
Белок Without children	APLTLDSDLPPGAGSSK	APLTLDSDLPPSAGSSK ^{b)}	Q9VB55	S949G	Woc		✓
FI01003p	QQLGSILMSAPR	QQLGSILMNAPR ^{b)}	Q9VPX4	N111S	Tfb4		✓ ✓
Ионные каналы							
Транзиторийный рецептор потенциал-зависимого катионного канала trpm, ортолог подсемейства M	rTPEAAQAGQAK	PSQTAAEAVPQTPEAAQAGQAK	A8DYE2	Q1933R	Trpm	✓	
Различные ферменты							
Аденозинкиназа (LP07155p)	LAVHEIVDANGAGDAFVGGFLSQFVQ GK	LAVHEIVDTNGAGDAFVGGFLSQFVQ GK ^{a), b)}	Q86NZ5	T297A	AdenoK	✓	✓
Кальпаин-Б	GALSSAGYHLNNR	GALNSAGYHLNNR ^{a), b)}	Q9VT65	N856S	CalpB	✓	✓
RH40737p (RH63541p)	IMQAGTK	IMQADTK	Q9VZJ8	D546G	CG1159 4	✓	
RE40534p	AIAGVLTPLNPLWALR	AIAR, VLTPLNPLWALR ^{a), b)}	Q5U191	R173G	CG1882 (pumI)	✓	✓
Гомолог белка 5, содержащего тиоредоксиновый домен	GYPTLLWIEDGR	GYPTLLWIEDGK ^{a), b)}	M9PGX2	K249R	prtp		✓

Без аннотации						
GH12158p	ILDVLFATGR	ILDVLFATG Q MK	Q9VFT7	Q723R	CG1084 1	✓
Неохарактеризованный белок (IP11775p)	TYNEQALV M LGR	TYNEQALV M LER	Q1EC94	E131G	CG1414 2	✓
Белок UPF0585	LVEMPSNNK	LV K , MPSNNK	Q9VLF6	K209E	CG1866 1	✓
GEO11026p1 (IP16896p, IP17096p)	CAQDWQPLCMESDLR	CTQDWQPLCMESDLR	A1A6Q4	T95A	CG3424 4	✓
GH16729p	ETIVTHDDGAPAVGMGGK	ETIVTHDDGAP T VGMGGK	Q9VMU1	T285A	CG4230	✓ ✓
Неохарактеризованный белок, изоформа C, изоформа E	rGLITDLV A LK	IAEQAYASPLGSSNNVLLMIMYNIYISG HGLITDLV A LK	B7YZJ4	H132R	CG4975 (Atxn10)	✓
GH01257p	TLE L IAK	TL K , LIAK	Q9VBQ4	K175E	CG5107	✓

Сайты были обнаружены в наборах данных плодовой мушки, представляющих протеомы в разные моменты времени в течение жизненного цикла и эмбрионального развития с номерами доступа PXD005691 и PXD005713 соответственно [182]. Список был отфильтрован, как описано в разделе 3.11. Обнаружили 14 общих сайтов перекодирования с данными из первой части работы (таблица 4). Для пептидов, обнаруженных в расщепленных и недорасщепленных формах, показаны оба варианта, звездочками обозначены пептиды с более высокой интенсивностью, выбранные для дальнейшего анализа. Строчные буквы, выделенные жирным курсивом, указывают на сайт редактирования, расположенный справа от пептида и не представленный в самом пептиде. ^{a)} Перекодированный пептид принадлежит набору данных PXD005691. ^{b)} Перекодированный пептид принадлежит набору данных PXD005713.

4.1.7. Распределение перекодированных белков по стадиям жизненного цикла

Используя протеомные данные, мы определили интенсивности перекодированных пептидов и уровни экспрессии соответствующих белков, чтобы сопоставить эти значения между стадиями развития. Все перекодированные сайты кластеризовали с использованием интенсивности иона-предшественника (MS1) как меры концентрации триптического пептида. Примечательно, что кластеризация «без учителя» распределила перекодированные белки в точном соответствии с фазами жизненного цикла (рис. 6). Можно предположить, что это распределение отражает уровни экспрессии соответствующих белков, поскольку фазовоспецифические белки распознавались раньше [29]. Однако те же самые белки, сгруппированные с использованием уровня их экспрессии, оцененного с использованием индекса нормированных спектральных коэффициентов обилия (normalized spectral abundance factors, NSAF) [151], не показали такого хорошего соответствия с фазами жизненного цикла (рис. П5 в Приложении). Кроме того, эти кластеры NSAF имели другой состав по сравнению с кластерами MS1. В большинстве случаев коэффициент корреляции Пирсона между интенсивностями MS1 триптических пептидов, содержащих перекодированные сайты, и значениями NSAF соответствующих белков значимо не отличался от нуля (уровень значимости $< 0,05$; серые столбцы справа, рисунок 6). Все эти данные ясно показывают, что несинонимичное редактирование РНК изменяет последовательности белков и, предположительно, их функции в зависимости от фазы жизненного цикла.

Мы можем только предполагать, как достигается это фазозависимое редактирование, принимая во внимание, что все изоформы ADAR у насекомого образует единственный ген. Более тщательного исследования в контексте перекодирования белков требуют такие факторы, как модуляция его ферментативной активности с помощью онтогенетических белковых регуляторов и переключение сплайс-изоформ [72].

Кластер белков, перекодированных преимущественно на эмбриональной стадии, включает эмбрионально-специфичный рецептор вителлогена, кодируемый геном *yl*. Субъединица сигнального комплекса COP9 (*CSN7*), перекодированная на этой стадии, как показано, участвует как в оогенезе, так и в эмбриогенезе [184]. Связанные с цитоскелетом белки гельзолин (*Gel*) и эхиноид (*ed*), участвующие в морфогенезе, также представляют интерес в этой группе. Эхиноидный белок и неохарактеризованный продукт гена *CG4975* перекодируются не только в фазе яйца, но и у личинок. Некоторые продукты, перекодированные в эмбриональной фазе, также имеют небольшой пик у взрослых особей (рис. 6). Для этого анализа мы изучили имаго обоих полов вместе. Когда поиск был проведен отдельно для самцов и самок, стало ясно,

что специфические для яиц идентификации были найдены у взрослых самок с эмбрионами, но не у самцов.

Личиночным стадиям, по-видимому, присуще перекодирование белков мышечного аппарата. Они включают титин (*sls*), мультиплексин (*Mp*) и мышечно-специфический белок (*Msp300*). Из трех белков кластера куколки продукт гена *scrib* был примечателен своим участием в морфогенезе крыла [185].

Самый большой кластер белков, кодируемый у поздних куколок и имаго насекомых, соответствовал ранее описанному всплеску редактирования ADAR, связанного с нейронными функциями у взрослых животных [72]. В кластере видны хорошо известные мишени ADAR, такие как белки, связанные с функцией синаптических пузырьков и их обновлением [26]. Это компоненты комплекса SNARE - комплексин (*cpx*), stoned protein B (*stnB*), а также эндофилин А (*EndoA*), регулятор кривизны мембран. Более того, РНК-связывающий продукт гена *Atx2* может представлять собой связь между редактированием РНК и регуляцией циркадных ритмов [186]. Помимо нейрональных компонентов, перекодированные мышечные белки, такие как *thin* (*tn*), парамиозин (*Prm*) и прожектин (*bt*), могут указывать на участие редактирования РНК в формировании мускулатуры взрослого насекомого.

Другой набор данных из работы [182], в котором подробно описывается эмбриогенез плодовой мушки, дал гораздо меньше перекодированных участков, чем основной набор данных жизненного цикла (таблица 7). Кластеризация двенадцати участков таким же образом, как на рисунке 6, ожидаемо не смогла установить каких-либо закономерностей (см. рис. П6 и П7 в Приложении) из-за малого количества компонентов для анализа. В этих данных уровни трех перекодированных редактированием сайтов коррелировали с содержанием целого белка.

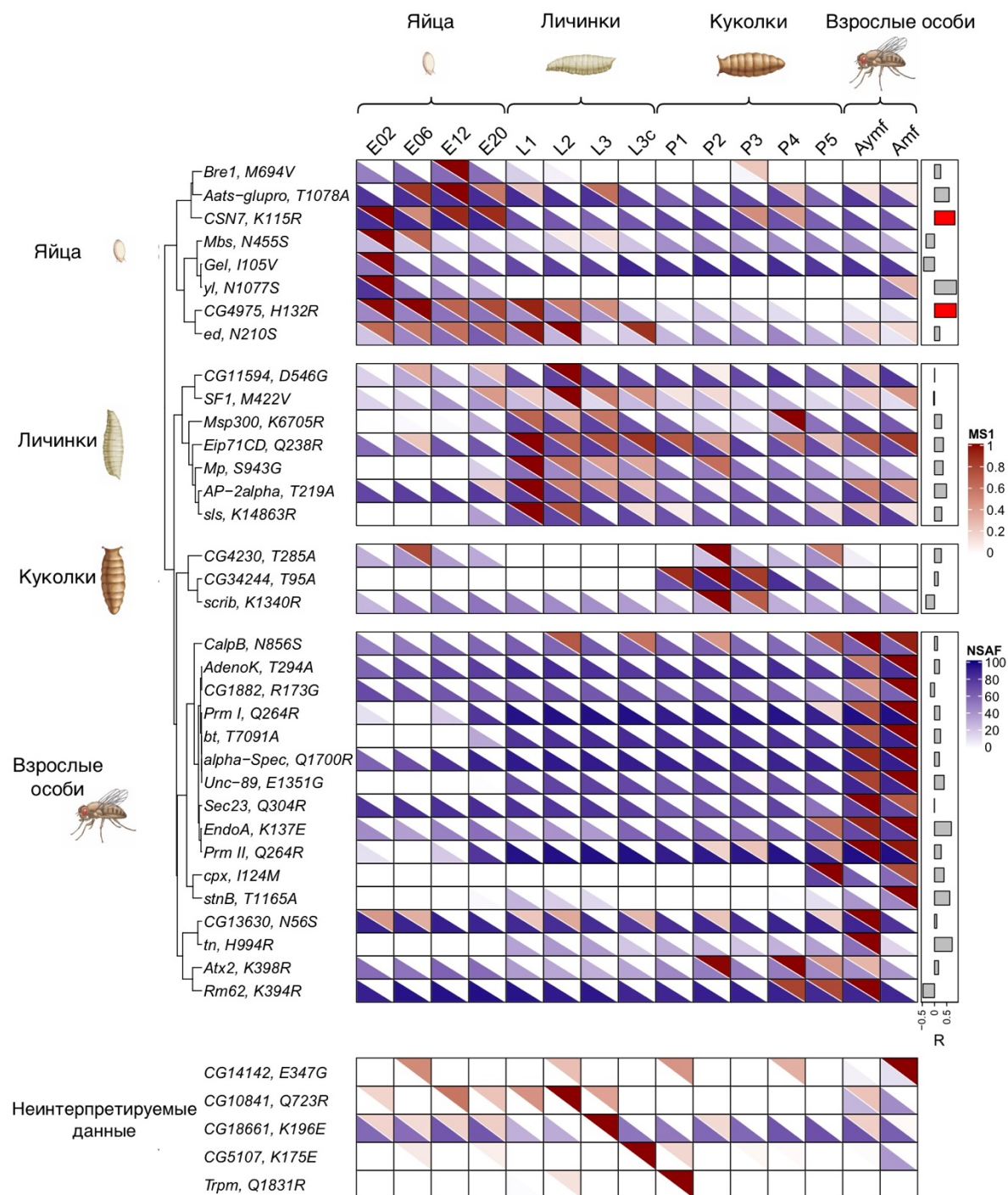


Рисунок 6. Тепловая карта относительной представленности перекодированных участков в виде триптических пептидов и соответствующих белков на различных стадиях жизненного цикла плодовой мушки. Уровень перекодирования (красная половина каждой ячейки диаграммы) показана в значениях интенсивности MS1 для соответствующих вариантных пептидов, нормированных индивидуально для каждой строки. Количество белка (фиолетовая половина каждой клетки) показана как процентиль NSAF [151]. Коэффициенты корреляции Пирсона между значениями MS1 и NSAF показаны справа. Для каждого коэффициента вычислен уровень значимости его отличия от нуля. Значения, выделенные красным, представляют собой коэффициенты Пирсона с поправкой на множественное сравнение ($FDR < 0,05$). Кластеризация проводилась с использованием интенсивностей MS1. Если перекодированный редактированием участок был идентифицирован, а содержание белка невозможно было оценить, данные считались неинтерпретируемыми. Протеомные данные по стадиям жизненного цикла взяты из работы Casas-Vila и соавт. [182]. Пептиды SLLESSLHR и VEIELDSVR, принадлежащие парамиозину, показаны как Prm_l (left) Prm_r (right) соответственно.

4.1.8. Перекодированные с помощью ADAR участки в транскриптомных и протеомных данных жизненного цикла плодовой мушки

На следующем этапе мы сравнили протеомные последствия редактирования РНК с заменой аденозина на инозин с транскриптомными данными для жизненного цикла *D. melanogaster*, находящимися в свободном доступе. В одном из наиболее полных исследований, включающих 30 временных точек жизни насекомых, было зарегистрировано 972 события редактирования ADAR в масштабе всего транскриптома [35]. Эти сайты редактирования большей частью входили в базу данных, используемую для протеомного поиска [110,134]. В предыдущих исследованиях транскриптомов продемонстрировано, что редактирование в большинстве сайтов начинается на стадии поздней куколки, несмотря на то, что некоторые транскрипты редактируются и в ранней эмбриональной стадии [35]. Мы отобрали 624 несинонимичных события редактирования для сравнения с 40 перекодированными участками, полученными выше из повторно проанализированных протеомных данных. Только 12 сайтов были общими для двух наборов данных (рис. 7). Низкий уровень совпадения зарегистрированных сайтов может объясняться как биологическими, так и техническими факторами. В самом деле, ранее было показано, что корреляция между количеством транскриптов мРНК и соответствующих белковых продуктов может быть низкой из-за разной продолжительности жизни этих молекул в клетках и жидкостях организма [172]. Методы транскриптомики и протеомики, которые различаются по чувствительности, помимо других аналитических характеристик, также вызывают несоответствие результатов. Ожидается, что мультиомиксный подход, при котором одни и те же образцы используются для выделения РНК и белков, улучшит воспроизводимость, хотя в исследованиях редактирования РНК он до сих пор не использовался.

Тепловая карта сравнения уровней редактирования транскриптов с относительным уровнем представленности соответствующих участков белка показана на рисунке 7. Можно отметить, что только сайты, перекодированные на стадиях поздней куколки и у взрослых особей, имеют транскриптомные аналоги. Редактирование транскриптов в большинстве случаев обнаруживалось на несколько этапов раньше, чем белковые изменения. Это можно объяснить ограниченной чувствительностью масс-спектрометрии, используемой для обнаружения представляющих интерес пептидов, по сравнению с методом амплификации нуклеиновых кислот. Постепенно, с увеличением количества протеоформ, кодируемых редактированием, их становится возможно обнаружить на уровне белка.

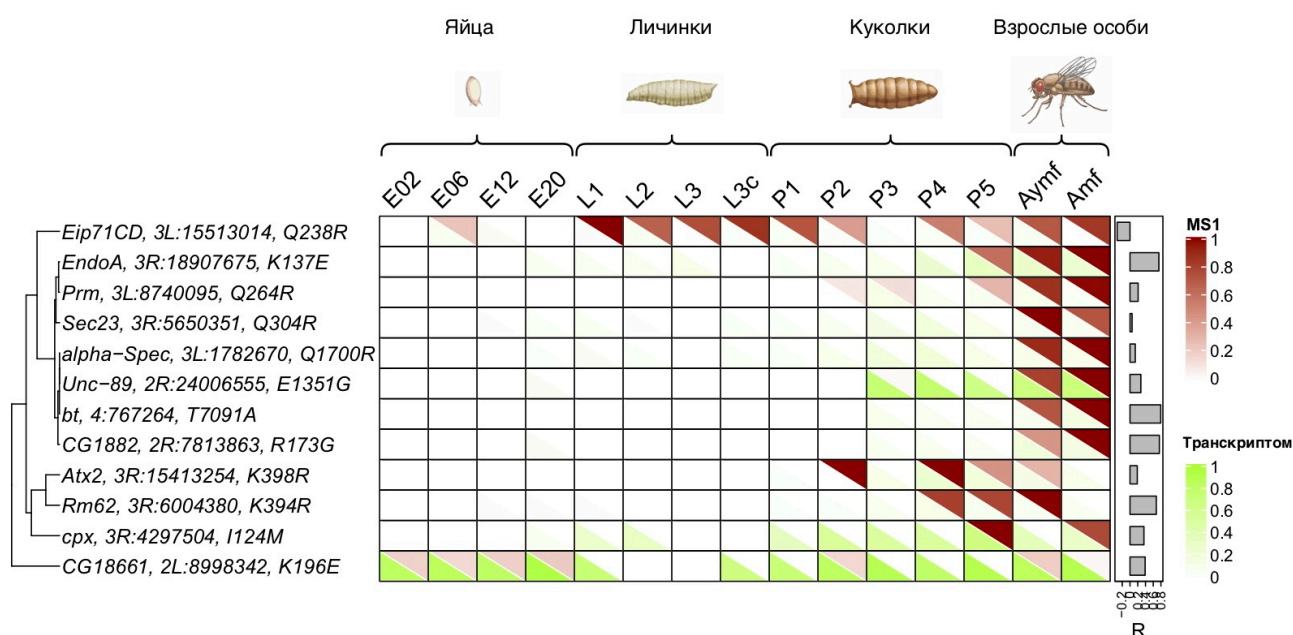


Рисунок 7. Представленность перекодированных белковых участков и соответствующих им отредактированных транскриптов на различных стадиях жизненного цикла плодовой мушки. Степень перекодирования белка (красная верхняя половина каждого прямоугольника) показывает интенсивности MS1 соответствующих пептидов, нормализованные для каждой строки отдельно. Представленность транскриптов (зеленая нижняя половина каждого прямоугольника) показывает процент редактирования, взятый из более ранней работы [35]. Данные, сгруппированные на основе коэффициентов корреляции Пирсона между значения MS1 и процентом перекодирования, показаны справа. Для каждого коэффициента вычислен уровень значимости его отличия от нуля; он был выше 0,05 во всех случаях (нет значимой корреляции). Кластеризация проводилась с использованием интенсивностей MS1. Единственное выпадающее значение MS1 было исключено из данных для гена *CG18661*. Стадии жизненного цикла были взяты из публикаций Casas-Vila и соавт.[182]

4.1.9. Количественное определение специфичных для мозга белков, перекодированных редактированием РНК, на разных стадиях жизненного цикла плодовой мушки

После анализа перекодированных участков белков целых тел насекомых, мы выполнили исследование перекодирования в мозге плодовой мушки. Таргетный протеомный анализ для измерения перекодированных редактированием и кодируемых в геноме сайтов в выбранных белках были предложен на основе их повторной идентификации в протеомах мозга и головы, а также их связи с известными функциями [26]. Мы количественно оценили уровни перекодирования этих белковых сайтов в мозге, выделенном из *D. melanogaster*, на разных фазах жизненного цикла, включающих личинку, куколку и имаго.

Белки, перекодированные редактированием РНК, в основном были связаны с синаптической передачей сигнала и, в частности, с путями, которые обеспечивают слияние синаптических пузырьков с пресинаптической мембраной для высвобождения первых [165]. Были исследованы синтаксин 1A (*Syx1A*), комплексин (*cpx*), эндофилин А (*EndoA*), продукты

генов *Cadps* и *stol*, а также гомолог атаксина-2 (*Atx2*), чьи функции подробно описаны в разделе 4.1.5.

Сравнительные уровни перекодированных и кодируемых в геноме сайтов шести выбранных белков были количественно определены в мозге *D. melanogaster* на трех этапах жизненного цикла с использованием таргетной масс-спектрометрии на основе мониторинга множественных реакций (MRM) с использованием меченых стабильными изотопами стандартов пептидов. Стандарты были разработаны и ранее протестированы для измерения редактирования белка в мозге взрослых насекомых в первой части работы (раздел 4.1.5). Все биологические эксперименты были выполнены на двух независимых культурах дрозофил с использованием двух разных хромато-масс-спектрометров с детектором типа тройного квадруполь.

Таргетная масс-спектрометрия предназначена для абсолютного количественного определения представляющих интерес белков. Однако этот метод измеряет концентрацию протеолитического пептида, и во время выделения белка из образца и расщепления его протеазой в анализ могут быть внесены технические погрешности. Фактическое абсолютное количественное определение методом достигается, когда все этапы эксперимента строго стандартизированы, включая роботизацию и решение методологических проблем [187]. В настоящей работе наблюдалась разница между абсолютными количественными данными, нормализованными по общему содержанию белка, полученными для разных биологических повторов (рис. П8 Приложения). В то же время результаты были согласованными в каждой повторности, а относительные уровни измеренных продуктов были достаточно сходными. Уровень перекодирования каждого белкового сайта выражали в процентах перекодированного продукта (рис. 8) в соответствии с формой представления, установленной для результатов секвенирования РНК, чтобы оценить то же самое на уровне белка [188]. Уровень перекодирования рассчитывался как концентрация варианта, перекодированного редактированием РНК, деленная на сумму концентраций перекодированных редактированием и закодированных в геноме форм.

Из шести рассмотренных в данном исследовании сайтов перекодированных белков, пять дали вполне интерпретируемые результаты (рис. 8). К сожалению, как перекодированные, так и кодируемые в геноме формы *Atx2* были обнаружены только в одном биологическом повторе со значениями, близкими к пределу обнаружения. Таким образом, результаты по данному участку перекодирования было трудно охарактеризовать. Остальные сайты продемонстрировали тенденцию к увеличению уровня редактирования в мозге при переходе от стадии личинки к взрослой особи (рис. 8). Сайт эндофилина А с заменой лизина на глутаминовую кислоту в 137 положении перекодировался в наибольшей степени, достигая 75-80% в мозге взрослой особи. В

то же время уровень перекодирования этого сайта также был относительно высоким даже у личинок.

Из двух компонентов комплекса SNARE [166] перекодирование участка комплексина (*срх*) значительно увеличивалось от стадии личинки до куколки, а затем стабилизировалось или даже немного снижалось. Что касается синтаксина 1A (*Syx1A*), то уровень его перекодирования был низким на всех этапах с тенденцией к увеличению в одном из биологических повторов.

Кальций-зависимый активатор секреции CADPS представляет собой белок, который функционально близок к комплексу SNARE, так как он также способствует высвобождению кальций-зависимых везикул [163]. Его умеренно перекодированный сайт также характеризуется увеличением процента перекодирования в течение жизненного цикла. Примечательно, что этот белок является одним из немногих, транскрипты которого также редактируются ADAR у млекопитающих. Однако его перекодированные сайты, обнаруженные в протеомах человека и мыши, неконсервативны и отличаются от замены метионина на валин в 1234 положении у плодовой мушки [26]. Другой связанный с нейротрансмиссией белок — это потенциал-зависимый кальциевый канал *stol* (*CG4587*). Несинонимичное редактирование РНК его ортологов, которые модулируют приток кальция через каналы, широко распространено у грызунов [189]. При этом описанная здесь замена аргинина на глицин в 489 положении для плодовой мушки не сохраняется в качестве сайта, перекодированного редактированием, у млекопитающих. Перекодирование этого сайта значительно увеличивается на каждой фазе жизненного цикла, что, вероятно, указывает на его функциональную важность. Этот тип замены в результате редактирования РНК изменяет размер и электростатический заряд аминокислотного остатка и, как было показано, модулирует электрофизиологические свойства субъединиц глутаматных каналов крыс [190].

В соответствии с транскриптомными данными, таргетный анализ белков мозга, перекодированных редактированием, показал, что метаморфоз насекомых сопровождается увеличением уровня перекодирования, повторяющим всплеск редактирования РНК, описанный ранее [72] и наблюдаемый во всех приведенных выше данных. Примечательно, что по крайней мере пять из шести белков были перекодированы в личиночной фазе, что дополнительно подтверждает предположение о существовании базального уровня активности ADAR в раннем

развитии.

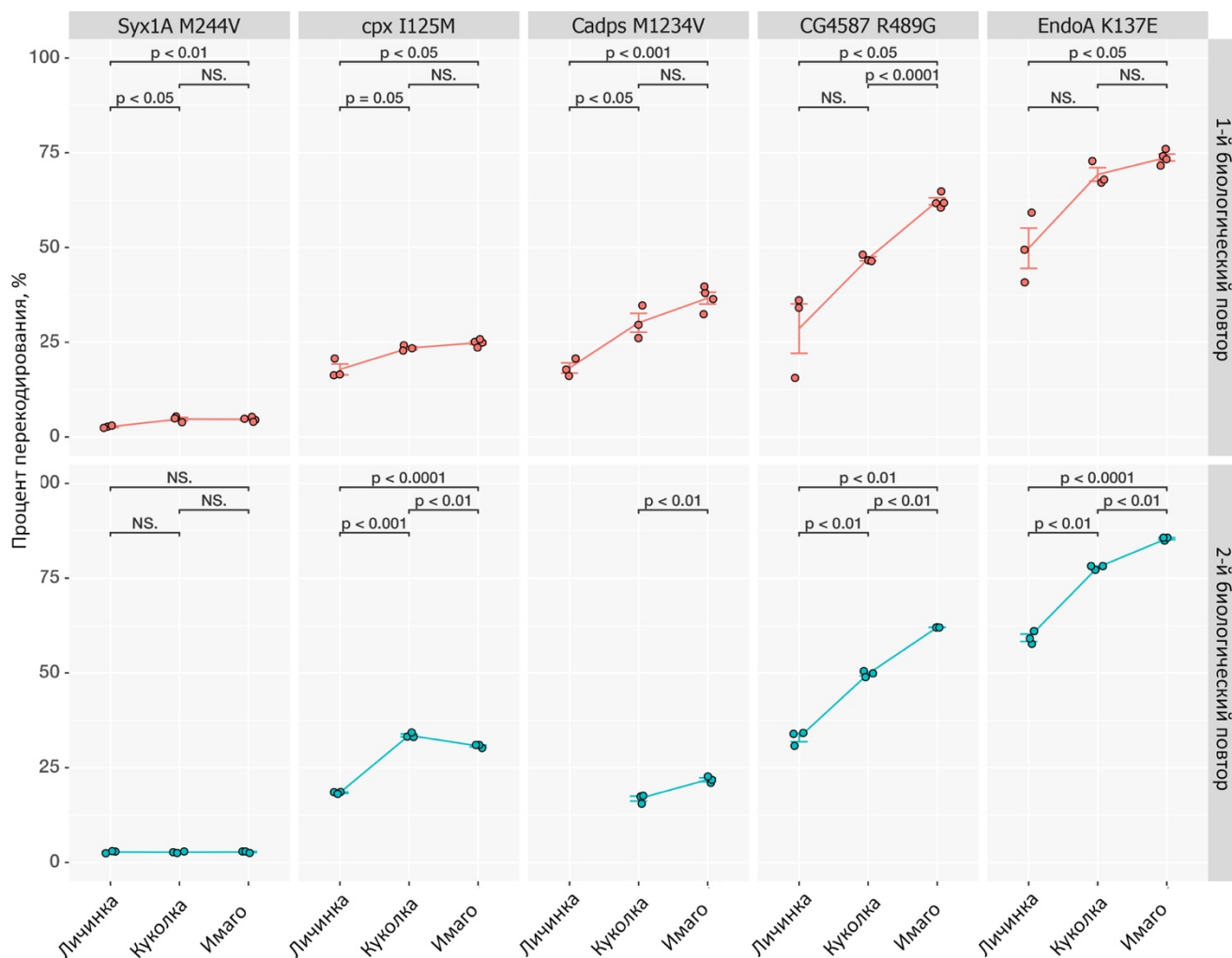


Рисунок 8. Динамика уровня перекодирования выбранных участков белков в мозге на разных фазах жизненного цикла *D. melanogaster*. Данные показаны независимо для двух биологических повторов.

Процент перекодирования рассчитывался как концентрация варианта, перекодированного редактированием РНК, деленная на сумму концентраций перекодированных редактированием и кодируемых в геноме форм. Уровень значимости Т-теста представлен для оценки разницы в уровне перекодирования между фазами жизненного цикла. Планки погрешностей представляют собой стандартное отклонение для трех технических повторов. Перекодированный редактированием сайт *Cadps* во втором биологическом повторе не был обнаружен на личиночной стадии.

Вскоре после открытия [191] ADAR-опосредованного редактирования РНК в модельных организмах, оно было изучено с использованием ингибирования или модификации активности ADAR с помощью генетических мутаций и наблюдения за эффектами на различных уровнях организации животного. Недавно были обнаружены некоторые механизмы, приводящие к нейродегенерации в отсутствие активности ADAR с использованием штамма *D. melanogaster* с инактивированным ферментом [71]. Кроме того, была описана точечная мутация, индуцированная в гене ADAR плодовой мушки, которая сохраняла уровень экспрессии полученного белка, но подавляла его ферментативную активность [23]. Используя этих

модельных животных, была показана уникальная двойная функция фермента ADAR дрозофилы, который разделяет черты изоформы ADAR1 млекопитающих, участвующей в подавлении врожденного иммунитета и в то же время редактирующей мРНК и, следовательно, белковые последовательности изоформы ADAR2 [23].

В дополнение к исследованиям функциональности с использованием мутантных моделей, молекулярные эффекты активности ADAR у плодовой мушки были изучены на уровне транскриптома [134,192] и - в этой работе - на уровне протеома. Показано, что фатальный дефицит ADAR2 у мышей может быть устранен путем геномного редактирования одного сайта в субъединице [32] рецептора глутамата GRIA2. Однако эти результаты не могут быть напрямую применены к вопросам, касающимся здоровья человека, из-за другой продолжительности жизни и физиологии грызунов. Недавнее всестороннее исследование мутантов плодовых мушек показало, что локомоторная недостаточность, связанная с нарушением ферментативной функции ADAR, опосредуется перекодированием белков [23]. В то же время остается неизвестным, как именно перекодированные белки модулируют клеточные и физиологические функции.

Идентификация протеоформ, возникающих при редактировании РНК, предоставляет инструмент для выяснения роли редактируемых сайтов в штаммах *D. melanogaster* дикого типа и мутантных. Здесь предоставлены такие данные для фаз жизненного цикла плодовой мушки. Эти результаты продемонстрировали, что фазовоспецифичное редактирование белков может происходить независимо от уровней их экспрессии. Однако остается неясным, какие факторы регулируют это перекодирование в течение жизненного цикла. В соответствии с ранее полученными в других работах транскриптомными данными, мы также наблюдали всплеск несинонимичного перекодирования, вызванного редактированием РНК, на уровне протеома во время метаморфоза насекомых. Положительная динамика перекодирования белков была впервые подтверждена с помощью таргетной масс-спектрометрии белков мозга плодовой мушки.

4.2. Поиск сайтов редактирования РНК в протеоме мыши и человека

4.2.1. Анализ находящихся в открытом доступе результатов

транскриптомных и протеомных данных для биоинформатического поиска редактирования РНК

Протеомы мышинного и человеческого мозга, полученные на приборах высокого разрешения, по три для каждого вида, были отобраны из репозитория открытых данных Proteomexchange [124]. Выбрано два наиболее объемных набора данных, доступных для мозга мыши и человека: PXD001250 [125] и PXD005445 [126], соответственно. Эти наборы

использовались в качестве базовых эталонных данных, другие два протеома для каждого организма были добавлены к ним для проверки результатов. Дополнительные данные мышинных протеомов взяты для моделей болезни Гентингтона [127] и нейропатической боли [128], представленных под номерами доступа PXD003442 и PXD004087, соответственно. Дополнительные данные протеомов человека были взяты из исследования спинномозговой жидкости (CSF) [129] под номером доступа PXD009646 и из протеоеномного исследования дорсолатеральной префронтальной коры мозга [130] под номером доступа PXD004143. Исследование спинномозговой жидкости было выбрано для проверки, могут ли отредактированные белки быть идентифицированы в образце, который обычно лучше доступен в клинических условиях, чем ткани мозга. Характеристики выбранных для проекта протеомов приведены в таблице 8.

Таблица 8. Характеристика наборов данных, используемых для идентификации событий редактирования

ProteomeXchange ID	Организм	Биоматериал	Суммарное число PSM	Суммарное число PSM при 1% FDR
PXD001250, [125]	мышь	Различные ткани мозга и первичные клеточные культуры здоровых мышей линии C57BL/6	15 949 908	8 633 512
PXD003442, [127]		Полосатые тела нокаутированных мышей с увеличенным количеством повторов кодона CAG (модель болезни Гентингтона)	9 625 663	4 641 359
PXD004087, [128]		Префронтальная кора стареющих мышей с дефицитом програнулина или избыточной экспрессией програнулина в нейронах	1 099 253	701 336
PXD005445, [126]	человек	Различные ткани мозга 16 человек без явной патологии	6 623 202	3 969 816
PXD009646, [129]		Образец спинномозговой жидкости из пула 96 здоровых доноров	558 627	60 764
PXD004143, [130]		Дорсолатеральная префронтальная кора двух индивидуумов	64 876	27 521

4.2.2. ADAR-опосредованное редактирование мРНК в протеоме мышинного мозга

В соответствии со схемой, представленной на рисунке 9 по базам данных RADAR³ и DARNED⁴, а также и из дополнительных данных недавней работы [21], содержащих геномные координаты сайтов редактирования мРНК, составлена база для поиска отредактированных белков. Поиск пептидов был произведен по хромато-масс-спектрометрическим данным для отделов мозга и первичных культур клеток мозга мышей линии C57BL/6 в программе X!Tandem с применением группоспецифичного FDR.



Рисунок 9. Схема интеграции данных секвенирования нуклеиновых кислот и протеомной хромато-масс-спектрометрии. Алгоритм действий включает в себя формирование базы данных для протеомного поиска отредактированных участков, основываясь на методах высокопроизводительного секвенирования. Следующий этап – получение данных панорамного протеомного анализа и поиск отредактированных пептидов с использованием программы X!Tandem. Для валидации полученных результатов может использоваться таргетный масс-спектрометрический анализ.

³ Коллекция строго аннотированных событий редактирования аденозина на инозин - a collection of rigorously annotated A-to-I events

⁴ DAtabase of RNa EEditing

В протеомах мышиногo мозга идентифицировали 20 отредактированных сайтов, принадлежащих продуктам 14 генов (таблица 9). 15 из этих сайтов характеризовались более чем двумя спектральными идентификациями ($PSM > 2$). Среди идентифицированных измененных пептидов можно найти примеры редактирования, хорошо известные в данной области по результатам ортогональных молекулярных методов. В таблице 9 показано, что профили идентифицированных отредактированных пептидов из меньших наборов данных [127,128], как правило, представляют собой подмножество самого большого набора данных [125].

Субъединицы глутаматного рецептора GRIA2 (GluA2) и GRIA3 (GluA3), идентифицированные в этом исследовании, являются классическими примерами событий перекодирования вследствие редактирования [86]. Для них были идентифицированы события перекодирования как во флип, так и во флоп слайс-изоформах с заменой аргинина на глицин (R/G) (таблица 9). Недавно было показано, что, по крайней мере, во флоп-изоформе редактирование ускоряет открытие и десенсibilизацию образованных GRIA2 ионных каналов [190]. Более узнаваемый сайт редактирования с заменой глутамина на аргинин (Q/R) в GRIA2, который, как было показано, полностью отредактирован в мозге взрослых млекопитающих [20], также был идентифицирован, но с меньшими количеством спектральных identификаций, чем для сайта R/G. Скорее всего, это обусловлено сложностью детекции данного пептида масс-спектрометрией. Последний содержит остаток глутамина на N-конце, что часто приводит к непредсказуемым химическим изменениям при фрагментации [193]. Редактирование РНК в субъединицах глутаматных рецепторов было тщательно изучено на разных молекулярных уровнях [78]. Для белков основным инструментом было создание рекомбинантных, принудительно редактируемых рецепторов в моделях клеток и организмов [190]. С помощью таргетной протеомики, основанной на наших результатах, можно отслеживать динамику отредактированной изоформы в интактных тканях, включая клинические образцы.

Цитоплазматический белок 2, взаимодействующий с FMR1 (The cytoplasmic fragile X mental retardation 1 interacting protein 2, CYFIP2), является белком в мышинном протеоме с основным перекодированным сайтом замены лизина на глутаминовую кислоту (K/E) в 320-м положении его последовательности. Данный сайт перекодирования выявлен с наибольшим числом спектральных identификаций. CYFIP2 является важным компонентом актин-регулирующего комплекса WAVE [194], который реорганизует актиновые филаменты в ответ на внешние раздражители с образованием ламеллоподий. Функционально показано, что этот комплекс участвует в морфогенезе позвоночника и пресинаптической модуляции [195]. Мутации зародышевой линии CYFIP2 могут приводить к эпилептической энцефалопатии с дебютом в раннем возрасте [196], а также иметь связь с переизданием [197]. Несмотря на накопление знаний

о роли этого белка в различных путях, точная функция его перекодирования с заменой К/Е остается неясной.

Альфа-филамин (филамин-А, FLNA) представляет собой крупный белок с множественными иммуноглобулино-подобными повторами, который обеспечивает пространственную организацию актиновых филаментов, а также связывает многие лиганды, включая цитоплазматические домены трансмембранных рецепторов [198]. Его перекодированный сайт с заменой глутамина на аргинин (Q/R), расположенный в 22-м филаминовом повторе и идентифицированный в этой работе, был описан ранее [199]. Функциональное значение этой замены остается неясным. Было высказано предположение, что редактирование может модифицировать связывание FLNA с его белком-партнером, бета-интегрином [199].

Альфа-субъединица коатомера (COPA) является компонентом специфического покрытия везикул (COP1) и обеспечивает ретроградный транспорт белков из комплекса Гольджи в эндоплазматический ретикулум [200]. Транскрипты этого белка являются хорошо известными мишенями для ADAR. Интересно, что поиск в больших данных ракового протеома обнаружил ту же замену изолейцина на валин [34]. Несмотря на консервативный характер замены, было показано, что, по крайней мере, в клеточной линии рака молочной железы это редактирование способствовало росту рака по еще неизвестному механизму [34].

Еще один идентифицированный белок - кальций-зависимый активатор секреции 1 (CADPS) - участвует в кальций-зависимом высвобождении синаптических везикул [201]. Это единственный белок из нашего списка, чей ортолог был также найден перекодированным в протеоме плодовой мушки. Тем не менее, в мозге мышей и насекомых CADPS был перекодирован в разных негомологичных местах. Набор перекодированных белков у исследованных млекопитающих и насекомого существенно отличается, причем компоненты пресинаптических механизмов высвобождения в основном редактируются у насекомых.

Из ранее описанных мишеней редактирования ADAR следует упомянуть инсулиноподобный белок 7, связывающий фактор роста (IGFBP7) с его сайтом замены аргинина на глицин, идентифицированный в наборе данных протеома полосатых тел [199]. Значение перекодированных остатков в этом белке также остается неясным.

Таблица 9. ADAR-опосредованные аминокислотные замены, идентифицированные в протеомах мышиного мозга [125,127,128]. Участки, идентифицированные по двум и более спектральным идентификациям (PSM) перечислены в верхней части таблицы, а сайты с одним PSM - в нижней части. Белковые сайты, которые также перекодируются в человеческом мозге, затенены.

Название белка	Последовательность пептида	Идентификатор Uniprot	А/к замена	Ген	Суммарное количество PSM		
					Шарман и соавт. [125]	Лангфельд и соавт. [127]	Альтман и соавт. [128]
Митохондриальная ацил-коэнзим А тиоэстераза	VHSEVFSLSGREHMTT NVFHFTFMSEK	Q32MW3	D381G	ACOT10	14	4	
Кальций-зависимый активатор секреции 1 (Кальций-зависимый активатор белка для секреции 1) (CAPS-1)	VNGEMYIER, <i>dk</i> VNGEMYER*	Q80TJ1	E1252G	CADPS	17	2	7
Кальций-зависимый активатор секреции 1 (Кальций-зависимый активатор белка для секреции 1) (CAPS-1)	YVDVPEPGMDVADAY VTFVR	Q80TJ1	K1226E	CADPS	8		
Альфа субъединица коаптомера	VWDVSGLR	F8WHL2	I164V	COPA	27		
Цитоплазматический белок 2, взаимодействующий с FMR1	YIETSAHYEENK, YIETSAHYEENK <i>sk</i>	Q5SQX6	K320E	CYFIP2	88	72	35
Цитоплазматический белок 2, взаимодействующий с FMR1	IQFPIEMGMPWILTDHIL ETK <i>epsmmeyvlypldlyndsay</i> <i>yaltk</i> , <i>r</i> IQFPIEMGMPWILTDHI LETK	Q5SQX6	S337G	CYFIP2	2		
Филамин-А	LTVSSLR, <i>r</i> LTVSSLR	B7FAU9	Q2333R	FLNA	15	1	
Глутаматный рецептор 2	QGCDISPR	E9QKC0	Q607R	GRIA2	13	10	
Глутаматный рецептор 3, флип изоформа	GSALGTPVNLAVLK	Q9Z2W9	R769G	GRIA3	26	39	20
Глутаматный рецептор 3, флоп изоформа	GSALGNAVNLAFLK	F2Z488	R769G	GRIA3	9	22	10
Глутаматный рецептор 2; Глутаматный рецептор 4, флип изоформа	GSSLGTPVNLAVLK	E9QKC0; Q9Z2W8	R765G, R764G	GRIA2; GRIA4	41	22	18
Инсулиноподобный белок, связывающий фактор роста 7	GEGEPCGGGAAGGGHC APGMECVK	E9Q5D9	R78G	IGFBP7		3	
Метаботропный глутаматный рецептор 4 (mGluR4)	SLTFVQALIEK	G3XA00	Q118R	GRM4	8		
Митохондриальный НАДФ-зависимый оксиянтарный фермент	LSNHVVFVQGAGEAAM GVAHLLVMALEK	Q8BMF3	I75V	ME3	7	2	
Потенциал-зависимая альфа-1G субъединица кальциевого канала Т-типа	VSRHIR**	Q5SUF9	K2078R	CACNA1 G	3		
Кальций-активируемая субъединица альфа-1 калиевого канала	EEVVAAEVGWMTGVK	J3QMT8	S36G	KCNMA 1	1		

Цитоплазматический белок 2, взаимодействующий с FMR1	VMGFGLYLMDGNVSN YR	Q5SQX6	K283R	CYFIP2	1
Белок Son	AAELSVVSASVISEQSE QPMPGMLEPSMTK	H9KV15	T216A	SON	1
SH3 домен-связывающий белок 2	HPCGYAGPR	Q06649	Y597C	SH3BP2	1
Потенциал-зависимая альфа-1G субъединица кальциевого канала Т-типа	DPQGTR	Q5SUF9	E2099G	CACNA1 G	1

* В недорасщепленных пептидах добавления к перекодированным пептидам показаны строчными буквами.

** Показан недорасщепленный пептид, поскольку его короткие аналоги не могут быть видны из-за условий поиска.

Ряд событий перекодирования вследствие редактирования, идентифицированных в этой работе, никогда не обсуждался в литературе, и в то же время они идентифицированы по меньшему количеству спектральных идентификаций, чем было обнаружено в известных редактируемых сайтах. Следует упомянуть два ионных канала: метаботропный глутаматный рецептор 4 (GRM4) и потенциал-зависимую альфа-1G субъединицу кальциевого канала Т-типа (CACNA1G). Последняя функционально связана с белком stol плодовой мушки, который, как было показано в данной работе, тоже перекодировался. Неожиданно среди перекодированных белков мышинного мозга были обнаружены два фермента с митохондриальной локализацией, ACOT10 и ME3.

Из полученных данных протеоеномного исследования без подтверждения другими ортогональными методами можно только утверждать, что события перекодирования могут быть задетектированы. Высокая чувствительность молекулярных методов может обеспечить идентификацию тех сайтов, для которых не показана функциональная значимость и поможет понять, являются ли эти сайты результатом побочных эффектов ферментативных реакций с помощью ADAR. Леванон и соавт. [199] также предположили, что события редактирования, приводящие к аминокислотным заменам, могут быть побочным эффектом альтернативной регуляции сплайсинга с помощью ADAR, например, в филамине альфа, как было показано ранее для глутаматного рецептора [202].

4.2.2.1. Распределение аминокислотных замен вследствие редактирования мРНК в тканях мозга

Используя хромато-масс-спектрометрические данные высокого разрешения микроглии мозга молодых и взрослых мышей, гранулярных мозжечковых и кортикальных нейронов (все первичные культуры) и фракционированного целого мозга [125], перекодированные пептиды были оценены по количеству идентифицированных спектров без использования изотопной метки. Для анализа были отобраны те образцы, в которых количество спектральных идентификаций (peptide-spectrum match, PSM) для отредактированных пептидов каждого представляющего интерес белка было не менее 5. Для клеточных культур были найдены сайты перекодирования, характеризующиеся наибольшим количеством спектральных идентификаций. В образцах тканей, содержащих много стромальных внеклеточных компонентов, перекодированные сайты были идентифицированы с низким количеством PSM. Из всех исследованных тканей мозга в работе Шарма и соавторов [125], для протеома целого мозга, проанализированного фракциями, показано наибольшее количество спектральных идентификаций для перекодированных продуктов.

Используя пять наборов данных с большим количеством перекодированных сайтов, было исследовано, соответствует ли степень перекодирования уровню каждого белка. Другими словами, если имеется высокое значение PSM для перекодированных пептидов глутаматного рецептора в кортикальных нейронах и низкое значение в микроглии, означает ли это, что рецептор высоко экспрессируется в нейронах? Чтобы ответить на этот вопрос, были пересчитаны параметры LFQ (intensity-based label-free quantification, количественный анализ без использования изотопной метки) для белков в пяти представляющих интерес наборах данных и получены процентные доли индекса нормированных спектральных коэффициентов обилия (normalized spectral abundance factors, NSAF), [151]. Далее эти индексы были нанесены на график для тех белков, которые достигли пяти или более PSM перекодированных пептидов, по крайней мере в одном из интересующих наборов данных (рис. 10).

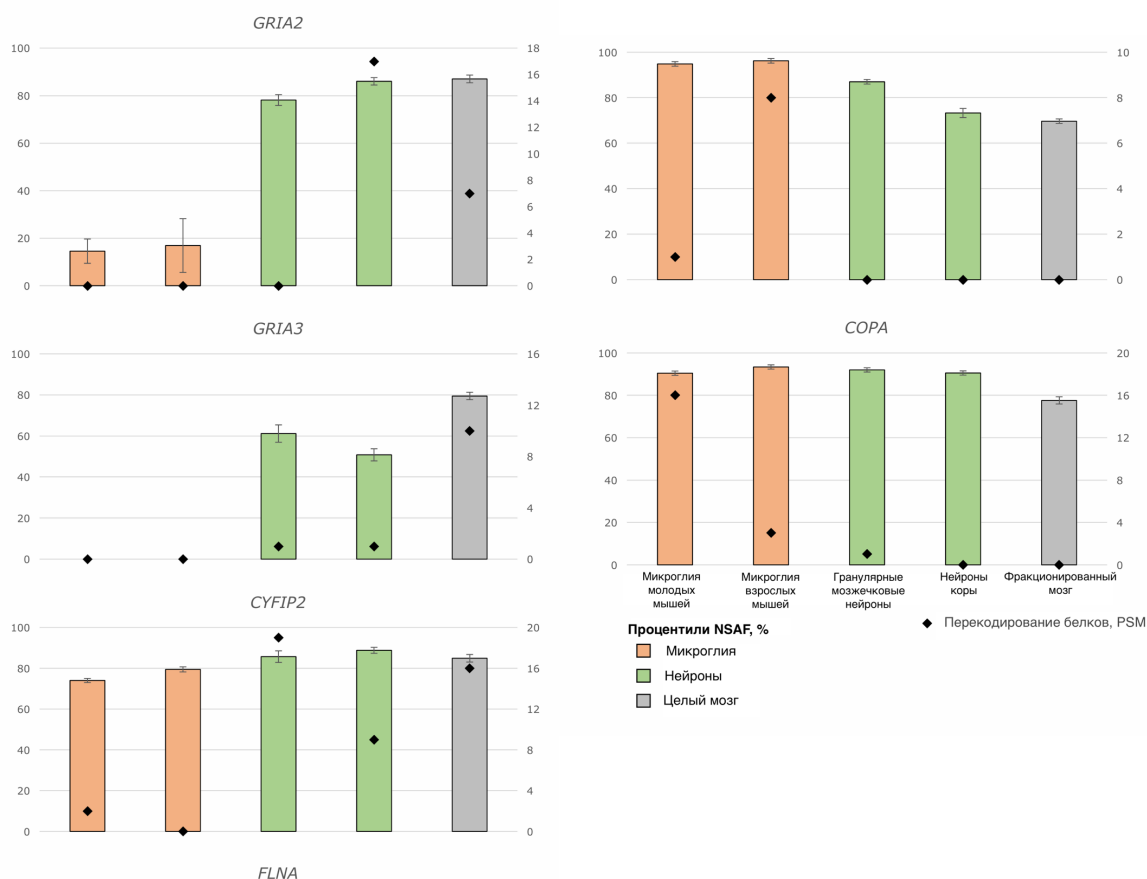


Рисунок 10. Уровни перекодирования и экспрессии выбранных белков в субпротеомах микроглии мозга молодых и взрослых мышей, гранулярных мозжечковых, кортикальных нейронах (все первичные культуры) и фракционированного целого мозга. Степень перекодирования выражается в PSM перекодированных пептидов, просуммированных для каждого белка, если у него было более одного перекодированного сайта. Экспрессия белка выражается в процентилях NSAF [151] с планками погрешностей, выражающих стандартное отклонение для трех биологических повторов. Субпротеомы из работы Шарма и соавт. [125] были отобраны, если хотя бы один из идентифицированных перекодированных сайтов имел более пяти PSM. Субъединица глутаматного рецептора GRIA2 имеет общий перекодированный сайт с GRIA4, однако перекодирование сообщается в первую очередь для GRIA2.

В некоторых случаях простая логика «нет белка - нет перекодирования редактированием» действительно имела место. Таким образом, в микроглии практически отсутствовали рецепторы глутамата GRIA2 и GRIA3, и поэтому в этом типе клеток не было выявлено перекодированных сайтов. Для других белков наблюдался разный уровень перекодирования при сопоставимых уровнях экспрессии. Например, перекодирование продукта гена *GRIA2* было специфичным для нейронов коры головного мозга без перекодирования в нейронах гранулярных слоя мозжечка (рис. 10). Продукт гена *GRIA3* имел низкие уровни перекодирования в обоих типах нейронов, несмотря на высокие уровни для целого мозга, что может означать его перекодирование в других типах нейронов, не охваченных конкретным набором данных [125].

Из полученных данных следует, что продукт гена *CYFIP2* дифференцированно перекодируется в нейрональных клетках с низким уровнем перекодирования в микроглии, несмотря на относительно высокую представленность белка во всех типах клеток. Напротив, филамин-А (*FLNA*) и альфа-субъединица коатомера (*COPA*) перекодированы в основном в микроглии. Примечательно, что редактирование продукта гена *COPA*, характерное для микроглии новорожденных мышей, в то же время, как было показано, увеличивает пролиферацию, миграцию и инвазию раковых клеток *in vitro* на примере клеточных линий рака молочной железы и почки [34]. Данный перекодированный белок отвечает за ретроградный транспорт из аппарата Гольджи в эндоплазматическую сеть. Вероятно, его специфическое перекодирование с заменой изолейцина на валин способствует изменениям в обороте транспортных белков, характерным для растущих клеток.

Наши результаты убедительно свидетельствуют о том, что мишени ADAR в разных типах клеток мозга по-разному перекодируются. Скорее всего, эта дифференциальная схема обеспечивается ADAR2, который характеризуется таргетной специфичностью, еще не идентифицированной [203]. Отличные друг от друга паттерны перекодирования белков требуют дальнейшего изучения, например, таргетной протеомикой, чтобы выяснить их возможную роль в заболевании. Например, микроглия, которая, очевидно, является объектом такого специфического редактирования РНК, может вызывать гибель нейронов при боковом амиотрофическом склерозе.

В то время как в этой работе идентифицированы несколько классических примеров перекодированных белков, таких как глутаматные рецепторы, некоторые другие найти не удалось. Одним из таких белков является серотониновый рецептор, точнее, 5-гидрокситриптаминный рецептор 2C (HTR2C) [204]. К сожалению, один из основных перекодируемых сайтов этого рецептора с заменой изолейцина на метионин в 156 положении (I156V) находится в коротком триптическом пептиде, выходящем за рамки параметров поиска. В то же время сам белок HTR2C не был идентифицирован ни в наборах данных для клеточных линий, обсуждавшихся выше, ни в образцах мозга. Вероятнее всего, панорамный протеомный анализ исследуемых данных не обеспечил достаточную чувствительность для идентификации этого рецептора, специфичного для выбранных типов нейронов и областей мозга, например, миндалины [205].

4.2.3. ADAR-опосредованное редактирование мРНК в протеоме мозга человека

Для идентификации событий перекодирования в протеоме человеческого мозга, использовался тот же дизайн исследования, что и для мышинных данных. Более объемный набор данных ожидаемо обеспечил большинство identifications перекодированных пептидов [126]. В статье Карлайла и соавт. утверждалось, что эта работа по мозгу человека была аналогом исследования мышинного мозга, проведенного Шармой и соавт. [125]. Однако данные относились к большему количеству субъектов и в то же время не имели того количества identifications, которое было достигнуто при исследовании мозга лабораторных животных. Более того, не было проанализировано клеточных культур, которые, в случае мышинного протеома мозга, давали лучший охват перекодированных продуктов. Анализируя данные человека, также необходимо было рассмотреть полиморфизм генома, который не может наблюдаться у инбредных животных и, в свою очередь, иногда может перекрывать события перекодирования. База данных для человека содержала более 1800 событий перекодирования, что в три раза больше мышинной, вероятно, в связи с лучшей изученностью событий редактирования РНК человека.

В общей сложности, в протеомах человеческого мозга идентифицировали 38 перекодированных сайтов, принадлежавших продуктам 33 генов, причем только 19 сайтов характеризовались двумя или более спектральными identifications ($PSM \geq 2$). Совокупность результатов, общих для мыши и человека, включала 8 участков: основной участок редактирования, кодируемый геном *CYFIP2*, три участка для глутаматных рецепторов *GRIA2* и *GRIA3*, участки *COPA*, *FLNA* и *CADPS*, описанные выше (таблица 9 и 10). В отличие от мышинных данных для субъединицы рецептора *GRIA2* также была найдена отредактированная флор-изоформа с заменой аргинина на глицин.

В этой части исследования мы намеренно включили поиск по протеому спинномозговой жидкости [129]. Ликвор является наиболее доступным образцом, связанным с человеческим мозгом, который может быть извлечен без сложного оперативного вмешательства. Как и следовало ожидать, классические сайты редактирования РНК, такие как в глутаматных рецепторах, в этой жидкости идентифицировать не удалось. Вместо этого был обнаружен сайт замены глутамина на аргинин простагландин-Н2D-изомеразы (PTGDS). PTGDS является многофункциональным ферментом и нейромодулятором, также известным как биомаркер ликвора, применяемый для диагностики ликвореи [206]. Кроме того, в ликворе также идентифицировали сайт редактирования белка-участника каскада инсулиноподобного фактора роста IGFBP7, до этого обнаруженного в протеоме полосатых тел мозга мыши [127].

Еще один набор данных [130], который относился к протеоеномному исследованию префронтальной коры, имел недостаточное для анализа количество идентификаций и показал единственный отредактированный сайт (PSM=1) белка 3, взаимодействующего с полимеразой дельта. Даже в человеческом протеоме с наибольшим количеством идентификаций, полученным Карлайлом и соавт. [126], не удалось найти того количества перекодированных пептидов, как в мышинном наборе данных Шармы и соавт. [125]. Таким образом, для данных человеческого мозга мы не смогли поддержать или опровергнуть результаты дифференциального перекодирования белков для разных клеток, показанные для мышинного мозга в разделе 4.2.2.1.

Таблица 10. Перекодированные вследствие редактирования РНК участки, идентифицированные в протеомах мозга человека [126,129,130]. Участки, найденные по двум и более спектральным идентификациям (PSM), представлены в верхней части таблицы, а сайты с PSM=1 - в нижней части. Значения в правом столбце представляют собой суммарные значения PSM пептидов для каждого перекодированного сайта во всех проанализированных образцах, если не указано иное. Сайты белка, которые также перекодируются в мышинном мозге, затенены.

Белок	Последовательность пептида	Идентификатор Uniprot	А/к замена	Ген	Количество PSM
Анкириновый повтор и белок 1A, содержащий домен смерти	SQLTAASASR	Q495B1-5	T507A	ANKDD1A	5
белок 2B, содержащий домен АТФазного семейства AAA	GLSVTSEQINPR	ATAD2B	H1092R	ATAD2B	2
Кальций-зависимый активатор секреции 1	AGGGRPSGSPSPVSEK	Q9ULU8	S91G	CADPS	5
Кальций-зависимый активатор секреции 1	VNGEMYIER, dkVNGEMYIER*	Q9ULU8	E1241G	CADPS	3
Каспаза 1	DFIAFCSSAPDNVSWR	P29466	T313A	CASP1	2
Альфа субъединица коаомера	VWDVSGLR	P53621	I164V	COPA	3
Цитоплазматический белок 2, взаимодействующий с FMR1	YIETSAHYEENK	Q96F07	K124E	CYFIP2	73
Филамин-А	LTVSSLR, rLTVSSLR	P21333	Q2341R	FLNA	10
Митохондриальная глутаматдегидрогеназа 2	GFIGPGVDVPAPDMSTGER	P49448	N227S	GLUD2	2
Глутаматный рецептор 3 (GluR-3), флорп изоформа	GSALGNAVNLAVLK	P42263	R775G	GRIA3	26
Глутаматный рецептор 3 (GluR-3), флип изоформа	GSALGTPVNLAVLK	P42263	R775G	GRIA3	10
Глутаматный рецептор 2 (GluR2); Глутаматный рецептор 4 (GluR4), флорп изоформа	GSSLGNAVNLAVLK	P48058; P42262	R765G (Gria4); R764G (Gria2)	GRIA4; GRIA2	13
Глутаматный рецептор 2 (GluR2); Глутаматный рецептор 4 (GluR4), флип изоформа	GSSLGTPVNLAVLK	P48058; P42262	R765G (Gria4); R764G (Gria2)	GRIA4; GRIA2	43
Белок теплового шока 1L, 70 кДа	QTQIFTTYSDSQPGVLIQVY EGER	P34931	N436S	HSPA1L	4
Инсулиноподобный белок, связывающий фактор роста 7 (IGFBP-7)**	GEGEPCGGGGAGGGYCAP GMECVK	Q16270	R78G	IGFBP7	4

Несприн-1	DIQQTEQTIEQR	Q8NF91	K4121R	SYNE1	4
Простогландин-H2 D-изомераза***	AAPEAQVSVQPNFQQDK	P41222	Q21R	PTGDS	3
Субъединица 1 комплекса WASH	ATSQGGDLMSDLFNK	A8K0Z3	H407D	WASHC1	7
Цинковый палец 587B	TSLGNMVK	E7ETH6	I396M	ZNF587B	4
40S рибосомальный белок S18	VITIMQSPR	P62269	N73S	RPS18	1
Белок 1, подобный АДФ-рибозилирующему фактору	IGISESELVAMLEEEELR	P40616	K104E	ARL1	1
Анкириновый повтор и белок 1A, содержащий домен смерти	SQAATSASR	Q495B1-5	T505A	ANKDD1A	1
Белок 40, содержащий спираль-спиральный домен	LAGILNR	Q4G0X9	S336G	CCDC40	1
Циклин-I	QLLHCVACNQLLQFR	Q14094	M189V	CCNI	1
Dapper, гомолог 3	QRPPDASPPGGARPAR	Q96B18	S298G	DACT3	1
Филамин-B	VNQPAFAIR	O75369	S2283G	FLNB	1
Глутаматный рецептор 3 (GluR-3), флип изоформа	LSEQGILDKLENK	P42263	K795E	GRIA3	1
Гистоновая деметилаза UTY	SPVTVIR	O14607	N957S	UTY	1
Полиаденилат-связывающий белок 3	APPSGYFMTAVPQTQNR	Q9H361	H413R	PABPC3	1
Белок 3, взаимодействующий с полимеразой дельта****	ALANMSR	Q9BY77	T268A	POLDIP3	1
Вероятный глобальный активатор транскрипции SNF2L1	EEMLVIR	P28370	M656V	SMARCA1	1
Альфа-цепь митохондриальной пропионил-КоА-карбоксилазы,	LLIEEFIDNPR	P05165	K236E	PCCA	1
Альфа-цепь митохондриальной пропионил-КоА-карбоксилазы,	LSSQEASGFGDDR	P05165	S252G	PCCA	1
Гомолог белка numb (h-Numb)	VTAATEQAER	P49757	T199A	NUMB	1
Белок 1, содержащий сорбин и SH3 домены	LECSGTVIAHCSLR	Q9BX66	K457R	SORBS1	1
Сортирующий нексин-14	IPVFCIDVGR	Q9Y5W7	E545G	SNX14	1
Белок цинковых пальцев 432	TLESNLSLVNQNK	O94892	K138E	ZNF432	1
Белок цинковых пальцев 714	SGVQDRPGQHGK	Q96N38	Q530R	ZNF714	1

*В недорасщепленных пептидах добавленная часть показана строчным буквами курсивом.

** У человека этот сайт был обнаружен только в протеоме спинномозговой жидкости [129]. В то же время, он также был обнаружен в наборе данных протеомов полосатых тел мыши [127].

*** Тот же сайт был обнаружен в протеоме спинномозговой жидкости Макрона и соавт. [129], (PSM = 2).

**** Сайт был обнаружен только в наборе данных Винго и соавт. [130].

4.2.4. Анализ редактирования РНК на уровне протеома головного мозга трансгенных мышей - моделей заболеваний ЦНС

При сравнении идентификаций перекодированных пептидов мыши и человека выяснилось, что восемь из них, наблюдаемых с самой высокой интенсивностью, полностью совпадают. Основные участки редактирования между видами оказались консервативными. Поэтому был получен набор для таргетного анализа этих участков для мыши и человека. Набор был испытан на мозжечке и коре больших полушарий трех линий трансгенных мышей и контрольных линий.

Количественный анализ перекодированных белков был проведен на трех линиях трансгенных мышей - моделях заболеваний, описанных ниже. Чтобы понять количественные изменения уровня редактирования для исследуемых белков, были взяты 3 группы трансгенных мышей и соответствующие им контрольные группы. Образцы мозжечка и коры были проанализированы в трех биологических повторах для каждого образца.

Первая группа – самки мышей линии CD1 с мутацией в гене FUS (fused in sarcoma), кодирующем ДНК/РНК-связывающий белок [207]. Возраст каждой особи группы составлял 108 дней. В геном мышей введен конструктор, кодирующий патогенную форму белка FUS под контролем нейроспецифического промотора Thy1. Данные мыши получались при скрещивании особей дикого типа с гомозиготами FUS. В качестве контроля использовались самки мышей той же линии, полученные при скрещивании гемизиготных мышей с особями дикого типа. Возраст каждой особи контрольной группы составлял 96 дней. Аберрантная форма данного белка лишена области, ответственной за распознавание и связывание РНК, теряет ядерную локализацию, становясь преимущественно цитоплазматической. Такая форма склонна к агрегации и образованию неамилоидных включений в пораженных нервных тканях. У трансгенных мышей наблюдались повреждения моторных нейронов, их аксонов, а также нейровоспалительная реакция. Данные патологические изменения вызвали внезапное развитие тяжелого моторного фенотипа в возрасте 2,5-4,5 месяцев и смерть животных в течение нескольких дней после начала. Паттерн патологии таких мышей повторял несколько ключевых особенностей бокового амиотрофического склероза человека с динамикой прогрессирования заболевания [208].

Вторая группа – трансгенные самцы мышей с таупатией. Возраст каждой особи группы составлял 154 дня. В качестве контроля использовались самцы линии C57Bl/6J возрастом 146 дней. Белок тау экспрессируется преимущественно в нейронах и на более низких уровнях в астроцитах и олигодендроцитах. Играет жизненно важную роль в регуляции цитоскелета, связывая тубулин и контролируя таким образом основные функции микротрубочек [209]. Таупатия включает в себя

формирование филаментов, содержащих гиперфосфорилированную форму этого белка, что уменьшает связывание тау с микротрубочками и приводит к дегенерации нервных клеток. Такие трансгенные мыши повторяли ключевые признаки таких нейродегенеративных заболеваний, как болезнь Паркинсона, болезнь Пика или болезнь Альцгеймера, где тау участвует в образовании внутриклеточных агрегатов [210].

Третья группа мышей – трансгенные мыши обоих полов, повторяющие признаки шизофрении. Возраст особей этой группы составлял 200 дней. Данная линия трансгенных мышей получена в лаборатории генетического моделирования нейродегенеративных процессов ИФАВ РАН [211]. Особи этой группы несут в геноме конструкт, кодирующий короткую изоформу NEAT1_1 человека под контролем промотора *Thy1*, обеспечивающих экспрессию в нервной ткани. Встройка трансгена была подтверждена ПЦР-анализом. Трансгенные животные оказались жизнеспособными, фертильными и не имели внешних признаков патологии; при скрещивании с мышами дикого типа в потомстве наблюдалось расщепление 1:1, что соответствовало ожидаемому при менделевском типе наследования. Экспрессия трансгена в тканях головного и спинного мозга была подтверждена методом количественной ПЦР с обратной транскрипцией (кОТ-ПЦР), с использованием праймеров, специфически узнающих человеческий вариант короткой изоформы гена NEAT1. Данная линия переведена на генетический фон линии C57Bl/6J. NEAT1 (nuclear paraspeckle assembly transcript 1) - длинная некодирующая РНК (нкРНК), являющаяся структурным компонентом субъядерных тел в межхроматиновых областях. В недавнем исследовании было показано, что экспрессия длинной некодирующей РНК NEAT1 снижается при шизофрении [212]. Изучение поведения трансгенных мышей этой линии позволило выявить наличие определенного фенотипа, характеризующегося такими признаками, как стресс-индуцируемая гиперактивность, сниженная тревожность, изменение ритмов активности и нарушение социального поведения, при этом моторные функции животных не были нарушены. Данный фенотип сохранялся при старении. Так как длинная изоформа *Neat1_2* в норме экспрессируется только в определенных органах и тканях, например, в кишечнике и не обнаруживается в ЦНС [213], то фенотип нокаутных мышей может быть объяснен потерей экспрессии короткой изоформы *Neat1_1*. В качестве контроля были использованы мыши обоих полов линии C57Bl, возраст каждой особи составлял 200 дней.

Уровень перекодирования был оценен для аминокислотных замен, представленных в таблице 11.

Таблица 11. Пептиды дикого типа и соответствующие им отредактированные, проверенные с помощью мониторинга множественных реакций (MRM) с использованием меченых стабильными изотопами стандартов пептидов

№	Название белка	Идентификатор Uniprot	Пептид дикого типа	Отредактированный пептид	Аминокислотная замена
1	Кальций-зависимый активатор секреции 1, CADPS	Q80TJ1	AGGGRPSSPSPS VVSEK	AGGGRPSGPSP SVVSEK	Сер-89-Гли
			VNEEMYIER	VNGEMYIER	Глу-1252-Гли
2	Субъединица 3 глутаматного рецептора, GRIA3	F2Z488	NAVNLAVLK, GSALRNAVNLA VLK	GSALGNAVNL AVLK	Арг-769-Гли
3	Субъединицы 2 и 4 глутаматного рецептора, GRIA2 и GRIA4	E9QKC0 Q9Z2W8	TPVNLAVLK, GSSLRTPVNLA VLK	GSSLGTPVNLA VLK	Арг-764-Гли Арг-765-Гли
4	Цитоплазматический белок 2, взаимодействующий с FMR1, Cyfip2	Q5SQX6	TSAHYEENK	YIETSAHYEEN K	Лиз-320-Глу
5	Альфа-субъединица коамера, Сора	F8WHL2	VWDISGLR	VWDVSGLR	Иле-164-Вал

Исследовали ткань коры головного мозга и мозжечка, в которых, по имеющимся данным, максимально представлено редактирование РНК. В результате были получены новые данные об отношении отредактированных и кодируемых в геноме участков на белковом уровне. При анализе результатов мы опирались на долю отредактированных участков исследуемых белков. На рис. 11 представлены основные результаты. Доля перекодирования варьировала от 100% перекодированных пептидов, например, для продуктов генов GRIA3 (глутаматный рецептор) и CYFIP2, до менее 10% для COPA и CADPS в мозжечке. Как оказалось, профиль перекодирования был совершенно одинаковым во всех трех линиях трансгенных мышей по сравнению с контролем. Таким образом, экспрессия генов *FUS*, *tau* и *NEAT1* в головном мозге мышей никак не влияла на редактирование РНК, детектируемое на уровне протеома. Из рис. 11 также видно, что профили перекодирования разных анализируемых сайтов воспроизводимо отличаются между корой головного мозга и мозжечком всех животных. Значение этих различий также может быть поводом для дальнейших исследований.

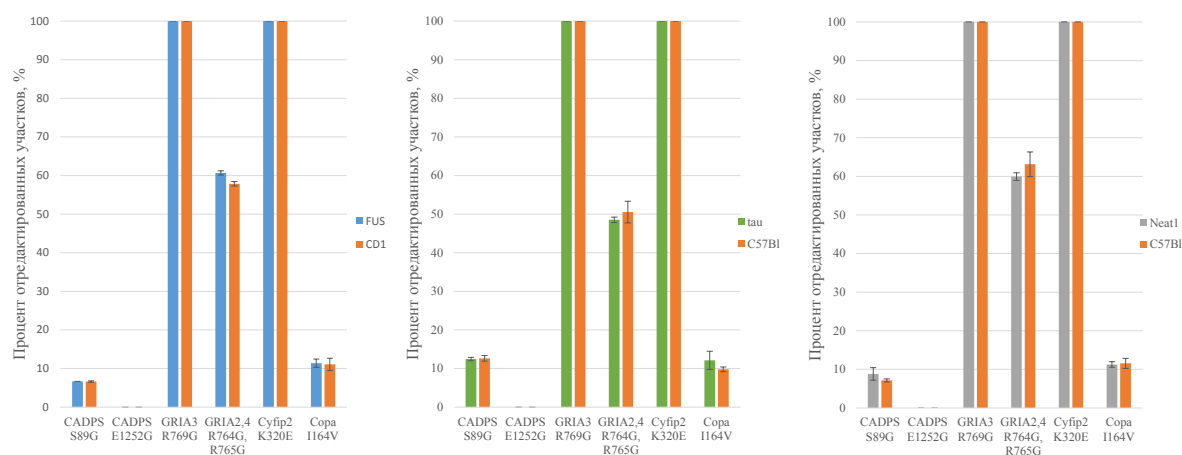


Рисунок 11. Процент редактирования исследуемых пептидов в мозжечках трансгенных мышей с различными нейродегенеративными заболеваниями, отклонениями в поведении по типу шизофрении и соответствующие им контрольные образцы. Проверенные участки редактирования белков включали пептиды, принадлежащие кальций-зависимому активатору секреции 1 (CADPS) с заменами Сер-89-Гли и Глу-1252-Гли; субъединице 3 глутаматного рецептора (GRIA3) с заменой Арг-769-Гли; субъединицам 2 и 4 глутаматного рецептора (GRIA2, GRIA4) с соответствующими им заменами Арг-764-Гли и Арг-765-Гли; цитоплазматическому белку 2, взаимодействующему с FMR1 (Cyfip2) с заменой Лиз-320-Глу и альфа-субъединице коатомера (Copa) с заменой Иле-164-Вал. Высота столбца указывает средний процент перекодирования пептида по трем биологическим повторам. Планки погрешностей показывают стандартное отклонение. Гистограммы представлены для трех патологических состояний. Группа FUS, обозначенная синими столбцами – модель бокового амиотрофического склероза. Группа tau, показанная салатовым цветом – таупатия с образованием внутриклеточных агрегатов по типу болезни Альцгеймера. Процент редактирования интересующих участков в группе Neat1, соответствующей описанной недавно модели шизофрении, изображен в виде серых столбцов. Оранжевыми столбцами обозначены мыши из контрольных групп.

4.2.5. Анализ редактирования РНК на уровне протеома цереброспинальной жидкости

Описанный выше набор для таргетного анализа участков перекодирования также был испытан на 25 образцах от пациентов с нейродегенеративными заболеваниями и нарушениями мозгового кровообращения. Как и ожидалось из анализа панорамных протеомов этой биологической жидкости на предыдущем этапе (раздел 4.2.1), результат был отрицательным. Ни один из анализируемых сайтов не был зарегистрирован в образцах ликвора. Единственный участок редактируемого белка, обнаруженный в ней на предыдущем этапе, не анализировали методом MRM из-за сложностей в синтезе пептидного стандарта.

5. ЗАКЛЮЧЕНИЕ

Масс-спектрометрия высокого разрешения добилась успеха в идентификации и количественном определении белковых продуктов генов в виде так называемых «мастерных» белков. На сегодняшний день анализ с использованием хромато-масс-спектрометрии высокого

разрешения позволяет получить протеом, охватывающий около 50% человеческих кодирующих белки генов, из одного образца [214,215]. Это соотношение еще выше у модельных организмов, которые могут характеризоваться открытыми геномами с более высокими числами экспрессируемых белков [216]. Соответственно, следующая цель протеомики состоит в том, чтобы каталогизировать протеоформы каждого белка, то есть подмножества белковых форм, происходящих из одного гена [217,218]. В данной работе были зарегистрированы протеоформы плодовой мушки, мыши и человека, которые произошли из-за редактирования мРНК посредством ферментов класса ADAR.

Как и ожидалось, не каждая аминокислотная замена, предсказанная по данным РНК, может быть обнаружена. Частично это связано с технологическим пределом обнаружения, который в протеомике намного выше, чем в исследованиях нуклеиновых кислоты на основе их амплификации. Более того, многие отредактированные РНК могут вообще не транслироваться. Предполагается, что фермент ADAR изначально был участником каскада распознавания двухцепочечных структур РНК в составе вирусных геномов. На сегодняшний день иммунорегуляторные функции обычно ассоциируются с ADAR1, по крайней мере, в клетках млекопитающих [9]. Таким образом, многие отредактированные РНК могут потерять свою стабильность и элиминироваться без трансляции. Точно так же случайно перекодированные белковые продукты могут оказаться нестабильными и уничтожаться системами протеостаза.

Редактирование РНК, опосредованное ферментами ADAR, является широко распространенной посттранскрипционной модификацией и критически значимо для многоклеточных организмов, включая человека. Дезаминазная активность ферментов семейства ADAR преобразует аденозиновые нуклеозиды РНК в инозиновые, что может приводить к изменению ее информационного значения и, как следствие, аминокислотным заменам в белке. С использованием протеогеномного подхода с дополненными базами данных для протеомного поиска в данных панорамной протеомики обнаружены 62 участка перекодирования в 54 белках плодовой мушки. Также перекодирование, опосредованное редактированием РНК, было изучено для фаз жизненного цикла плодовой мушки. Результаты продемонстрировали, что перекодирование белков на разных стадиях метаморфоза может не иметь прямой зависимости от уровней их экспрессии. Однако остается неясным, какие факторы регулируют это перекодирование в течение жизненного цикла. В соответствии с ранее полученными в других работах транскриптомными данными мы также наблюдали всплеск перекодирования, вызванного редактированием РНК, во время метаморфоза насекомых. Перекодирование выбранных участков было подтверждено с помощью таргетной масс-спектрометрии белков мозга плодовой мушки. Показано, что некоторые белки с известной нейрональной функцией,

например, эндофилин, присутствовали в мозге взрослого насекомого большей частью в перекодированном виде.

С использованием протеогеномного подхода были проанализированы хромато-масс-спектрометрические данные высокого разрешения мозга мыши [125,127,128] и человека [126,129,130], выбранные из общедоступных источников. Насколько нам известно, в этой работе протеогеномный подход впервые использовали для поиска ADAR-опосредованного редактирования в протеомах мозга. Результаты по числу идентифицированных сайтов перекодирования были более скромными, чем для дрозофилы – хромато-масс-спектрометрические данные с большим количеством идентификаций (более 9000 белков) были необходимы для нахождения более десяти отредактированных пептидов в мозге млекопитающих.

В центральной нервной системе мыши было определено, что одни и те же мишени ADAR в нейрональных и микроглиальных клетках редактируются по-разному. Скорее всего, это различие обеспечивается ADAR2, который характеризуется еще не идентифицированной таргетной специфичностью. В мышинном мозге идентифицированы 20 отредактированных белковых сайтов, принадлежащих 14 генам. Для человеческого мозга найдено 37 отредактированных участков, относящихся к 33 генам. При сравнении идентификаций перекодированных пептидов мыши и человека выяснилось, что восемь из них, наблюдаемых с самой высокой интенсивностью, полностью совпадают по последовательности. При этом найденные перекодированные белки у исследованных млекопитающих и насекомого существенно отличались. Было показано, что компоненты пресинаптических механизмов высвобождения нейромедиаторов в основном редактируются у насекомых. Единственным ортологичным геном, который перекодировался в мозге млекопитающих и насекомых, был CADPS. Однако, и этот белок был изменен в разных негомологичных местах.

Так как основные участки перекодирования вследствие редактирования у мыши и человека оказались консервативными, был получен набор для таргетного анализа этих участков. Его использование на мышинных моделях нескольких нейродегенеративных заболеваний показало, что введение трансгенов *FUS*, *tau* и *NEAT1* не оказывает воздействия на профиль перекодирования. При этом выявлены отличия в уровне перекодирования анализируемых сайтов в коре и мозжечке всех исследованных линий. Функциональное значение этих различий должно быть исследовано в дальнейших работах. Перспективными для дальнейшего исследования функции перекодирования белков могут оказаться и культуры клеток мозга различного происхождения - нейронов и микроглии.

6. ВЫВОДЫ

1. При исследовании протеома плодовой мушки впервые обнаружены 62 перекодированных в результате редактирования РНК участка, принадлежащих 54 белкам. Эти белки в основном являлись компонентами цитоскелета или были связаны с синаптической передачей сигнала.
2. Таргетный протеомный анализ белков мозга насекомых на разных стадиях развития позволил подтвердить и количественно измерить изменения в пяти белках, связанных с синаптической проводимостью, кодируемых генами *EndoA*, *stol*, *Cadps*, *cpx* и *Syx1A*. В соответствии с изменениями транскриптома, всплеск перекодирования белков был выявлен на стадии метаморфоза при переходе от поздней куколки к раннему имаго.
3. На протеомном уровне мышинного мозга выявлено 20 участков перекодирования, принадлежащих 14 белкам. Для наборов данных человеческого мозга найдено 38 перекодированных участков в 33 белках. Полученный список последовательностей содержал 8 общих для двух видов участков, в том числе главный сайт перекодирования, кодируемый геном *CYFIP2*, три сайта глутаматных рецепторов *GRIA2* и *GRIA3*, сайты белков *COPA*, *FLNA* и *CADPS*. В первичных культурах мышинных клеток мозга для перекодированных белков, ответственных за пространственную организацию актиновых филаментов (*CYFIP2*, *FLNA*) и компонента специфического покрытия везикул (*COPA*), было показано дифференциальное перекодирование между нейронами и микроглией.
4. На трех линиях трансгенных мышей проанализировано влияние на степень перекодирования белков при таких состояниях, как FUS-опатия (отчасти моделирует нейродегенерацию, например, при боковом амиотрофическом склерозе), таупатия (гиперфосфорилирование белка тау, имитирующее признаки болезней Паркинсона, Пика и Альцгеймера), а также моделирующий психическое расстройство нокаут гена не кодирующей РНК *NEAT1*. Выяснилось, что эти модели не вносят заметного вклада в перекодирование исследуемых белковых сайтов, однако в профилях перекодирования анализируемых белков показано отличие между корой головного мозга и мозжечком.

7. СПИСОК ЛИТЕРАТУРЫ

1. Мошковский С.А., Лобас А.А., Горшков М.В. Протеогеномика единичных клеток – ближайшая перспектива // Биохимия. 2020. Vol. 85, № 2. P. 165–173.
2. Lobas A.A. et al. Proteogenomics of Malignant Melanoma Cell Lines: The Effect of Stringency of Exome Data Filtering on Variant Peptide Identification in Shotgun Proteomics // J. Proteome Res. American Chemical Society, 2018. Vol. 17, № 5. P. 1801–1811.
3. Patterson J.B., Samuel C.E. Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. // Mol. Cell. Biol. 1995. Vol. 15, № 10. P. 5376–5388.
4. Гончаров А.О. и соавторы. Редактирование РНК аденозиндезаминазами ADAR: от молекулярной пластичности белков нервной системы до патогенеза злокачественных опухолей человека // Биохимия. 2019. Vol. 84, № 8. P. 1129–1138.
5. Van Blerkom L.M. Role of viruses in human evolution. // Am. J. Phys. Anthropol. 2003. Vol. Suppl 37. P. 14–46.
6. Jin Y., Zhang W., Li Q. Origins and evolution of ADAR-mediated RNA editing // IUBMB Life. 2009. Vol. 61, № 6. P. 572–578.
7. Vitali P., Scadden A.D.J. Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis // Nat. Struct. Mol. Biol. 2010. Vol. 17, № 9. P. 1043–1050.
8. Liddicoat B.J.B.J.B.J. et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. // Science. 2015. Vol. 349, № 6252. P. 1115–1120.
9. Mannion N.M. et al. The RNA-Editing Enzyme ADAR1 Controls Innate Immune Responses to RNA // Cell Rep. Cell Press, 2014. Vol. 9, № 4. P. 1482–1494.
10. Heigwer F., Port F., Boutros M. RNA Interference (RNAi) Screening in Drosophila // Genetics. 2018. Vol. 208, № 3. P. 853–874.
11. Thomas J.M., Beal P.A. How do ADARs bind RNA? New protein-RNA structures illuminate substrate recognition by the RNA editing ADARs // BioEssays. 2017. Vol. 39, № 4.
12. Michael Lavigne G. et al. Autocrine and paracrine interferon signalling as ‘ring vaccination’ and ‘contact tracing’ strategies to suppress virus infection in a host // Proc. R. Soc. B Biol. Sci. 2021. Vol. 288, № 1945. P. 20203002.

13. Yang J.H. et al. Intracellular Localization of Differentially Regulated RNA-specific Adenosine Deaminase Isoforms in Inflammation // *J. Biol. Chem.* 2003. Vol. 278, № 46. P. 45833–45842.
14. Samuel C.E. ADARs: viruses and innate immunity. // *Curr. Top. Microbiol. Immunol.* 2012. Vol. 353. P. 163–195.
15. Sherry B. Rotavirus and reovirus modulation of the interferon response. // *J. Interferon Cytokine Res.* 2009. Vol. 29, № 9. P. 559–567.
16. Kwak S. et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself // *Anal. Chem.* 2017/11/10 / ed. Cox J. Stanford University, Biophysics Program, Stanford, California, United States of America., United States: Nature Publishing Group, 2017. Vol. 17, № 1. P. 882.
17. Hsiao Y.H.E. et al. RNA editing in nascent RNA affects pre-mRNA splicing // *Genome Res.* 2018. Vol. 28, № 6. P. 812–823.
18. Yang Y., Zhou X.X., Jin Y.F. ADAR-mediated RNA editing in non-coding RNA sequences // *Sci. China Life Sci.* 2013. Vol. 56, № 10. P. 944–952.
19. Duan Y. et al. Adaptation of A-to-I RNA editing in *Drosophila* // *PLoS Genet.* 2017. Vol. 13, № 3. P. e1006648.
20. Bass B.L. RNA editing by adenosine deaminases that act on RNA // *Annu. Rev. Biochem.* 2002. Vol. 71. P. 817–846.
21. Tan M.H. et al. Dynamic landscape and regulation of RNA editing in mammals // *Nature.* 2017. Vol. 550, № 7675. P. 249–254.
22. Keegan L. et al. ADAR RNA editing below the backbone // *Rna.* 2017. Vol. 23, № 9. P. 1317–1328.
23. Deng P. et al. Adar RNA editing-dependent and -independent effects are required for brain and innate immune functions in *Drosophila* // *Nat. Commun.* 2020. Vol. 11, № 1. P. 1580.
24. Rosenthal J.J.C., Seeburg P.H. A-to-I RNA Editing: Effects on Proteins Key to Neural Excitability // *Neuron.* Elsevier Inc., 2012. Vol. 74, № 3. P. 432–439.
25. Hoopengardner B. et al. Nervous system targets of RNA editing identified by comparative genomics // *Science* (80-.). 2003. Vol. 301, № 5634. P. 832–836.
26. Kuznetsova K.G. et al. Proteogenomics of Adenosine-to-Inosine RNA Editing in the Fruit Fly // *J. Proteome Res.* American Chemical Society, 2018. Vol. 17, № 11. P. 3889–3903.

27. Levitsky L.I. et al. Adenosine-to-Inosine RNA Editing in Mouse and Human Brain Proteomes // *Proteomics*. John Wiley & Sons, Ltd, 2019. Vol. 19, № 23. P. 1900195.
28. Keegan L.P. et al. Tuning of RNA editing by ADAR is required in *Drosophila* // *EMBO J.* 2005. Vol. 24, № 12. P. 2183–2193.
29. Oakes E. et al. Adenosine deaminase that acts on RNA 3 (adar3) binding to glutamate receptor subunit B Pre-mRNA Inhibits RNA editing in glioblastoma // *J. Biol. Chem.* 2017/02/06. American Society for Biochemistry and Molecular Biology, 2017. Vol. 292, № 10. P. 4326–4335.
30. Schumacher J.M. et al. Distribution of Tenr, an RNA-Binding Protein, in a Lattice-Like Network within the Spermatid Nucleus in the Mouse1 // *Biol. Reprod. United States*, 2005. Vol. 52, № 6. P. 1274–1283.
31. SAUNDERS L.R., BARBER G.N. The dsRNA binding protein family: critical roles, diverse cellular functions // *FASEB J.* 2003. Vol. 17, № 9. P. 961–983.
32. Chalk A.M. et al. The majority of A-to-I RNA editing is not required for mammalian homeostasis // *Genome Biol.* 2019. Vol. 20, № 1. P. 268.
33. Jain M. et al. RNA editing of Filamin A pre- mRNA regulates vascular contraction and diastolic blood pressure // *EMBO J.* 2018. Vol. 37, № 19.
34. Peng X. et al. A-to-I RNA Editing Contributes to Proteomic Diversity in Cancer // *Cancer Cell*. Elsevier Inc., 2018. Vol. 33, № 5. P. 817-828.e7.
35. Graveley B.R. et al. The developmental transcriptome of *Drosophila melanogaster* // *Nature*. 2011. Vol. 471, № 7339. P. 473–479.
36. Wahlstedt H. et al. Large-scale mRNA sequencing determines global regulation of RNA editing during brain development // *Genome Res.* Cold Spring Harbor Laboratory Press, 2009. Vol. 19, № 6. P. 978–986.
37. Tariq A. et al. RNA-interacting proteins act as site-specific repressors of ADAR2-mediated RNA editing and fluctuate upon neuronal stimulation // *Nucleic Acids Res.* 2013. Vol. 41, № 4. P. 2581–2593.
38. Levanon E.Y. et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome // *Nat. Biotechnol.* 2004. Vol. 22, № 8. P. 1001–1005.
39. Kim D.D.Y.Y. et al. Widespread RNA Editing of Embedded Alu Elements in the Human Transcriptome // *Genome Res.* 2004. Vol. 14, № 9. P. 1719–1725.

40. Athanasiadis A., Rich A., Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome // *PLoS Biol.* 2004. Vol. 2, № 12. P. e391.
41. Blow M. et al. A survey of RNA editing in human brain // *Genome Res.* 2004. Vol. 14, № 12. P. 2379–2387.
42. Richard Shen M., Batzer M.A., Deininger P.L. Evolution of the master Alu gene(s) // *J. Mol. Evol.* 1991. Vol. 33, № 4. P. 311–320.
43. García M.A. et al. Impact of Protein Kinase PKR in Cell Biology: from Antiviral to Antiproliferative Action // *Microbiol. Mol. Biol. Rev.* 2006. Vol. 70, № 4. P. 1032–1060.
44. George C.X. et al. Editing of cellular self-RNAs by adenosine deaminase ADAR1 suppresses innate immune stress responses // *J. Biol. Chem.* 2016. Vol. 291, № 12. P. 6158–6168.
45. Clerzius G. et al. ADAR1 Interacts with PKR during Human Immunodeficiency Virus Infection of Lymphocytes and Contributes to Viral Replication // *J. Virol.* 2009. Vol. 83, № 19. P. 10119–10128.
46. Yang J.-H. et al. Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation. // *Immunology.* 2003. Vol. 109, № 1. P. 15–23.
47. Nie Y. et al. ADAR1 Interacts with NF90 through Double-Stranded RNA and Regulates NF90-Mediated Gene Expression Independently of RNA Editing // *Mol. Cell. Biol.* 2005. Vol. 25, № 16. P. 6956–6963.
48. Hartner J.C. et al. ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. // *Nat. Immunol.* 2009. Vol. 10, № 1. P. 109–115.
49. Porath H.T. et al. Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance // *Genome Biol. BioMed Central*, 2017. Vol. 18, № 1. P. 185.
50. Pestal K. et al. Isoforms of RNA-Editing Enzyme ADAR1 Independently Control Nucleic Acid Sensor MDA5-Driven Autoimmunity and Multi-organ Development // *Immunity.* 2015. Vol. 43, № 5. P. 933–944.
51. Crow Y.J. Aicardi-Goutières syndrome. // *Handb. Clin. Neurol.* 2013. Vol. 113. P. 1629–1635.
52. Casey J.L. Control of ADAR1 Editing of Hepatitis Delta Virus RNAs. 2011. P. 123–143.
53. Sommer B. et al. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels // *Cell. Elsevier*, 1991. Vol. 67, № 1. P. 11–19.

54. Gannon H.S. et al. Identification of ADAR1 adenosine deaminase dependency in a subset of cancer cells // *Nat. Commun.* 2018. Vol. 9, № 1. P. 5450.
55. Zhang M. et al. RNA editing derived epitopes function as cancer antigens to elicit immune responses // *Nat. Commun.* Nature Publishing Group UK, 2018. Vol. 9, № 1. P. 3919.
56. Orecchini E. et al. ADAR1 restricts LINE-1 retrotransposition // *Nucleic Acids Res.* 2017. Vol. 45, № 1. P. 155–168.
57. Rice G.I. et al. Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature // *Nat. Genet.* 2012. Vol. 44, № 11. P. 1243–1248.
58. Miyamura Y. et al. Mutations of the RNA-Specific Adenosine Deaminase Gene (DSRAD) Are Involved in Dyschromatosis Symmetrica Hereditaria // *Am. J. Hum. Genet.* 2003. Vol. 73, № 3. P. 693–699.
59. Suspene R. et al. Double-Stranded RNA Adenosine Deaminase ADAR-1-Induced Hypermutated Genomes among Inactivated Seasonal Influenza and Live Attenuated Measles Virus Vaccines // *J. Virol.* 2010/12/15. American Society for Microbiology (ASM), 2011. Vol. 85, № 5. P. 2458–2462.
60. Zahn R.C. et al. A-to-G Hypermutation in the Genome of Lymphocytic Choriomeningitis Virus // *J. Virol.* 2006/10/04. American Society for Microbiology, 2006. Vol. 81, № 2. P. 457–464.
61. Taylor D.R. et al. New Antiviral Pathway That Mediates Hepatitis C Virus Replicon Interferon Sensitivity through ADAR1 // *J. Virol.* American Society for Microbiology, 2005. Vol. 79, № 10. P. 6291–6298.
62. Gandy S.Z. et al. RNA Editing of the Human Herpesvirus 8 Kaposin Transcript Eliminates Its Transforming Activity and Is Induced during Lytic Replication // *J. Virol.* 2007/10/03. American Society for Microbiology (ASM), 2007. Vol. 81, № 24. P. 13544–13551.
63. Iizasa H. et al. Editing of Epstein-Barr virus-encoded BART6 microRNAs controls their dicer targeting and consequently affects viral latency // *J. Biol. Chem.* 2010/08/17. American Society for Biochemistry and Molecular Biology, 2010. Vol. 285, № 43. P. 33358–33370.
64. Toth A.M. et al. RNA-specific adenosine deaminase ADAR1 suppresses measles virus-induced apoptosis and activation of protein kinase PKR // *J. Biol. Chem.* 2009/08/25. American Society for Biochemistry and Molecular Biology, 2009. Vol. 284, № 43. P. 29350–29356.
65. Gelinas J.-F. et al. Enhancement of Replication of RNA Viruses by ADAR1 via RNA Editing and

- Inhibition of RNA-Activated Protein Kinase // J. Virol. American Society for Microbiology, 2011. Vol. 85, № 17. P. 8460–8466.
66. Picardi E. et al. REDportal: A comprehensive database of A-to-I RNA editing events in humans // Nucleic Acids Res. 2017. Vol. 45, № D1.
 67. Ganem N.S., Ben-Asher N., Lamm A.T. In cancer, A-to-I RNA editing can be the driver, the passenger, or the mechanic. // Drug Resist. Updat. 2017. Vol. 32. P. 16–22.
 68. Xu L. Di, Öhman M. ADAR1 editing and its role in cancer // Genes (Basel). MDPI, 2019. Vol. 10, № 1. P. 12.
 69. Ishizuka J.J. et al. Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade // Nature. 2019. Vol. 565, № 7737. P. 43–48.
 70. Palladino M.J. et al. A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity // Cell. 2000. Vol. 102, № 4. P. 437–449.
 71. Khan A. et al. Membrane and synaptic defects leading to neurodegeneration in Adar mutant *Drosophila* are rescued by increased autophagy // BMC Biol. 2012/01/05 / ed. Levy Y.K. Stanford University, Biophysics Program, Stanford, California, United States of America.: Nature Publishing Group, 2020. Vol. 18, № 1. P. 15.
 72. Paro S. et al. Regulation and functions of ADAR in *Drosophila* // Current Topics in Microbiology and Immunology. 2012. Vol. 353, № 1. 221–236 p.
 73. Palladino M.J. et al. dADAR, a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing // Rna. 2000. Vol. 6, № 7. P. 1004–1018.
 74. Heale B.S.E. et al. Editing independent effects of ADARs on the miRNA/siRNA pathways. // EMBO J. 2009. Vol. 28, № 20. P. 3145–3156.
 75. Behm M., Öhman M. RNA Editing: A Contributor to Neuronal Dynamics in the Mammalian Brain // Trends Genet. 2016. Vol. 32, № 3. P. 165–175.
 76. Young P.E. et al. Epigenetic differences between monozygotic twins discordant for amyotrophic lateral sclerosis (ALS) provide clues to disease pathogenesis. // PLoS One. 2017. Vol. 12, № 8. P. e0182638.
 77. Gallo A. et al. ADAR RNA editing in human disease; more to it than meets the I // Hum. Genet. 2017. Vol. 136, № 9. P. 1265–1278.

78. Filippini A. et al. The Good and the Bad of Glutamate Receptor RNA Editing // *Mol. Neurobiol.* 2017. Vol. 54, № 9. P. 6795–6805.
79. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. // *Nat. Rev. Mol. Cell Biol.* 2016. Vol. 17, № 2. P. 83–96.
80. Peng P.L. et al. ADAR2-Dependent RNA Editing of AMPA Receptor Subunit GluR2 Determines Vulnerability of Neurons in Forebrain Ischemia // *Neuron.* 2006. Vol. 49, № 5. P. 719–733.
81. Kawahara Y. et al. RNA editing and death of motor neurons: There is a glutamate-receptor defect in patients with amyotrophic lateral sclerosis // *Nature.* 2004. Vol. 427, № 6977. P. 801.
82. Kwak S., Kawahara Y. Deficient RNA editing of GluR2 and neuronal death in amyotrophic lateral sclerosis // *J. Mol. Med.* 2005. Vol. 83, № 2. P. 110–120.
83. Yamashita T. et al. A role for calpain-dependent cleavage of TDP-43 in amyotrophic lateral sclerosis pathology // *Nat. Commun.* 2012. Vol. 3, № 1. P. 1307.
84. Akbarian S., Smith M.A., Jones E.G. Editing for an AMPA receptor subunit RNA in prefrontal cortex and striatum in Alzheimer's disease, Huntington's disease and schizophrenia. // *Brain Res.* 1995. Vol. 699, № 2. P. 297–304.
85. Maas S. et al. Underediting of glutamate receptor GluR-B mRNA in malignant gliomas // *Proc. Natl. Acad. Sci.* 2001. Vol. 98, № 25. P. 14687–14692.
86. Lomeli H. et al. Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. // *Science.* 1994. Vol. 266, № 5191. P. 1709–1713.
87. Barbon A. et al. Acute spinal cord injury persistently reduces R/G RNA editing of AMPA receptors. // *J. Neurochem.* 2010. Vol. 114, № 2. P. 397–407.
88. Bonini D. et al. Chronic glutamate treatment selectively modulates AMPA RNA editing and ADAR expression and activity in primary cortical neurons // *RNA Biol.* 2015. Vol. 12, № 1. P. 43–53.
89. Russo I. et al. AMPA Receptor Properties are Modulated in the Early Stages Following Pilocarpine-induced Status Epilepticus // *NeuroMolecular Med.* 2013. Vol. 15, № 2. P. 324–338.
90. Vollmar W. et al. RNA editing (R/G site) and flip–flop splicing of the AMPA receptor subunit GluR2 in nervous tissue of epilepsy patients // *Neurobiol. Dis.* 2004. Vol. 15, № 2. P. 371–379.
91. Barbon A. et al. Regulation of Editing and Expression of Glutamate α -Amino-Propionic-Acid (AMPA)/Kainate Receptors by Antidepressant Drugs // *Biol. Psychiatry.* 2006. Vol. 59, № 8. P.

713–720.

92. Barbon A. et al. Chronic antidepressant treatments induce a time-dependent up-regulation of AMPA receptor subunit protein levels // *Neurochem. Int.* 2011. Vol. 59, № 6. P. 896–905.
93. Khmresh K. et al. Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease // *RNA*. 2016. Vol. 22, № 2. P. 290–302.
94. Gaisler-Salomon I. et al. Hippocampus-specific deficiency in RNA editing of GluA2 in Alzheimer's disease // *Neurobiol. Aging*. 2014. Vol. 35, № 8. P. 1785–1791.
95. Caracciolo L. et al. Kainate Receptor RNA Editing is Markedly Altered by Acute Spinal Cord Injury // *J. Mol. Neurosci.* 2013. Vol. 51, № 3. P. 903–910.
96. Kortenbruck G. et al. RNA Editing at the Q/R Site for the Glutamate Receptor Subunits GLUR2, GLUR5, and GLUR6 in Hippocampus and Temporal Cortex from Epileptic Patients // *Neurobiol. Dis.* 2001. Vol. 8, № 3. P. 459–468.
97. Brande-Eilat N. et al. Acquisition of conditioned fear is followed by region-specific changes in RNA editing of glutamate receptors // *Stress*. 2015. Vol. 18, № 3. P. 309–318.
98. Egebjerg J., Heinemann S.F. Ca²⁺ permeability of unedited and edited versions of the kainate selective glutamate receptor GluR6. // *Proc. Natl. Acad. Sci. U. S. A.* 1993. Vol. 90, № 2. P. 755–759.
99. Köhler M. et al. Determinants of Ca²⁺ permeability in both TM1 and TM2 of high affinity kainate receptor channels: diversity by RNA editing. // *Neuron*. 1993. Vol. 10, № 3. P. 491–500.
100. Silberberg G. et al. Deregulation of the A-to-I RNA editing mechanism in psychiatric disorders // *Hum. Mol. Genet.* 2012. Vol. 21, № 2. P. 311–321.
101. Li B. et al. Cell Type-Specific Gene Expression and Editing Responses to Chronic Fluoxetine Treatment in the In Vivo Mouse Brain and Their Relevance for Stress-Induced Anhedonia // *Neurochem. Res.* 2012. Vol. 37, № 11. P. 2480–2495.
102. Uchida H. et al. RNA editing enzyme ADAR2 is a mediator of neuropathic pain after peripheral nerve injury. // *FASEB J.* 2017. Vol. 31, № 5. P. 1847–1855.
103. Zaidan H. et al. A-to-I RNA editing in the rat brain is age-dependent, region-specific and sensitive to environmental stress across generations. // *BMC Genomics*. 2018. Vol. 19, № 1. P. 28.
104. Peng Z. et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. // *Nat. Biotechnol.* 2012. Vol. 30, № 3. P. 253–260.

105. Ramaswami G. et al. Identifying RNA editing sites using RNA sequencing data alone. // Nat. Methods. 2013. Vol. 10, № 2. P. 128–132.
106. Sakurai M. et al. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. // Nat. Chem. Biol. Nature Publishing Group, 2010. Vol. 6, № 10. P. 733–740.
107. Zhang R. et al. Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. // Nat. Methods. 2014. Vol. 11, № 1. P. 51–54.
108. Suzuki T. et al. Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. // Nat. Protoc. 2015. Vol. 10, № 5. P. 715–732.
109. Mengel-Jørgensen J., Kirpekar F. Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. // Nucleic Acids Res. 2002. Vol. 30, № 23. P. e135.
110. Ramaswami G., Li J.B. RADAR: A rigorously annotated database of A-to-I RNA editing // Nucleic Acids Res. 2014. Vol. 42, № D1. P. D109–13.
111. Kiran A.M. et al. Darned in 2013: inclusion of model organisms and linking with Wikipedia. // Nucleic Acids Res. 2013. Vol. 41, № Database issue. P. D258–61.
112. Wilhelm M. et al. Mass-spectrometry-based draft of the human proteome // Nature. 2014. Vol. 509, № 7502. P. 582–587.
113. Smith L.M. et al. Proteoform: a single term describing protein complexity // Nat. Methods. 2013. Vol. 10, № 3. P. 186–187.
114. Ansong C. et al. Proteogenomics: needs and roles to be filled by proteomics in genome annotation. // Brief. Funct. Genomic. Proteomic. 2008. Vol. 7, № 1. P. 50–62.
115. Nesvizhskii A.I. Proteogenomics: concepts, applications and computational strategies. // Nat. Methods. 2014. Vol. 11, № 11. P. 1114–1125.
116. Menon R. et al. Distinct splice variants and pathway enrichment in the cell-line models of aggressive human breast cancer subtypes. // J. Proteome Res. 2014. Vol. 13, № 1. P. 212–227.
117. Li J. et al. A bioinformatics workflow for variant peptide detection in shotgun proteomics. // Mol. Cell. Proteomics. 2011. Vol. 10, № 5. P. M110.006536.
118. Low T.Y. et al. Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. // Cell Rep. 2013. Vol. 5, № 5. P. 1469–1478.

119. Liscovitch-Brauer N. et al. Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods // *Cell. Elsevier*, 2017. Vol. 169, № 2. P. 191–202.e11.
120. Garrett S., Rosenthal J.J.C. RNA editing underlies temperature adaptation in K⁺ channels from polar octopuses // *Science* (80-.). 2012/01/05. 2012. Vol. 335, № 6070. P. 848–851.
121. Li M. et al. Widespread RNA and DNA sequence differences in the human transcriptome. // *Science*. 2011. Vol. 333, № 6038. P. 53–58.
122. Xing X. et al. Qualitative and quantitative analysis of the adult *Drosophila melanogaster* proteome. // *Proteomics*. 2014. Vol. 14, № 2–3. P. 286–290.
123. Aradska J. et al. Gel-free mass spectrometry analysis of *Drosophila melanogaster* heads. // *Proteomics*. 2015. Vol. 15, № 19. P. 3356–3360.
124. Vizcaíno J.A. et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination // *Nature Biotechnology*. 2014. Vol. 32, № 3. P. 223–226.
125. Sharma K. et al. Cell type- and brain region-resolved mouse brain proteome. // *Nat. Neurosci*. 2015. Vol. 18, № 12. P. 1819–1831.
126. Carlyle B.C. et al. A multiregional proteomic survey of the postnatal human brain // *Nat. Neurosci*. 2017. Vol. 20, № 12. P. 1787–1795.
127. Langfelder P. et al. Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. // *Nat. Neurosci*. 2016. Vol. 19, № 4. P. 623–633.
128. Altmann C. et al. Progranulin overexpression in sensory neurons attenuates neuropathic pain in mice: Role of autophagy. // *Neurobiol. Dis*. 2016. Vol. 96. P. 294–311.
129. Macron C. et al. Deep Dive on the Proteome of Human Cerebrospinal Fluid: A Valuable Data Resource for Biomarker Discovery and Missing Protein Identification // *J. Proteome Res*. 2018. Vol. 17, № 12. P. 4113–4126.
130. Wingo T.S. et al. Integrating Next-Generation Genomic Sequencing and Mass Spectrometry To Estimate Allele-Specific Protein Abundance in Human Brain. // *J. Proteome Res*. 2017. Vol. 16, № 9. P. 3336–3347.
131. Fukutomi Y. et al. Methods for staging pupal periods and measurement of wing pigmentation of *drosophila guttifera* // *J. Vis. Exp*. 2018. Vol. 2018, № 131.
132. Wiechelman K.J., Braun R.D., Fitzpatrick J.D. Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation. // *Anal. Biochem*. 1988. Vol.

- 175, № 1. P. 231–237.
133. Sambrook J., Russell D.W. Purification of Nucleic Acids by Extraction with Phenol:Chloroform // Cold Spring Harb. Protoc. 2006. Vol. 2006, № 1. P. pdb.prot4455.
 134. St Laurent G. et al. Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in *Drosophila* // Nat. Struct. Mol. Biol. Nature Publishing Group, 2013. Vol. 20, № 11. P. 1333–1339.
 135. Gramates L.S. et al. FlyBase at 25: looking to the future. // Nucleic Acids Res. 2017. Vol. 45, № D1. P. D663–D671.
 136. Hinrichs A.S. et al. UCSC Data Integrator and Variant Annotation Integrator // Bioinformatics. 2016. Vol. 32, № 9. P. 1430–1432.
 137. Yu Y. et al. The Landscape of A-to-I RNA Editome Is Shaped by Both Positive and Purifying Selection. // PLoS Genet. 2016. Vol. 12, № 7. P. e1006191.
 138. Lobas A.A. et al. Exome-based proteogenomics of HEK-293 human cell line: Coding genomic variants identified at the level of shotgun proteome. // Proteomics. 2016.
 139. Levitsky L.I. et al. Unbiased False Discovery Rate Estimation for Shotgun Proteomics Based on the Target-Decoy Approach // J. Proteome Res. 2017. Vol. 16, № 2. P. 393–397.
 140. Goloborodko A.A. et al. Pyteomics - A python framework for exploratory data analysis and rapid software prototyping in proteomics // J. Am. Soc. Mass Spectrom. 2013. Vol. 24, № 2. P. 301–304.
 141. Adusumilli R., Mallick P. Data conversion with proteoWizard msConvert // Methods Mol. Biol. 2017. Vol. 1550. P. 339–368.
 142. Levitsky L.I. et al. IdentiPy: An Extensible Search Engine for Protein Identification in Shotgun Proteomics // J. Proteome Res. 2018. Vol. 17, № 7. P. 2249–2255.
 143. Teleman J. et al. Dinosaur: A Refined Open-Source Peptide MS Feature Detector // J. Proteome Res. 2016. Vol. 15, № 7. P. 2143–2151.
 144. Ivanov M. V et al. Scavenger: A Versatile Postsearch Validation Algorithm for Shotgun Proteomics Based on Gradient Boosting // Proteomics. 2019. Vol. 19, № 3. P. 1800280.
 145. Ma C. et al. Improved Peptide Retention Time Prediction in Liquid Chromatography through Deep Learning // Anal. Chem. 2018. Vol. 90, № 18. P. 10881–10888.

146. Bittremieux W. Spectrum-utils: A Python Package for Mass Spectrometry Data Processing and Visualization // *Anal. Chem.* 2020. Vol. 92, № 1. P. 659–661.
147. Gessulat S. et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning // *Nat. Methods.* Springer US, 2019. Vol. 16, № 6. P. 509–518.
148. Levitsky L.I. et al. Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework // *J. Proteome Res.* 2019. Vol. 18, № 2. P. 709–714.
149. Craig R., Beavis R.C. A method for reducing the time required to match protein sequences with tandem mass spectra // *Rapid Commun. Mass Spectrom.* 2003. Vol. 17, № 20. P. 2310–2316.
150. Schaeffer M. et al. The neXtProt peptide uniqueness checker: a tool for the proteomics community. // *Bioinformatics.* 2017. Vol. 33, № 21. P. 3471–3472.
151. Zybaylov B.L., Florens L., Washburn M.P. Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors // *Mol. Biosyst.* 2007. Vol. 3, № 5. P. 354–360.
152. Szklarczyk D. et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets // *Nucleic Acids Res.* 2019. Vol. 47, № D1. P. D607–D613.
153. Gu Z., Eils R., Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data // *Bioinformatics.* 2016. Vol. 32, № 18. P. 2847–2849.
154. Sheynkman G.M. et al. Proteogenomics: Integrating Next-Generation Sequencing and Mass Spectrometry to Characterize Human Proteomic Variation. // *Annu. Rev. Anal. Chem.* (Palo Alto, Calif). 2016. Vol. 9, № 1. P. 521–545.
155. Moshkovskii S.A. et al. Identification of Single Amino Acid Substitutions in Proteogenomics // *Biochem.* 2018. Vol. 83, № 3. P. 250–258.
156. Szklarczyk D. et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. // *Nucleic Acids Res.* 2017. Vol. 45, № D1. P. D362–D368.
157. Wang Z., Chapman E.R. Rat and Drosophila synaptotagmin 4 have opposite effects during SNARE-catalyzed membrane fusion. // *J. Biol. Chem.* 2010. Vol. 285, № 40. P. 30759–30766.
158. Fergestad T., Broadie K. Interaction of stoned and synaptotagmin in synaptic vesicle endocytosis. // *J. Neurosci.* 2001. Vol. 21, № 4. P. 1218–1227.
159. González-Gaitán M., Jäckle H. Role of Drosophila alpha-adaptin in presynaptic vesicle recycling.

- // Cell. 1997. Vol. 88, № 6. P. 767–776.
160. Guichet A. et al. Essential role of endophilin A in synaptic vesicle budding at the *Drosophila* neuromuscular junction // EMBO J. 2002. Vol. 21, № 7. P. 1661–1672.
 161. Verstreken P. et al. Endophilin mutations block clathrin-mediated endocytosis but not neurotransmitter release. // Cell. 2002. Vol. 109, № 1. P. 101–112.
 162. Coyle I.P. et al. Nervous wreck, an SH3 adaptor protein that interacts with Wsp, regulates synaptic growth in *Drosophila*. // Neuron. 2004. Vol. 41, № 4. P. 521–534.
 163. Renden R. et al. *Drosophila* CAPS is an essential gene that regulates dense-core vesicle release and synaptic vesicle fusion // Neuron. 2001. Vol. 31, № 3. P. 421–437.
 164. Pongs O. et al. Frequentin--a novel calcium-binding protein that modulates synaptic efficacy in the *Drosophila* nervous system. // Neuron. 1993. Vol. 11, № 1. P. 15–28.
 165. Söllner T. et al. SNAP receptors implicated in vesicle targeting and fusion // Nature. 1993. Vol. 362, № 6418. P. 318–324.
 166. Hu K. et al. Action of complexin on SNARE complex // J. Biol. Chem. 2002. Vol. 277, № 44. P. 41652–41656.
 167. Neely G.G. et al. TrpA1 regulates thermal nociception in *Drosophila* // PLoS One. 2011. Vol. 6, № 8. P. e24343.
 168. Heinrich L., Ryglewski S. Different functions of two putative *Drosophila* $\alpha 2\delta$ subunits in the same identified motoneurons. // Sci. Rep. 2020. Vol. 10, № 1. P. 13670.
 169. Oswald D. et al. Cooperation of Syd-1 with Neurexin synchronizes pre- with postsynaptic assembly. // Nat. Neurosci. 2012. Vol. 15, № 9. P. 1219–1226.
 170. Dubowy C., Sehgal A. Circadian rhythms and sleep in *Drosophila melanogaster* // Genetics. 2017. Vol. 205, № 4. P. 1373–1397.
 171. Bakthavachalu B. et al. RNP-Granule Assembly via Ataxin-2 Disordered Domains Is Required for Long-Term Memory and Neurodegeneration // Neuron. 2018. Vol. 98, № 4. P. 754–766.e4.
 172. Liu Y., Beyer A., Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance // Cell. 2016. Vol. 165, № 3. P. 535–550.
 173. Chen X. et al. Three-dimensional structure of the complexin/SNARE complex. // Neuron. 2002. Vol. 33, № 3. P. 397–409.

174. Neely G.G. et al. A genome-wide *Drosophila* screen for heat nociception identifies $\alpha 2\delta 3$ as an evolutionarily conserved pain gene. // *Cell*. 2010. Vol. 143, № 4. P. 628–638.
175. Griffin N.M. et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. // *Nat. Biotechnol.* 2010. Vol. 28, № 1. P. 83–89.
176. Paoletti A.C. et al. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors // *Proc. Natl. Acad. Sci.* 2006. Vol. 103, № 50. P. 18928–18933.
177. Ishihama Y. et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. // *Mol. Cell. Proteomics*. 2005. Vol. 4, № 9. P. 1265–1272.
178. Weissenhorn W. Crystal structure of the endophilin-A1 BAR domain. // *J. Mol. Biol.* 2005. Vol. 351, № 3. P. 653–661.
179. Stanishneva-Konovalova T.B. et al. The Role of BAR Domain Proteins in the Regulation of Membrane Dynamics. // *Acta Naturae*. Vol. 8, № 4. P. 60–69.
180. Masuda M. et al. Endophilin BAR domain drives membrane curvature by two newly identified structure-based mechanisms. // *EMBO J.* 2006. Vol. 25, № 12. P. 2889–2897.
181. Kuznetsova K.G. et al. Brain Proteome of *Drosophila melanogaster* Is Enriched with Nuclear Proteins // *Biochem.* 2019. Vol. 84, № 1. P. 71–78.
182. Casas-Vila N. et al. The developmental proteome of *Drosophila melanogaster* // *Genome Res.* 2017. Vol. 27, № 7. P. 1273–1285.
183. Rolfs Z. et al. Comment on “a subset of HLA-I peptides are not genomically templated: Evidence for cis- And trans-spliced peptide ligands” // *Sci. Immunol.* 2019. Vol. 4, № 38.
184. Doronkin S., Djagaeva I., Beckendorf S.K. The COP9 signalosome promotes degradation of Cyclin E during early *Drosophila* oogenesis // *Dev. Cell*. 2003. Vol. 4, № 5. P. 699–710.
185. Gui J., Huang Y., Shimmi O. Scribbled Optimizes BMP Signaling through Its Receptor Internalization to the Rab5 Endosome and Promote Robust Epithelial Morphogenesis // *PLoS Genet.* 2016. Vol. 12, № 11. P. e1006424.
186. Zhang Y. et al. A role for *Drosophila* ATX2 in activation of PER translation and circadian behavior // *Science (80-.)*. 2013. Vol. 340, № 6134. P. 879–882.
187. Percy A.J. et al. A standardized kit for automated quantitative assessment of candidate protein

- biomarkers in human plasma // *Bioanalysis*. 2015. Vol. 7, № 23. P. 2991–3004.
188. Lo Giudice C. et al. Investigating RNA editing in deep transcriptome datasets with REDIttools and REDIportal // *Nat. Protoc.* 2020. Vol. 15, № 3. P. 1098–1131.
 189. Huang H. et al. RNA Editing of the IQ Domain in Ca v1.3 Channels Modulates Their Ca²⁺-Dependent Inactivation // *Neuron*. 2012. Vol. 73, № 2. P. 304–316.
 190. Wen W., Lin C.Y., Niu L. R/G editing in GluA2Rflop modulates the functional difference between GluA1 flip and flop variants in GluA1/2R heteromeric channels // *Sci. Rep.* 2017. Vol. 7, № 1. P. 13654.
 191. Bass B.L., Weintraub H. An unwinding activity that covalently modifies its double-stranded RNA substrate. // *Cell*. 1988. Vol. 55, № 6. P. 1089–1098.
 192. Sapiro A.L. et al. Illuminating spatial A-to-I RNA editing signatures within the Drosophila brain // *Proc. Natl. Acad. Sci. U. S. A.* 2019. Vol. 116, № 6. P. 2318–2327.
 193. Godugu B. et al. Effect of N-terminal glutamic acid and glutamine on fragmentation of peptide ions // *J. Am. Soc. Mass Spectrom.* 2010. Vol. 21, № 7. P. 1169–1176.
 194. Chen Z. et al. Structure and control of the actin regulatory WAVE complex. // *Nature*. 2010. Vol. 468, № 7323. P. 533–538.
 195. Han K. et al. Fragile X-like behaviors and abnormal cortical dendritic spines in cytoplasmic FMR1-interacting protein 2-mutant mice. // *Hum. Mol. Genet.* 2015. Vol. 24, № 7. P. 1813–1823.
 196. Nakashima M. et al. De novo hotspot variants in CYFIP2 cause early-onset epileptic encephalopathy. // *Ann. Neurol.* 2018. Vol. 83, № 4. P. 794–806.
 197. Kirkpatrick S.L. et al. Cytoplasmic FMR1-Interacting Protein 2 Is a Major Genetic Factor Underlying Binge Eating // *Biol. Psychiatry*. 2017. Vol. 81, № 9. P. 757–769.
 198. Loo D.T., Kanner S.B., Aruffo A. Filamin Binds to the Cytoplasmic Domain of the β 1 -Integrin // *J. Biol. Chem.* 1998. Vol. 273, № 36. P. 23304–23312.
 199. Levanon E.Y. Evolutionarily conserved human targets of adenosine to inosine RNA editing // *Nucleic Acids Res.* 2005. Vol. 33, № 4. P. 1162–1168.
 200. Orci L. et al. Bidirectional transport by distinct populations of COPI-coated vesicles. // *Cell*. 1997. Vol. 90, № 2. P. 335–349.
 201. O’Neil R.T. et al. Comparative analysis of A-to-I editing in human and non-human primate brains

- reveals conserved patterns and context-dependent regulation of RNA editing // *Mol. Brain*. 2017. Vol. 10, № 1. P. 11.
202. BRATT E. Coordination of editing and splicing of glutamate receptor pre-mRNA // *RNA*. 2003. Vol. 9, № 3. P. 309–318.
 203. Eggington J.M., Greene T., Bass B.L. Predicting sites of ADAR editing in double-stranded RNA. // *Nat. Commun*. 2011. Vol. 2. P. 319.
 204. Burns C.M. et al. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. // *Nature*. 1997. Vol. 387, № 6630. P. 303–308.
 205. Hackler E.A. et al. 5-HT(2C) receptor RNA editing in the amygdala of C57BL/6J, DBA/2J, and BALB/cJ mice. // *Neurosci. Res*. 2006. Vol. 55, № 1. P. 96–104.
 206. Bachmann G., Petereit H. Beta-trace protein as sensitive marker for liquorrhea. // *Acta Neurol. Scand*. 2004. Vol. 110, № 5. P. 337–338; author reply 339.
 207. Vance C. et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. // *Science*. 2009. Vol. 323, № 5918. P. 1208–1211.
 208. Shelkovnikova T.A. et al. Fused in sarcoma (FUS) protein lacking nuclear localization signal (NLS) and major RNA binding motifs triggers proteinopathy and severe motor phenotype in transgenic mice. // *J. Biol. Chem*. 2013. Vol. 288, № 35. P. 25266–25274.
 209. Sotiropoulos I. et al. Atypical, non-standard functions of the microtubule associated Tau protein. // *Acta Neuropathol. Commun*. 2017. Vol. 5, № 1. P. 91.
 210. Roman A.Y. et al. Zinc Induces Temperature-Dependent Reversible Self-Assembly of Tau. // *J. Mol. Biol*. 2019. Vol. 431, № 4. P. 687–695.
 211. Kukharsky M.S. et al. Long non-coding RNA Neat1 regulates adaptive behavioural response to stress in mice // *Transl. Psychiatry*. 2020. Vol. 10, № 1. P. 171.
 212. Katsel P. et al. The expression of long noncoding RNA NEAT1 is reduced in schizophrenia and modulates oligodendrocytes transcription. // *NPJ Schizophr*. 2019. Vol. 5, № 1. P. 3.
 213. Nakagawa S. et al. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. // *J. Cell Biol*. 2011. Vol. 193, № 1. P. 31–39.
 214. Mann M. et al. The coming age of complete, accurate, and ubiquitous proteomes. // *Mol. Cell*. 2013. Vol. 49, № 4. P. 583–590.

215. Chick J.M. et al. A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. // Nat. Biotechnol. 2015. Vol. 33, № 7. P. 743–749.
216. Nagaraj N. et al. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. // Mol. Cell. Proteomics. 2012. Vol. 11, № 3. P. M111.013722.
217. Lisitsa A. et al. Profiling proteoforms: promising follow-up of proteomics for biomarker discovery // Expert Rev. Proteomics. 2014. Vol. 11, № 1. P. 121–129.
218. Toby T.K., Fornelli L., Kelleher N.L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. // Annu. Rev. Anal. Chem. (Palo Alto. Calif). 2016. Vol. 9, № 1. P. 499–519.

8. БЛАГОДАРНОСТИ

Автор благодарит руководителя и коллег за помощь в проведении работы и подготовке диссертации. В частности, автор благодарен проф. Згоде В.Г. и к.б.н. Новиковой С.Е. за проведение масс-спектрометрического анализа, к.б.н. Фарафоновой Т.Е. за проведение синтеза пептидов, к.б.н. Пятницкому М.А., Левицкому Л.И. и к.ф.-м.н. Иванову М.В. за консультации и помощь при обработке данных. Кроме того, автор благодарит к.ф.-м.н. Горшкова М.В., Гончарова А.О., Ильину И.Ю., к.б.н. Карпова Д.С., Кузнецову К.Г., проф. Медведева А.Е., к.б.н. Курбатова Л.К., к.б.н. Кухарского М.С. и других коллег, участвовавших в проекте или помогавших его проведению.

9. ФИНАНСИРОВАНИЕ

Работа поддержана грантом Российского научного фонда № 17-15-01229 (руководитель Мошковский С.А.).

10.ПРИЛОЖЕНИЕ

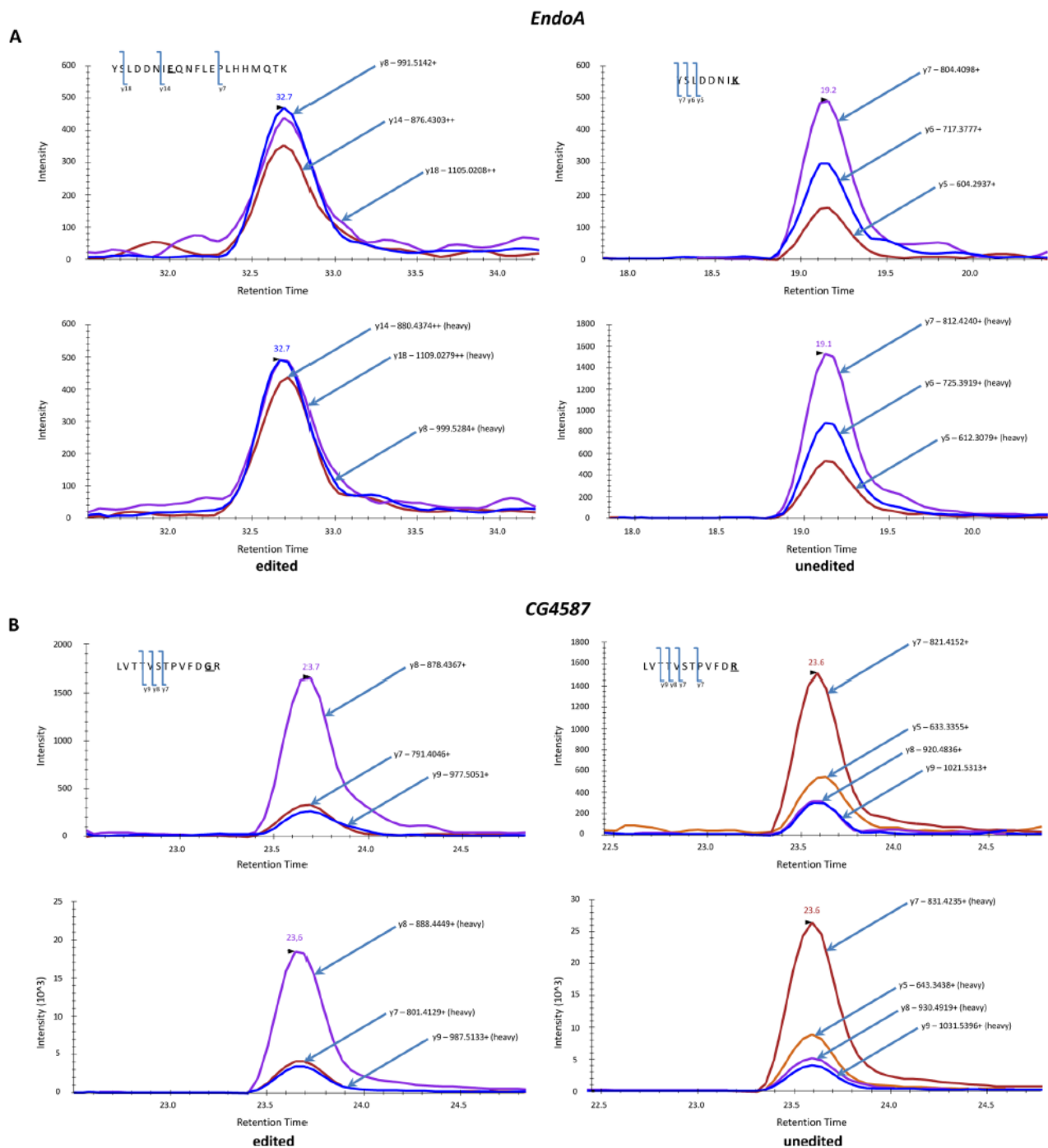
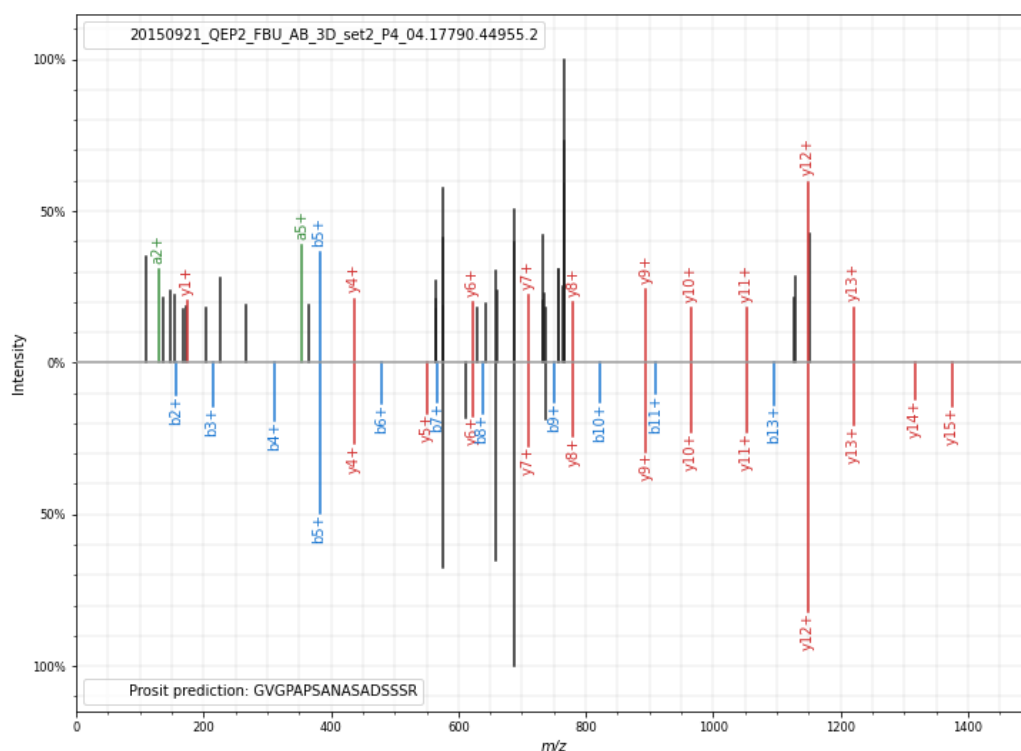
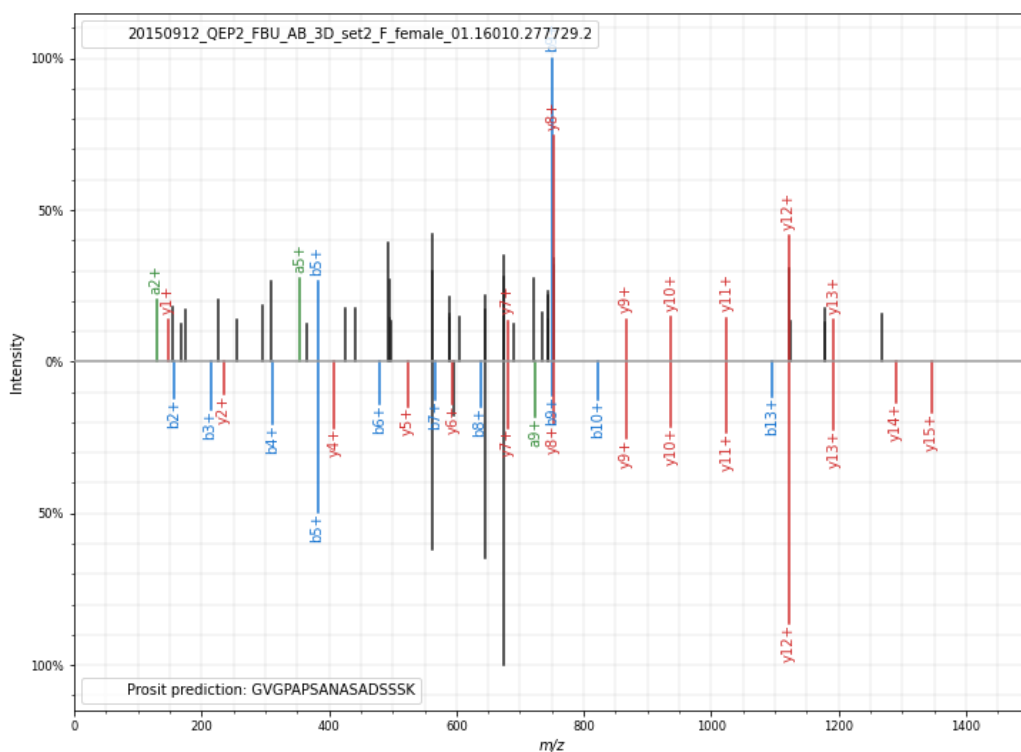


Рисунок П1. Сигнал от перекодированных (edited) и кодируемых в геноме (unedited) пептидов и их изотопно меченных стандартов для продуктов генов А) EndoA и В) stol (CG4587) в образцах головного мозга плодовой мушки. По оси х указано время удерживания (retention time), по оси у – интенсивность сигнала (intensity).

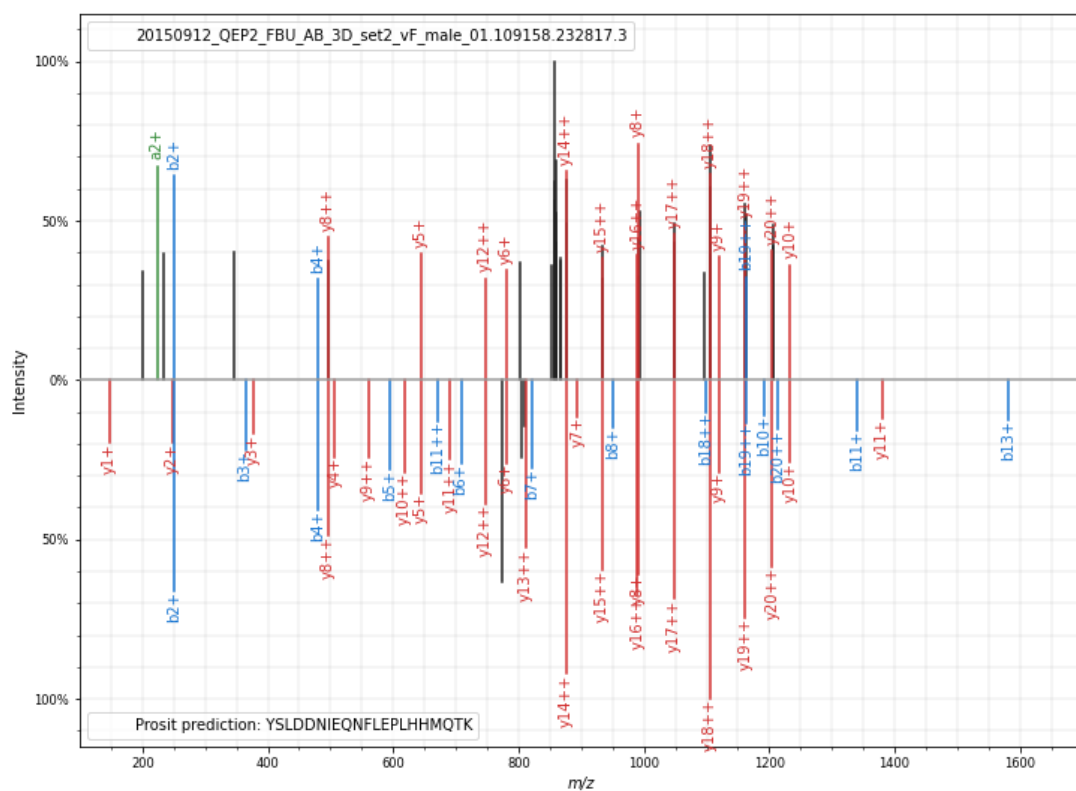


A

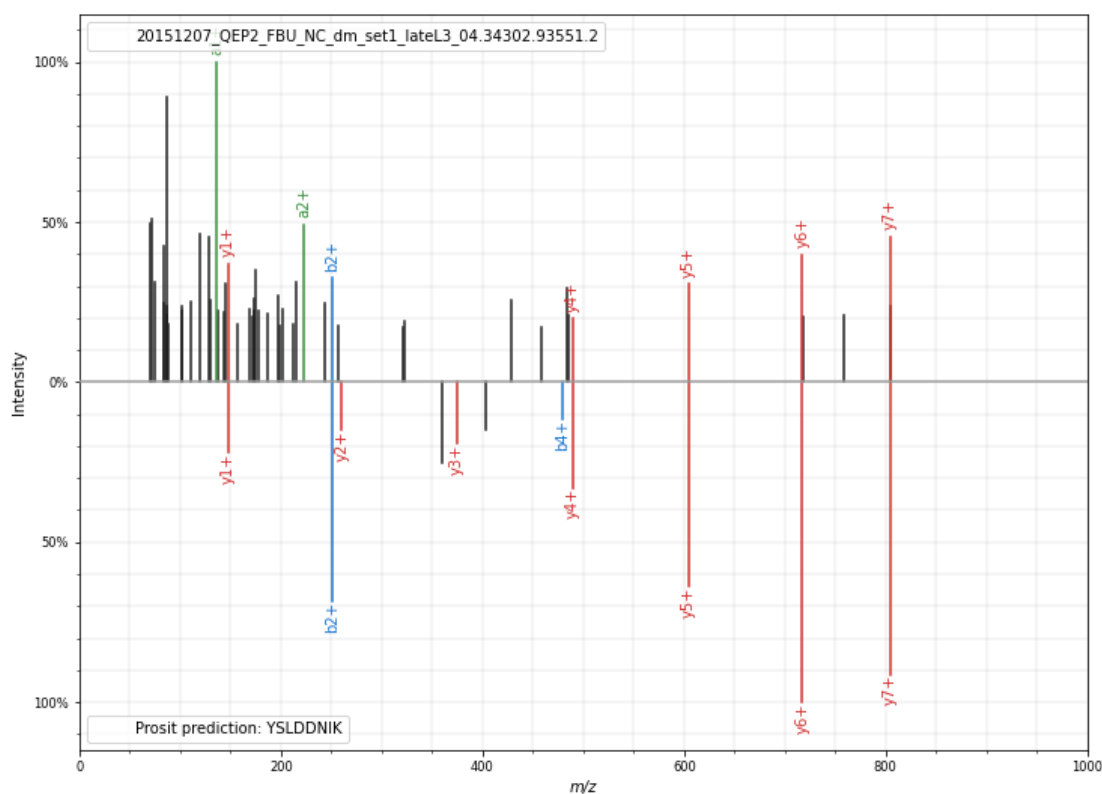


B

Рисунок П2. Типовые примеры экспериментальных МС/МС спектров и предсказанные спектры фрагментации для идентифицированных перекодированных пептидов и соответствующих им канонических вариантов. А – для перекодированного варианта продукта гена *Atx*. В – для кодируемого в геноме варианта продукта гена *Atx* (по оси *x* – величина *m/z*, по оси *y* – интенсивность сигнала).



A



B

Рисунок ПЗ. Типовые примеры экспериментальных МС/МС спектров и предсказанные спектры фрагментации для идентифицированных перекодированных пептидов и соответствующих им канонических вариантов. А – для перекодированного варианта продукта гена EndoA. В – для кодируемого в геноме варианта продукта гена EndoA (по оси x – величина m/z , по оси y – интенсивность сигнала).

Таблица П4. Перекодированные пептиды плодовой мушки, полученные двумя поисковыми системами по трем базам данных.

<i>Последовательность пептида</i>	Ген	Идентификатор UniProt	Геномные координаты	Ранг открытого поиска
<i>ADSLVESGQFDTAGIQEK</i>	alpha-Spec	M9PGV6	chr3L:1782670:A/G:Q1700R	-
<i>LVTTVSTPVFDGR</i>	stol	A8DZ06	chr2L:15651359:A/G:R489G	36
<i>IEYHVEHAVDYVQTATQDTK</i>	Syx1A	Q24547	chr3R:24103714:A/G:M244V	140
<i>GVGPAPSANASADSSSR</i>	Atx2	Q8SWR8-2	chr3R:15413305:A/G:K398R	-
<i>LVSVLESTLSK</i>	Cadps	Q9NHE5-2	chr4:1265107:A/G:M1234V	140
<i>NQMETQVNELK</i>	cpx	E1JJ33	chr3R:4297504:A/G:I125M	120
<i>YSLDDNIEQNFLPLHHMQTK</i>	EndoA	B5RIU6	chr3R:18907675:A/G:K137E	15
<i>YLLQIWPQQAQHQQR</i>	RhoGAP1 00F	Q9V9S7	chr3R:31837086:A/G:Q1153R	-
<i>MTVAHMMWFDNQIIEADTTEDQSGVQCDR</i>	Atpalpha	P13607-7	chr3R:20965039:A/G:Y429C	66
<i>EDIQANQLVMGEFEVSDVPGQI IDYVAR</i>	bai	Q8SXY6	chr3R:25039627:A/G:I60V	-
<i>AIAGVLTPLNPLWALR</i>	CG1882	Q5U191	chr2R:7813863:A/G:R240G	-
<i>FIYVSEEEELLAVAK</i>	CG5862	Q9VDD1	chr3R:21216016:A/G:K267E	-
<i>LAVHEIVDANGAGDAFVGGFL SQFVQ GK</i>	AdenoK	Q9VU39	chr3L:13020448:A/G:T294A	-
<i>IIHLLNDQHMGVVTTAAASLIDA LVK</i>	AP-2alpha	P91926	chr2L:411895:A/G:T207A	-
<i>LQQSSVGP GDPVR</i>	brp	A1Z7V1	chr2R:9534783:A/G:S1170G	93
<i>GALSSAGYHLNNR</i>	CalpB	M9NE73	chr3L:9892134:A/G:N856S	-
<i>LHAGDDEEADDPIR</i>	Calx	Q9VDG5	chr3R:20993853:A/G:Q429R	-
<i>TVGETETFLT GNEITIK</i>	CG10077	Q7KU78	chr3L:6954754:A/G:S142G	93
<i>TAGPLPGGR</i>	CG10508	M9PFU5	chr3L:21213316:A/G:S237G	-
<i>AIHALAAGASSSAEQDGAYNPW PHFR</i>	CG13630	Q9VC48	chr3R:24719046:A/G:N56S	-
<i>NQIETQVSELK</i>	cpx	E1JJ33	chr3R:4297518:A/G:N130S	247
<i>DLVATVLDMK</i>	eag	Q0KHS8	chrX:15004413:A/G:I990V	-
<i>LTEPLAPLK</i>	hig	Q09101-2	chr2R:9229936:A/G:Q716R	-
<i>ALQELEHIDDEADQLGIGFVK</i>	l(2)01289	Q0E9N2	chr2R:6733210:A/G:N196H	-
<i>LIILCVEALGQDR</i>	Mp	E1JIE1	chr3L:7051583:A/G:S969G	93
<i>LATIVEYMPALAGQLGQEFFDQ K</i>	Pp2A-29B	H5V8B8	chr2L:8368474:A/G:I426V	-
<i>SLLESSLHR</i>	Prm	A4V1N8	chr3L:8740095:A/G:Q264R	-
<i>DPWPVPR</i>	Sec23	Q9VNF8	chr3R:5650231:A/G:Q264R	-
<i>LAASQQVAVVHQHPSDQAGES GHK</i>	sif	Q8IQ62	chr3L:5740525:A/G:R2151G	36
<i>CTLNPYYNESFSFEVPFEQMOK</i>	Syt1	P21521-4	chr2L:2785542:A/G:I405M	-
<i>HDGGVIQVDDR</i>	Unc-89	A8DYP0	chr2R:24006555:A/G:E1351G	-
<i>RKPPQTPAEGEGDPEFIK</i>	up	M9NH07	chrX:13599412:A/G:K21R	-
<i>KKQQGGGGGGSGGGGGSGG GGGSR</i>	shf	Q9W3W5	chrX:6774650:A/G:N101S	-
<i>AEKPEPTENVLEPLHR</i>	CG2267	Q9VA51	chr3R:30628270:A/G:K170E	15
<i>EAPILVVAPHSSYVDSILVVASG PPSMVAK</i>	CG32699	Q0KHU5	chrX:9574655:A/G:I175M	-
<i>LPVNLGLSSILVPHGVDLDEPD VK</i>	stol	A8DZ06	chr2L:15652622:A/G:S169G	-

<i>ALELIAER</i>	CG8783	Q9VUB4	chr3L:14000997:A/G:T191A	-
<i>PQLVNNYSYVDDK</i>	dj-1beta	Q9VA37	chr3R:30807643:A/G:K128R	-
<i>IGHSLR</i>	Dyb	Q0E9B2	chr2R:12448756:A/G:Q>R	-
<i>VDLLFR</i>	GluClalph a	E1JIQ2	chr3R:19763634:A/G:K241R	-
<i>GLTNSIAVVVATSTCPR</i>	hig	Q09101	chr2R:9229936:A/G:Q716R	-
<i>MDITWFPFDDQHCEMK</i>	nAChRalp ha6	Q8IPE2	chr2L:9809297:A/G:I156M	-
<i>KTAGGSLEMFMTEDQK</i>	para	M9MS79	chrX:16473610:A/G:K1576T	270
<i>LSFRASGGSNGR</i>	Pde1c	B7YZV2	chr2L:11830365:A/G:S314R	-
<i>VEIELDsvR</i>	Prm	A4VIN8	chr3L:8740095:A/G:Q264R	-
<i>GFQPGPSVVGR</i>	pyd	A8JQV8	chr3R:8837191:A/G:S1247G	-
<i>NTCVFGGAPR</i>	Rm62	P19109-2	chr3R:6004380:A/G:K394R	-
<i>REGTTSgDKR</i>	rnhl	A1Z768	chr2R:7931822:A/G:K123E	-
<i>VTKGGAGATGAAGK</i>	RpL34b	Q9VHE5	chr3R:9400495:A/G:K150T	270
<i>SSNQAMSLAMLR</i>	Sh	P08510-3	chrX:17950057:A/G:I360M	-
<i>MDPMIYMfPR</i>	shakB	P33085	chrX:20773442:A/G:I232M	-
<i>MALTSYDQIPSELAPYAFVEFT</i>	stnB	X2JGF5	chrX:22982709:A/G:T1186A	-
<i>MPATQVSHATVR</i>	Syn	E2QCY9	chr3R:10195984:A/G:S110G	93
<i>EGLLQR</i>	Tfb4	Q9VPX4	chr2L:1150207:A/G:N111S	247
<i>QQLGSILMSAPR</i>	VhaSFD	Q9V3J1	chr2L:16724571:A/G:K292R	-
<i>DHCIAMVQCR</i>	VhaSFD	Q9V3J1-3	chr2L:16724571:A/G:K292T	270
<i>EQLATLMDTLR</i>	zip	Q99323-1	chr2R:24996026:A/G:K707T	-
<i>AHSDLTGVK</i>	Calx	Q9VDG5	chr3R:20993841:A/G:E425G	-
<i>SHHSSSRGKYER</i>	CG31211	Q86B99	chr3R:11953544:A/G:S770G	93
<i>QDSPSQPTPTIVVKDSSNAKLNH TK</i>	Cp190	Q24478	chr3R:15273480:A/G:I327V	-
<i>LKNQMETQVNELK</i>	cpx	E1JJ33	chr3R:4297504:A/G:I325M	-
<i>APNPLVYDPK</i>	Cyp6a8	Q27593	chr2R:14887484:A/G:T480A	-
<i>QQOQLQHSQQLPR</i>	InR	P09208	chr3R:21579766:A/G:N239S	-
<i>FLAGEADKK</i>	Tm1	P49455-2	chr3R:15296683:A/G:E156G	-
<i>ASTLSSSTSPSMSSSSASNHPGHS QSQR</i>	Rab3-GEF	Q9VXY2	chrX:15111879:A/G:Q2031R	-
<i>IMTFVGGPCSR</i>	Sec23	Q9VNF8	chr3R:5650351:A/G:Q304R	-
<i>RQTDGQWDLNGYHLVTLGDR</i>	syd	Q9GQF1-2	chr3L:7928680:A/G:S983G	93
<i>IAVMQNGKRLK</i>	Syt1	X2J4C1	chr2L:2785658:A/G:I367V	-
<i>excluded peptides</i>				
<i>VVETVTEdGEPK</i>	sls	Q9I7U4	chr3L:2063810:A/G:K>E	-
<i>HNNIGGGGGDGNNSNSISGSSNN K</i>	Tango5	Q9W2S1	chrX:10762318:A/G:N32D:- NM_001169245'	5
<i>IINLDGPQPTK</i>	CG42321	B7YZF5	chr2R:13422657:A/G:N222D:- NM_001144184'	
<i>DGLCQLCAPNEPGVFIGK</i>	Fatp	Q9VKU1	chr2L:10512299:A/G:N417D:- NM_164934'	
<i>DSDGDGVGDLNGIAEQLPYLK</i>	Mal-A5	Q5U124		
<i>DQMEGDYSGIEHLVVPGVMT IK</i>	mRpL13	Q9VJ39		
<i>DQMMAVCSHTIEK</i>	CG3483	Q9W172		

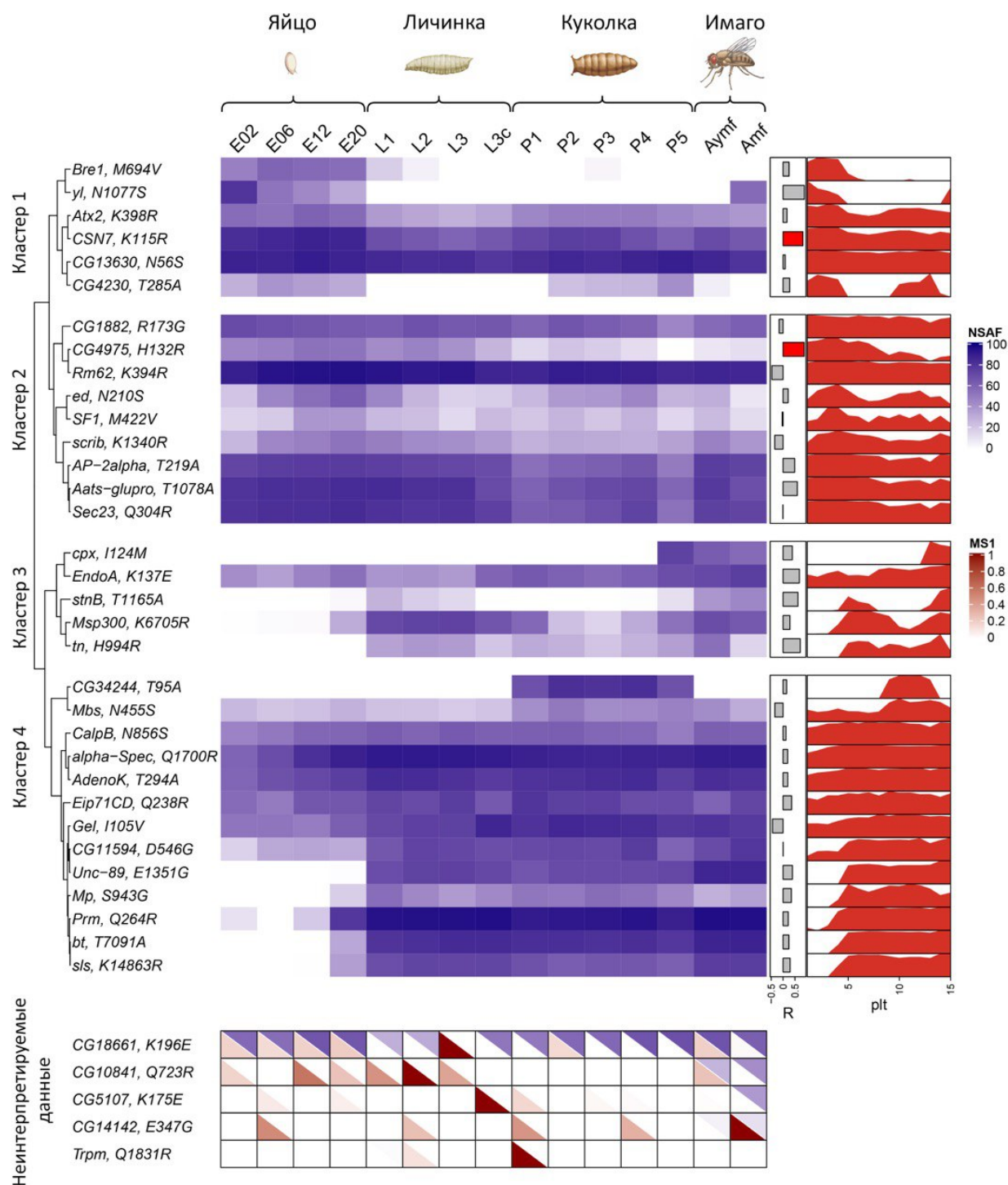


Рисунок П5. Тепловая карта относительной представленности перекодированных участков в виде триптических пептидов и соответствующих белков на различных стадиях жизненного цикла плодовой мушки. Количество белка (показано фиолетовым) выражено как процентиль NSAF [151].

Коэффициенты корреляции Пирсона между значениями MS1 и NSAF показаны справа. Для каждого коэффициента вычислен уровень значимости его отличия от нуля. Значения, выделенные серым, представляют собой коэффициенты Пирсона с поправкой на множественное сравнение $FDR < 0,05$.

Кластеризация проводилась с использованием значений NSAF. Если перекодированный редактированием участок был идентифицирован, а содержание белка невозможно было оценить, данные считались неинтерпретируемыми. Протеомные данные по стадиям жизненного цикла взяты из работы Casas-Vila и соавт. [182]. Красные графики справа также представляют изменение количества перекодированного белка на разных этапах жизненного цикла.

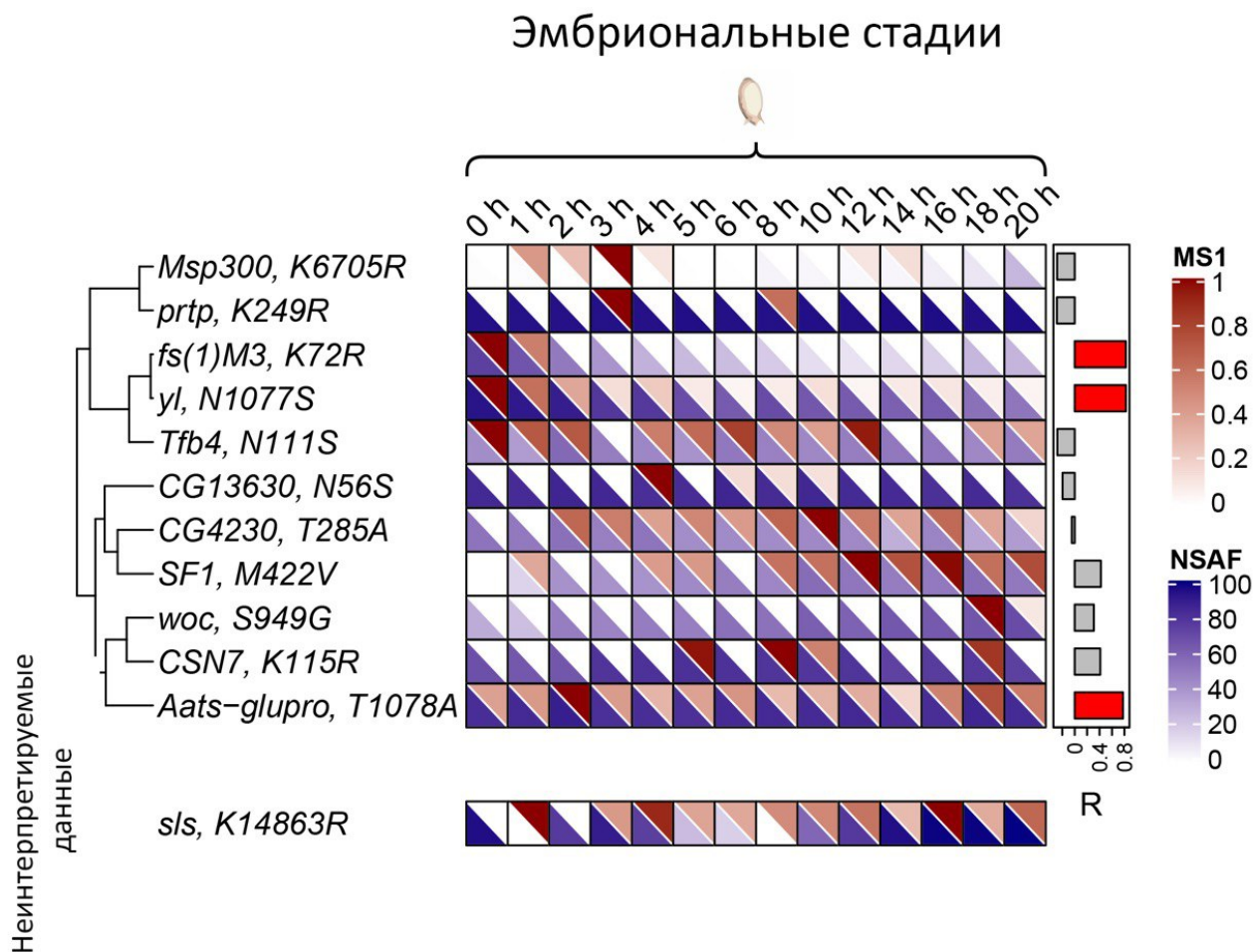


Рисунок П6. Тепловая карта относительной представленности перекодированных участков в виде триптических пептидов и соответствующих белков на различных стадиях эмбриогенеза. Уровень перекодирования (красная половина каждой ячейки диаграммы) показана в значениях интенсивности MS1 для соответствующих вариантных пептидов, нормированных индивидуально для каждой строки.

Количество белка (фиолетовая половина каждой клетки) показана как процентиль NSAF [151]. Коэффициенты корреляции Пирсона между значениями MS1 и NSAF показаны справа. Для каждого коэффициента вычислен уровень значимости его отличия от нуля. Значения, выделенные красным, представляют собой коэффициенты Пирсона с поправкой на множественное сравнение FDR < 0,05.

Кластеризация проводилась с использованием интенсивностей MS1. Если перекодированный редактированием участок был идентифицирован, а содержание белка невозможно было оценить, данные считались неинтерпретируемыми. Протеомные данные по стадиям жизненного цикла взяты из работы Casas-Vila и соавт. [182]. Пептиды SLLESSLHR и VEIELDSVR, принадлежащие парамиозину, показаны как Prm_l (left) Prm_r (right) соответственно.

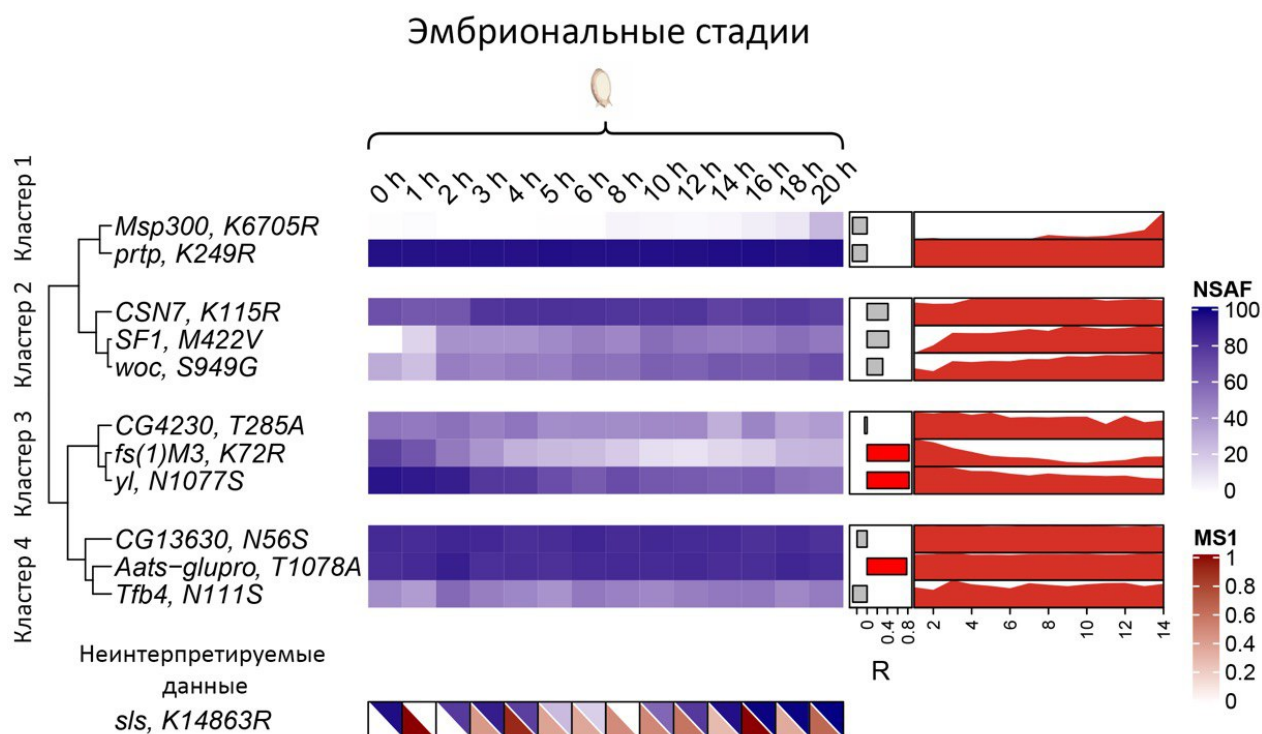


Рисунок П7. Тепловая карта относительной представленности перекодированных участков в виде триптических пептидов и соответствующих белков на различных эмбриогенеза. Количество белка (показано фиолетовым) выражено как процентиль NSAF [151]. Коэффициенты корреляции Пирсона между значениями MS1 и NSAF показаны справа. Для каждого коэффициента вычислен уровень значимости его отличия от нуля. Значения, выделенные серым, представляют собой коэффициенты Пирсона с поправкой на множественное сравнение $FDR < 0,05$. Кластеризация проводилась с использованием значений NSAF. Если перекодированный редактированием участок был идентифицирован, а содержание белка невозможно было оценить, данные считались неинтерпретируемыми. Протеомные данные по стадиям жизненного цикла взяты из работы Casas-Vila и соавт. [182]. Красные графики справа также представляют изменение количества перекодированного белка на разных этапах жизненного цикла.

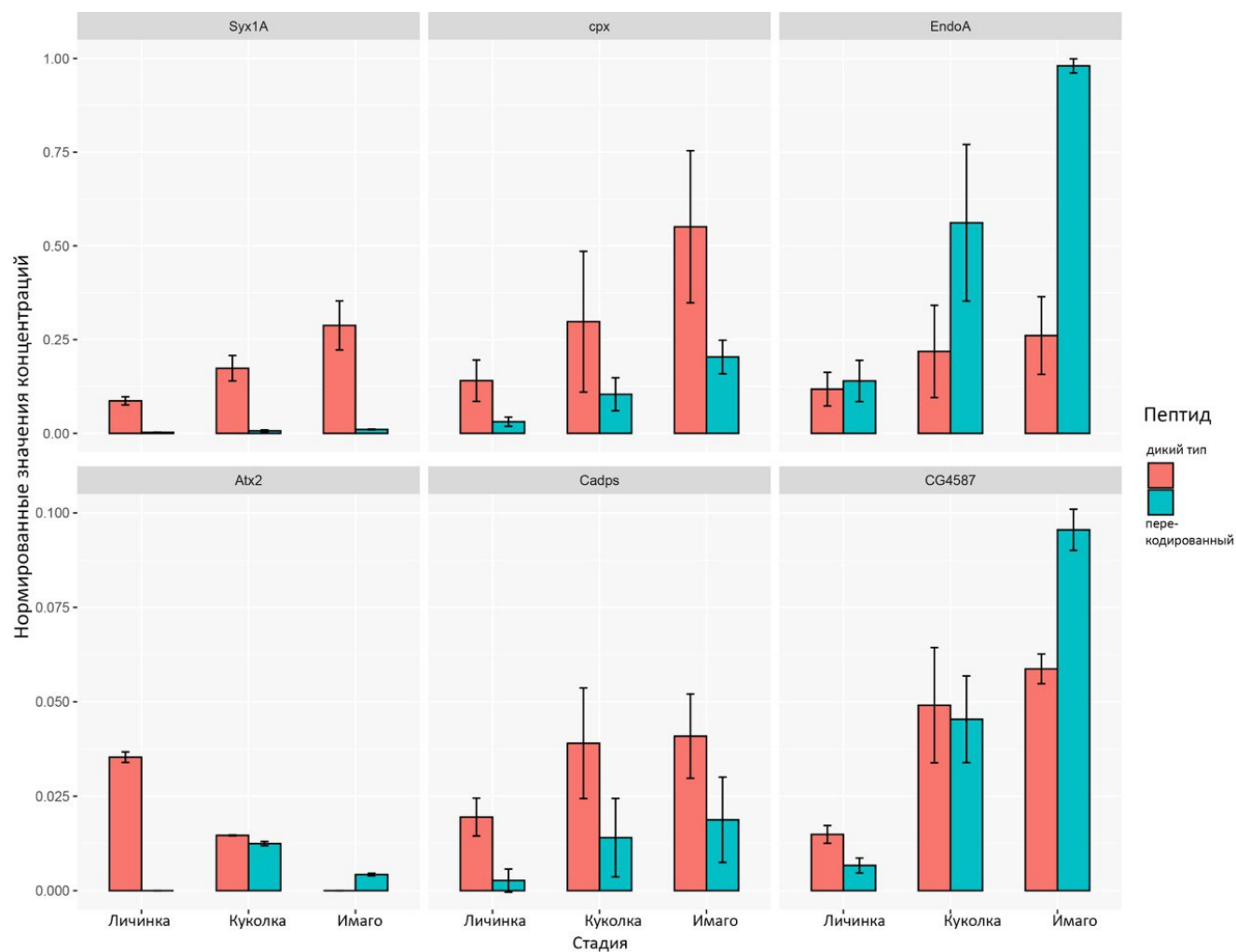


Рисунок П8. Концентрации перекодированных и кодируемых в геноме пептидов в головном мозге *D. melanogaster* на трех фазах жизненного цикла. Концентрации нормированы на общее содержание белка и получены для двух биологических повторов.