

ISSN 0026-8984

Том 57, Номер 1

Январь - Февраль 2023



# МОЛЕКУЛЯРНАЯ БИОЛОГИЯ

[www.sciencejournals.ru](http://www.sciencejournals.ru)



# СОДЕРЖАНИЕ

Том 57, номер 1, 2023

## ОБЗОРЫ

Интерлейкин-11 при патологиях нервной системы <i>М. И. Айрапетов, С. О. Ереско, П. Д. Игнатова, А. А. Лебедев, Е. Р. Бычков, П. Д. Шабанов</i>	3
Разнообразие функций хеликазы MLE в регуляции экспрессии генов у высших эукариот <i>Ю. В. Николенко, С. Г. Георгиева, Д. В. Копытова</i>	10
Регуляторный потенциал SNP-маркеров в генах, кодирующих белки систем репарации ДНК <i>Н. П. Бабушкина, А. Н. Кучер</i>	24

## ГЕНОМИКА. ТРАНСКРИПТОМИКА

Ассоциация уровня сывороточного пролактина с полиморфными вариантами генов <i>GRIN2A</i> , <i>GRM3</i> , <i>GRM7</i> у больных шизофренией, принимающих конвенциональные и атипичные антипсихотики <i>В. В. Тигунцев, В. И. Герасимова, Е. Г. Корнетова, О. Ю. Федоренко, А. Н. Корнетов, А. А. Гончарова, Е. Г. Полтавская, А. С. Бойко</i>	47
Структура и эволюция гена <i>AqE</i> у насекомых <i>Л. В. Пузакова, М. В. Пузаков</i>	56
Структура соматических мутаций в генах <i>KRAS</i> , <i>BRAF</i> , <i>PIK3CA</i> и клинические характеристики ВЭБ-ассоциированного и микросателлитно-нестабильного рака желудка <i>А. М. Данишевич, Н. И. Поспехова, А. М. Строганова, Д. А. Головина, М. П. Никулин, А. Е. Калинин, С. Э. Николаев, И. С. Стилиди, Л. Н. Любченко</i>	71
Линия трансгенных мышей APP <sup>swe</sup> /PS1 <sup>dE9</sup> /Vlg для моделирования церебрального амилоидоза при болезни Альцгеймера <i>Е. А. Лыскова, Е. В. Кузубова, А. И. Радченко, Е. А. Патраханов, К. Д. Чапров, М. В. Корокин, А. В. Дейкин, О. С. Гудырев, М. В. Покровский</i>	85
Идентификация фрагмента стигмаспецифичной экспрессии в промоторе гена хитиназы класса I сои <i>С. М. Zhao, Н. Hou, М. G. Xing, R.-G. Xue</i>	95
Необычная зависимость между экспрессией генов и отрицательным отбором у инфузорий рода <i>Euplotes</i> <i>М. А. Молдован, С. А. Гайдукова</i>	98

## МОЛЕКУЛЯРНАЯ БИОЛОГИЯ КЛЕТКИ

Активность антирестрикционного белка ArgD в отношении эндонуклеазы EcoAI <i>А. А. Кудрявцева, В. А. Алехин, М. Д. Лебедева, Е. Csefalvay, М. Weiserova, И. В. Манухов</i>	101
Метилофиопогонанон А ингибирует LPS/АТФ-индуцированный пироптоз макрофагов через путь ROS/NLRP3 <i>Н. В. Zeng, Л. Н. Zhang, D. P. Yuan, W. Wang, X. M. Su, W. X. Weng, R. Miao, J. Y. Xu, J. Long, Y. H. Song</i>	106

Инсуляторный белок Crp190 регулирует экспрессию генов дифференцировки сперматоцитов в клетках мужского зародышевого пути *Drosophila melanogaster*

*С. Е. Романов, В. В. Шлома, Д. Е. Коряков, С. Н. Белякин, П. П. Лактионов*

109

---

## **СТРУКТУРНО-ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ БИОПОЛИМЕРОВ И ИХ КОМПЛЕКСОВ**

Разделение фаз в препарате очищенного белка человека LSM4

*Н. Li, Y. Ju, W. W. Liu, Y. Y. Ma, H. Ye, N. Li*

124

---

## **БИОИНФОРМАТИКА**

Метод комплексного формирования предикторов для применения алгоритмов машинного обучения в задаче предсказания структуры и функций белка

*Ю. В. Мильчевский, В. Ю. Мильчевская, Ю. В. Кравацкий*

127

---

## **ПРОТЕОМИКА**

Влияние нокаута основных генов транссульфурации на паттерн белкового синтеза у *D. melanogaster*

*А. Ш. Заклута, В. Ю. Шилова, О. Г. Зацепина*

139

---

---

УДК 615.015.44

## ИНТЕРЛЕЙКИН-11 ПРИ ПАТОЛОГИЯХ НЕРВНОЙ СИСТЕМЫ

© 2023 г. М. И. Айрапетов<sup>a, b</sup>, С. О. Ереско<sup>b, c, \*</sup>, П. Д. Игнатова<sup>b</sup>, А. А. Лебедев<sup>b</sup>,  
Е. Р. Бычков<sup>b</sup>, П. Д. Шабанов<sup>b, d</sup>

<sup>a</sup>Санкт-Петербургский государственный педиатрический медицинский университет,  
Санкт-Петербург, 194100 Россия

<sup>b</sup>Институт экспериментальной медицины, Санкт-Петербург, 197376 Россия

<sup>c</sup>Санкт-Петербургский государственный химико-фармацевтический университет,  
Санкт-Петербург, 197101 Россия

<sup>d</sup>Военно-медицинская академия им. С.М. Кирова Министерства обороны Российской Федерации,  
Санкт-Петербург, 194044 Россия

\*e-mail: erescko.sergei@yandex.ru

Поступила в редакцию 08.05.2022 г.

После доработки 15.07.2022 г.

Принята к публикации 15.07.2022 г.

Изучение роли цитокинов при различных патологических состояниях организма — одно из актуальных направлений современной биомедицины. В 1990 году в фиброцитоподобных стромальных клетках костного мозга обнаружили интерлейкин-11 (IL-11), повышенный интерес к которому наблюдается в последнее время. IL-11 локально экспрессируется нервными клетками и участвует в патогенетических механизмах формирования ряда заболеваний нервной системы. В представленном обзоре обобщены данные, указывающие на участие IL-11 в развитии патологий головного мозга. Вероятно, этот цитокин сможет найти применение в коррекции механизмов, связанных с формированием патологических состояний нервной системы.

**Ключевые слова:** интерлейкин-11, IL-11, семейство IL-6, gp130, головной мозг, нейровоспаление, нейродегенерация

**DOI:** 10.31857/S0026898423010020, **EDN:** AYDLGW

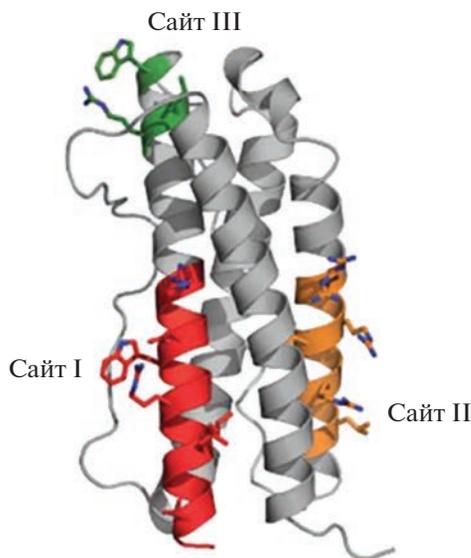
### ВВЕДЕНИЕ

Интерлейкин-11 (IL-11, он же AGIF) — белок, который относится к семейству IL-6-подобных цитокинов, впервые выделили в 1990 г. из фиброцитоподобных стромальных клеток костного мозга [1]. Первоначально показали, что IL-11 играет важную роль в процессах кроветворения, особенно в созревании мегакариоцитов [1]. В дальнейшем выявили участие этого цитокина и в других процессах.

IL-11 либо самостоятельно, либо вместе с другими цитокинами (IL-3, IL-4, IL-7, IL-12, IL-13, SCF, Flt-3 и GM-CSF) стимулирует гемопоэз мегакариоцитов, участвует в созревании В-клеток и активации Т-клеток, в пролиферации синовиальных фибробластов [1–3]. Кроме того, обнаружены негемопоэтические эффекты IL-11. Например, IL-11 обладает противовоспалительной активностью и оказывает защитное действие на эпителий слизистой оболочки желудочно-кишечного тракта и дыхательных путей [4]. Высокий уровень экспрессии IL-11 в дыхательных путях наблюдается при вирусиндуцированной бронхиальной астме,

а сверхэкспрессия приводит к ремоделированию дыхательных путей у мышей, воспалению и к появлению симптомов, сходных с симптомами астмы [5]. Передача сигналов от IL-11 имеет решающее значение для Th2-опосредованного воспалительного ответа в легких, а ингибирование передачи сигналов IL-11 снижает активацию провоспалительных путей [6]. Таким образом, регуляцию путей передачи сигналов от IL-11 можно рассматривать как потенциальную терапевтическую мишень при заболеваниях легких. Другая широко известная функция IL-11 — способность ингибировать адипогенез (отсюда второе название этого цитокина — AGIF, adipogenesis inhibitory factor) [7].

Получены многочисленные данные о том, что IL-11 вовлечен в процессы нейрогенеза [8], нейровоспаления и нейродегенерации [3, 4, 9–12]. За последние годы показано, что IL-11 вовлечен в развитие патологических состояний нервной системы. Роль IL-11 при болезни Альцгеймера, аутоиммунном энцефаломиелите, рассеянном склерозе, при ишемическом повреждении головного мозга и других заболеваниях изучена на жи-



**Рис. 1.** Структура IL-11. Отмечены сайты связывания цитокина с рецепторным комплексом. Рисунок построен по данным V. Barton и соавт. [17].

вотных моделях. В нашем обзоре обобщены данные о вовлеченности IL-11 в молекулярные нейроиммунные механизмы при патологических состояниях нервной системы.

### IL-11 И ВНУТРИКЛЕТОЧНЫЕ ПУТИ ПЕРЕДАЧИ СИГНАЛА

IL-11 — белок с молекулярной массой ~19 кДа, предшественник которого состоит из 199 аминокислотных остатков, включая лидерный пептид из 21 остатка. Ген *IL-11*, состоящий из пяти экзонов и четырех интронов, расположен на хромосоме 19 человека. IL-11 человека, крысы и мыши имеет высокую степень гомологии (более 80%), обогащен остатками пролина и не содержит остатков цистеина, что означает отсутствие дисульфидных связей во вторичной структуре. IL-11, как и другие члены семейства IL-6-подобных цитокинов, имеет форму пучка из четырех  $\alpha$ -спиралей. Во взаимодействии с рецептором участвует его С-концевая область (рис. 1) [3, 13–16].

IL-11 продуцируется клетками многих тканей [18]. В 1993 г. IL-11 обнаружили в нейронах гиппокампа, где он стимулирует пролиферацию прогениторных клеток, и показали тем самым его участие в процессах нейрогенеза [8]. Стало очевидным, что в головном мозге экспрессируются также компоненты рецепторного комплекса IL-11 (IL-11R $\alpha$ , gp130) [3, 4, 9–12, 19–21]. Экспрессия IL-11 выявлена в обонятельной луковице, миндалях, базальных ганглиях, таламусе, среднем мозге, мосте, продолговатом мозге, гиппокампе, коре головного мозга, мозжечке [18–21]. Имму-

ногистохимическим методом выявлена локализация ключевого компонента рецепторного комплекса — белка gp130 (он же IL6ST) — в головном мозге крысы [19]. При этом отмечена иммунореактивность gp130 как в глиальных, так и в нейрональных клетках. Окрашивание нейронов выявило две различные картины. На одной видно преимущественное окрашивание нейропилей в структурах конечного мозга, включая гиппокамп, кору головного мозга и хвостатое ядро. На другой окрашивание наблюдалась на цитоплазматической мембране тел нейронов, преимущественно в нижнем стволе мозга, в крупных нейронах ретикулярной формации, а также в спинномозговых и черепно-мозговых мотонейронах [19]. Электронно-микроскопический анализ показал, что оба типа иммунореактивности gp130 в основном связаны с цитоплазматической мембраной, но точно не локализуются в синаптических участках [19].

С помощью РНК-seq-анализа мРНК IL-11, IL-11R $\alpha$  и gp130 выявили также в различных структурах головного мозга. При этом в незначительных количествах мРНК IL-11 представлена во всех типах клеток нервной системы; мРНК IL-11R $\alpha$  и gp130 в наибольшем количестве обнаружена в астроцитах, микроглии и эндотелии [20, 21].

В настоящее время отсутствуют полученные с высоким разрешением структурные данные для сигнального комплекса IL-11, хотя имеются указания на сходство структур IL-11R $\alpha$  и IL-6R $\alpha$ . Методом иммунопреципитации показано, что сигнальные комплексы как IL-11, так и IL-6 имеют гексамерную структуру [17]. С помощью мутagenеза на IL-11 идентифицированы сайты I, II и III (рис. 1). Показано, что во взаимодействии IL-11 с рецепторным комплексом участвуют три сайта связывания (I, II и III) (рис. 1), расположенные аналогично сайтам I, II и III в IL-6. Сайт I позволяет IL-11 связываться с IL-11R, тогда как сайты II и III опосредуют связывание с gp130 [17]. Это стало еще одним фактом в пользу предположения о том, что активные сигнальные комплексы IL-6 и IL-11 используют в целом сходный механизм [17].

Взаимодействие IL-11 с рецепторным комплексом служит сигналом к активации JAK/SHP2-зависимого пути внутриклеточной передачи сигнала. Другой путь сигналинга начинается с фосфорилирования белков STAT1 и STAT3, что ведет к изменению скорости транскрипции генов *Socs3* и *Smad7*. Синтезируемый белок *Socs3* может, в свою очередь, ингибировать активность киназы JAK. Белок *Smad7* регулирует экспрессию антиапоптотических, пролиферативных и проангиогенных генов [3, 13–16]. При этом стоит отметить, что IL-11 немного длиннее, чем IL-6. Это указывает на различия в способе связывания и геометрии сигнального комплекса [22, 23], а внеклеточный домен IL-11R $\alpha$  имеет более гидрофоб-

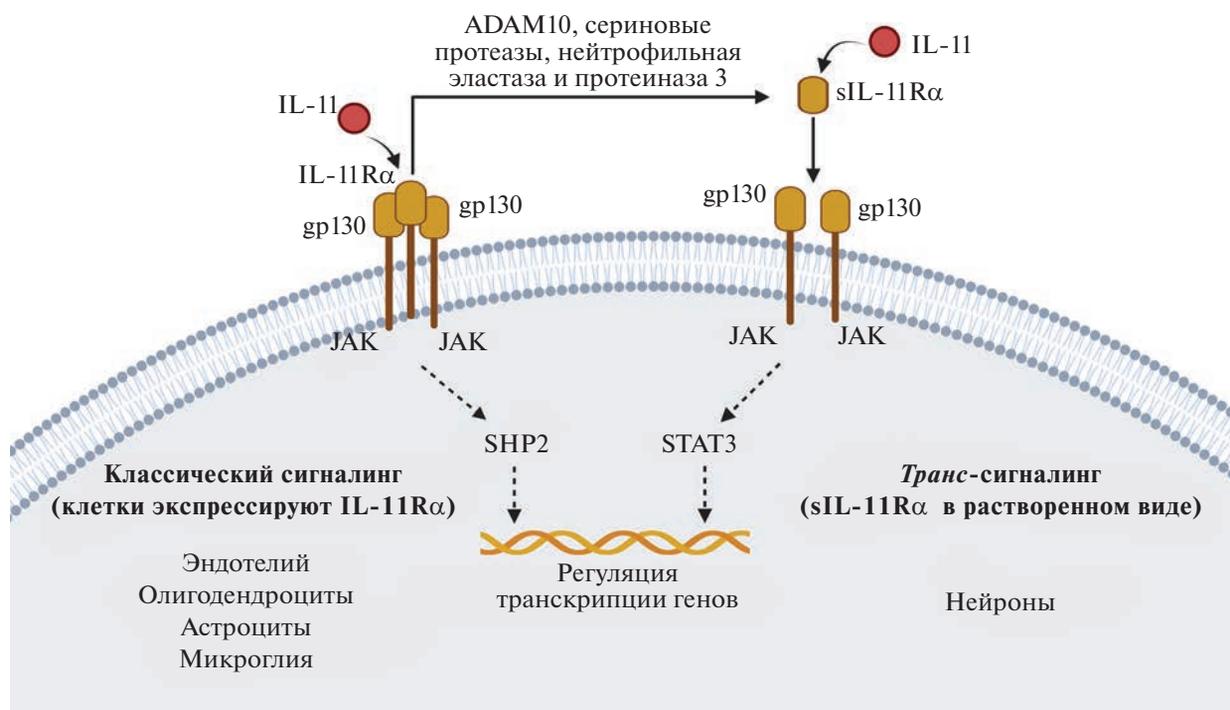


Рис. 2. Внутриклеточные пути передачи сигнала IL-11.

ный характер, чем IL-6R $\alpha$ , что предполагает существование различий в механизмах взаимодействия цитокинов IL-11 и IL-6 с рецепторным аппаратом [24].

Цитокины, в том числе IL-11, могут продуцироваться различными клетками, тогда как рецепторные субъединицы экспрессируются в более ограниченном наборе клеток, что определяет более специфичное влияние цитокинов на отдельные популяции клеток, которые могут активироваться непосредственно IL-11 [25]. Так, gp130 экспрессируется повсеместно, при этом IL-11R $\alpha$  представлен на меньшем числе клеток, хотя этот рецептор более распространен, чем IL-6R. IL-11R выявлен на многих типах клеток, включая мегакариоциты, макрофаги, Т-клетки, остеокласты, гепатоциты, эпителиальные и эндотелиальные клетки [25]. По данным РНК-seq мРНК IL-11R экспрессируется в астроцитах, микроглии и эндотелии [20, 21].

Кроме того, часть рецепторного комплекса IL-11R $\alpha$  может локализоваться не только на мембране клетки, но и существовать в растворимом виде (рис. 2) [25]. Металлопротеаза ADAM10, а также сериновые протеазы, нейтрофильная эластаза и протеиназа 3 могут отщеплять эктодомен от IL-11R $\alpha$ , а образовавшийся в ходе отщепления эктодомен представляет собой растворимую форму рецептора – sIL-11R (soluble IL-11R). sIL-11R проявляет биологическую активность и связывает IL-11 в межклеточном пространстве, увеличи-

вая число типов клеток, которые могут активироваться IL-11, но не экспрессируют собственный мембранный рецептор IL-11R $\alpha$ . Такой путь передачи сигнала может реализовываться нейронами головного мозга, в которых на высоком уровне экспрессируется белок и мРНК gp130, но на крайне низком – мРНК IL-11R $\alpha$  [20, 21].

Участие IL-11 в нескольких путях внутриклеточной передачи сигнала делает его плейотропным цитокином, вызывающим порой противоположные эффекты, однако выбор сигнального пути, как это часто бывает, определяется множеством причин и в зависимости от того или иного состояния организма может усиливаться либо замедляться тот или иной внутриклеточный путь передачи сигнала. Несмотря на то, что IL-11 играет значимую роль в нервной ткани, внутриклеточные сигнальные пути этого цитокина все еще остаются неизученными и представляют интерес для дальнейших исследований.

## ИНТЕРЛЕЙКИН-11 ПРИ ПАТОЛОГИЯХ НЕРВНОЙ СИСТЕМЫ

Показано, что IL-11 способен “корректировать” воспалительные пути в эпителиальных тканях дыхательной системы, где развиваются основные события при острой респираторной инфекции, вызываемой коронавирусом SARS-CoV-2. Рекомбинантный IL-11 используют в настоящее время для повышения числа тромбоцитов при миело-

диспластическом синдроме [26]. Подтверждена эффективность и безопасность IL-11 при лимфосаркомах и солидных опухолях [27], а также при злокачественных опухолях женских половых органов [28]. Можно рассчитывать, что дальнейшее изучение IL-11 приведет к более широкому применению этого цитокина в клинической практике.

В этом разделе обобщены данные, которые показывают роль IL-11 в механизмах, связанных с развитием патологий головного мозга (табл. 1). В одном из первых исследований, проведенном на культуре клеток нейробластомы (модель болезни Альцгеймера), показано участие IL-11 в механизмах развития этой патологии [2]. При болезни Альцгеймера наблюдается отложение агрегатов  $\beta$ -амилоида в головном мозге, а повышенная выработка  $\beta$ -амилоидного пептида считается одним из ранних событий в патогенезе этого заболевания, приводящего к прогрессирующей нейродегенерации и деменции. Показано, что  $\beta$ -амилоид активирует *L*-фосфосеринфосфатазу в нейрональных клетках (клетки нейробластомы крысы B104), а IL-11 дозозависимым образом (5, 50, 100 нг/мл) ингибирует вызываемую при этом нейротоксичность. Предположили, что *L*-фосфосеринфосфатаза может участвовать в изменении клеточного метаболизма при болезни Альцгеймера путем усиления нейротоксичности, тогда как IL-11 может действовать через свою рецепторную систему как нейропротектор [2].

Обнаружено повышение содержания IL-11 в астроцитах и IL-11R на олигодендроцитах при рассеянном склерозе. В ходе исследований, проведенных на культурах астроцитов и олигодендроцитов, сделан вывод, что путь IL-11/IL-11R может участвовать в защите олигодендроцитов. Так, добавление рекомбинантного IL-11 (10 нг/мл) к культуре клеток спинного мозга плода человека привело к увеличению числа олигодендроцитов, повышенному ветвлению олигодендроглии, снижению апоптотической гибели клеток; кроме того, IL-11 потенцировал образование миелина [29]. В последующих исследованиях наблюдали развитие выраженного нейровоспаления в тканях спинного мозга у мышей с нокаутом гена *IL-11R $\alpha$*  (*IL-11R $\alpha$ <sup>-/-</sup>* и *IL-11R $\alpha$ <sup>+/-</sup>*), которое характеризовалось инфильтрацией макрофагами с последующей демиелинизацией и снижением числа нервных клеток. Это данные свидетельствуют о значимости IL-11R $\alpha$  в регуляции нейровоспаления [10].

Влияние экзогенного IL-11 оценивали на культуре прогениторных олигодендроцитов (ОП) крысы. Применение IL-11 повышало выживаемость клеток, уменьшая число клеток, гибнущих по пути апоптоза, и потенцируя процесс деления клеток. Кроме того, экзогенный IL-11 способен

снижать активацию CD4 лимфоцитов, воздействуя на клетки CD11c<sup>++</sup>. IL-11 может регулировать продукцию эффекторных цитокинов, воздействуя на клетки CD11c<sup>++</sup> [10]. Изучено также влияние рекомбинантного IL-11 при аутоиммунном энцефаломиелите, который моделировали у мышей. После появления признаков заболевания животным вводили рекомбинантный IL-11 в дозе 25 или 50 мкг/кг/сут. Оказалось, что IL-11 в дозе 50 мкг/кг/сут снижал тяжесть заболевания: наблюдалось снижение степени демиелинизации и потери олигодендроцитов, а также воспаления и числа CD3 лимфоцитов. Более низкие дозы IL-11 не оказывали значимого эффекта [10].

Повышение концентрации IL-11 выявлено в ликворе и сыворотке пациентов с рассеянным склерозом. Сообщается, что основным источником IL-11 в крови являются клетки CD4<sup>+</sup>, а содержание таких клеток (IL-11+CD4<sup>+</sup>) повышено у пациентов с рассеянным склерозом. Предполагается, что повышенный уровень IL-11 может индуцировать дифференцировку и пролиферацию клеток Th17, так как IL-11 индуцированная дифференцировка Th17 ингибировалась антителом к IL-11. При этом антитело к IL-6 не ингибировало дифференцировку клеток Th17, т.е. IL-11-индуцированная дифференцировка Th17 не зависит от IL-6. Участие клеток Th17 в развитии рассеянного склероза подтверждается результатами нескольких исследований. Клетки Th17 человека служат источником эффекторных цитокинов IL-17A, IL-17F, IL-21, IL-22 и TNF- $\alpha$ , которые опосредуют продукцию воспалительных цитокинов, хемокинов и металлопротеиназ, а также способствуют привлечению гранулоцитов и лимфоцитов к местам воспаления. Миграцию клеток в ЦНС обеспечивает рецептор хемокинов CCR6. Высокий уровень экспрессии CCR6 на IL-11+ CD4<sup>+</sup> клетках может опосредовать их миграцию в ЦНС, что объясняет повышение уровня IL-11 в спинномозговой жидкости пациентов [30].

Появление очагов демиелинизации в белом веществе больных рассеянным склерозом считается результатом скоординированного аутоиммунного процесса, который включает активацию клеток микроглии в нервной ткани (местные резидентные макрофаги). На купризоневой модели рассеянного склероза у мышей, показано, что сверхэкспрессия IL-11, вызванная локальной доставкой в очаг повреждения гена *IL-11* с помощью вирусного вектора, снижала степень демиелинизации с одновременным ускорением процесса ремиелинизации. Поскольку известно, что микроглия участвует в фагоцитозе миелина, полученные результаты могут указывать на способность IL-11 вносить коррективы в этот процесс. Однако механизм того, каким образом микроглия участвует в эффектах, вызванных IL-11, еще предстоит выяснить. В эксперименте *in vitro* на клеточ-

Таблица 1. Участие IL-11 в патологиях нервной системы

Патологическое состояние	Объект	Функции IL-11	Источник
Болезнь Альцгеймера	Культура нейробластомы крысы B104	Выполняет нейропротекторную роль, ингибируя L-фосфосеринфосфатазу	[2]
Рассеянный склероз	Пациенты	Повышение IL-11 в астроцитах, IL-11R на олигодендроцитах, индуцирует дифференцировку и пролиферацию Th17, независимо от IL-6	[11, 29, 30]
	Первичная культура астроцитов и олигодендроцитов	Рекомбинантный IL-11 человека повышает число олигодендроцитов, снижает апоптоз, активирует ремиелинизацию	
	Мышь, культура клеток микроглии BV2	Ограничивает пролиферацию и способность микроглии к фагоцитозу миелина, замедляет демиелинизацию, способствует ускорению процесса ремиелинизации	
Аутоиммунный энцефаломиелит	Мышь, культура клеток OPCs	Замедляет нейродегенерацию, снижая потери олигодендроцитов и нейронов, снижает нейровоспаление и уровень цитотоксических лейкоцитов	[10]
Ишемия	Мышь	Снижает активацию микроглии, противовоспалительные цитокины (IL-1 $\beta$ , IL-6, TNF $\beta$ ) и окислительный стресс в мозге, повышает уровень противовоспалительного цитокина TGF- $\beta$ 1	[12]
Медуллобластома	Культура клеток медуллобластомы	Повышает химиорезистентность	[33]

ной линии микроглии BV2 показано [11], что добавление рекомбинантного IL-11 (1–30 нг/мл) дозозависимо уменьшало степень фагоцитоза миелина. Кроме того, использование IL-11 привело к увеличению толщины миелинового слоя, что указывает на способность IL-11 усиливать процесс ремиелинизации, а не только замедлять демиелинизацию, однако механизм этого еще предстоит изучить [11].

Демиелинизация в купризоновой модели рассеянного склероза также может зависеть от секреции микроглией провоспалительных цитокинов, токсичных для олигодендроцитов и нервных клеток. Опубликованы данные о способности IL-11 ингибировать синтез TNF- $\alpha$ , IL-1 $\beta$ , IL-12, IL-6 и продукцию NO активированными макрофагами [31, 32], однако уровень провоспалительных медиаторов в этом эксперименте не указан.

Изучено также участие IL-11 в ишемическом повреждении мозга, которое моделировали на мышцах с помощью окклюзии средней мозговой артерии. Показано снижение экспрессии мРНК и белка IL-11 в первые 24 ч ишемии с дальнейшим постепенным его увеличением. Внутривенное введение рекомбинантного IL-11 в эксперименте выявило противовоспалительный эффект этого цитокина в дозах 20 и 40 мкг/кг, который приво-

дил к снижению маркеров активации астроцитов и микроглии, а также уровня мРНК провоспалительных цитокинов (IL-1 $\beta$ , IL-6, TNF- $\beta$ ). При этом отмечено повышение уровня мРНК противовоспалительного цитокина (TGF- $\beta$ 1), уровня супероксид-дисмутазы и снижение уровня малонового диальдегида, что свидетельствует о снижении окислительного стресса в тканях мозга [12].

Стоит обратить внимание еще на один факт. В одном из недавних исследований показано, что цитокины семейства IL-6 могут опосредовать повышение химиорезистентности клеток медуллобластомы. Культуры клеток медуллобластомы, кондиционированные *in vitro* цитокинами IL-6, OSM, LIF или IL-11, проявляют повышенную активность JAK1/STAT3-сигнализации, при этом развивается резистентность к ряду препаратов. Применение ингибиторов gp130 или ингибиторов JAK-киназ позволило эффективно преодолеть резистентность экспрессирующих gp130 клеток медуллобластомы к винкристину [33]. Полученные данные могут указывать на ограниченную применимость рекомбинантного IL-11, однако необходимы дальнейшие исследования *in vivo*, а также изучение опосредованного влияния цитокинов семейства IL-6 на другие виды опухолевых клеточных линий.

## ЗАКЛЮЧЕНИЕ

IL-11 впервые выделили из фиброцитоподобных стромальных клеток костного мозга еще в 1990 г. и тогда же его нашли в головном мозге, однако о функциях этого цитокина в нервной системе известно совсем немного. IL-11 обладает нейрорепрессорными и противовоспалительными свойствами, замедляет скорость развития нейровоспаления и апоптотической гибели нервных клеток, которая наблюдается при различных патологических состояниях нервной системы [34, 35].

IL-11 активирует процессы нейрогенеза (ускоряет скорость деления прогениторных нейронов и олигодендроцитов), стимулирует процесс ремиелинизации аксонов при рассеянном склерозе. Необходимы дополнительные исследования, которые расширят наши знания о роли этого цитокина при патологии центральной нервной системы.

Исследование выполнено за счет средств ИЭМ в рамках государственного задания по теме “Фармакологический анализ действия нейротропных средств, изучение внутриклеточных мишеней и создание систем направленной доставки”, шифр 0557-2019-0004.

Настоящая статья не содержит каких-либо исследований с участием людей или животных в качестве объектов исследования.

Авторы заявляют об отсутствии конфликта интересов.

## СПИСОК ЛИТЕРАТУРЫ

- Paul S.R., Bennett F., Calvetti J.A., Kelleher K., Wood C.R., O'Hara R.M., Jr., Leary A.C., Sibley B., Clark S.C., Williams D.A., Yang Y.C. (1990) Molecular cloning of a cDNA encoding interleukin 11, a stromal cell derived lymphopoietic and hematopoietic cytokine. *Proc. Natl. Acad. Sci. USA*. **87**, 7512–7516.
- Heese K., Nagai Y., Sawada T. (2000) Induction of rat L-phosphoserine phosphatase by amyloid-beta (1-42) is inhibited by interleukin-11. *Neurosci. Lett.* **288**(1), 37–40.
- Гук К.Д., Купраш Д.В. (2011) Интерлейкин-11, член семейства IL-6-подобных цитокинов. *Молекуляр. биология*. **45**(1), 44–55.
- Wilde M.I., Faulds D. (1998) Oprelvekin: a review of its pharmacology and therapeutic potential in chemotherapy-induced thrombocytopenia. *BioDrugs*. **10**(2), 159–171.
- Einarsson O., Geba G.P., Zhu Z., Landry M., Elias J.A. (1996) Interleukin-11: stimulation *in vivo* and *in vitro* by respiratory viruses and induction of airways hyperresponsiveness. *J. Clin. Investig.* **97**(4), 915–924.
- Tang W., Geba G.P., Zheng T., Ray P., Homer R.J., Kuhn C. (1996) Targeted expression of IL-11 in the murine airway causes lymphocytic inflammation, bronchial remodeling, and airways obstruction. *J. Clin. Investig.* **98**(12), 2845–2853.
- Keller D.C., Du X.X., Srouf E.F., Hoffman R., Williams D.A. (1993) Interleukin-11 inhibits adipogenesis and stimulates myelopoiesis in human long-term marrow cultures. *Blood*. **82**(5), 1428–1435.
- Mehler M.F., Rozental R., Dougherty M., Spray D.C., Kessler J.A. (1993) Cytokine regulation of neuronal differentiation of hippocampal progenitor cells. *Nature*. **362**(6415), 62–65.
- Leng S.X., Elias J.A. (1997) Interleukin-11. *Int. J. Biochem. Cell. Biol.* **29**, 1059–1062.
- Gurfein B.T., Zhang Y., López C.B. (2009) IL-11 regulates autoimmune demyelination. *J. Immunol.* **183**(7), 4229–4240.
- Maheshwari A., Janssens K., Bogie J., Van Den Haute C., Struys T., Lambrechts I., Baekelandt V., Stinissen P., Hendriks J.J., Slaets H., Hellings N. (2013) Local overexpression of interleukin-11 in the central nervous system limits demyelination and enhances remyelination. *Mediators Inflamm.* **2013**, 685317.
- Zhang B., Zhang H.-X., Shi S.-T., Bai Y.-L., Zhe X., Zhang S.-J., Li Y.-J. (2019) Interleukin-11 treatment protected against cerebral ischemia/reperfusion injury. *Biomed. Pharmacother.* **115**, 108816.
- Горшкова Е.А., Недоспасов С.А., Шилов Е.С. (2016) Эволюционная пластичность цитокинов семейства IL-6. *Молекуляр. биология*. **50**(6), 1039–1048.
- McKinley D., Wu Q., Yang-Feng T., Yang Y.C. (1992) Genomic sequence and chromosomal location of human interleukin-11 gene (*IL11*). *Genomics*. **13**(3), 814–819.
- Yang Y.C., Yin T. (1992) Interleukin-11 and its receptor. *Biofactors*. **4**(1), 15–21.
- Stephen F.A., Thomas L.M., Alejandro A.S., Jinghui Z., Zheng Z., Webb M., David J.L. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucl. Acids Res.* **25**, 3389–3402.
- Barton V., Hall M., Hudson K.R., Heath J.K. (2000) Interleukin-11 signals through the formation of a hexameric receptor complex. *J. Biol. Chem.* **275**(46), 36197–36203.
- Knut & Alice Wallenberg Foundation. *IL-11. The Human Protein Atlas*.
- Watanabe D., Yoshimura R., Khalil M., Yoshida K., Kishimoto T., Taga T., Kiyama H. (1996) Characteristic localization of gp130 (the signal-transducing receptor component used in common for IL-6/IL-11/CNTF/LIF/OSM) in the rat brain. *Eur. J. Neurosci.* **8**(8), 1630–1640.
- Zhang Y., Chen K., Sloan S.A., Bennett M.L., Scholze A.R., O'Keeffe S., Phatnani H.P., Guarnieri P., Caneda C., Ruderisch N., Deng S., Liddelov S.A., Zhang C., Daneman R., Maniatis T., Barres B.A., Wu J.Q. (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**(36), 11929–11947.
- Zhang Y., Sloan S.A., Clarke L.E., Caneda C., Plaza C.A., Blumenthal P.D., Vogel H., Steinberg G.K., Edwards M.S., Li G., Duncan J.A., Cheshier S.H., Shuer L.M., Chang E.F., Grant G.A., Gephart M.G., Barres B.A. (2016) Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron*. **89**(1), 37–53.

22. Barton V., Hudson K.R., Heath J.K. (1999) Identification of three distinct receptor binding sites of murine interleukin-11. *J. Biol. Chem.* **274**(9), 5755–5761.
23. Tacke N., Dahmen H., Boisteau O., Minvielle S., Jacques Y., Grötzinger J. (1999) Definition of receptor binding sites on human interleukin-11 by molecular modeling-guided mutagenesis. *Eur. J. Biochem.* **265**(2), 645–655.
24. Metcalfe R.D., Aizel K., Zlatić C.O., Nguyen P.M., Morton C.J., Lio. (2020) The structure of the extracellular domains of human interleukin 11  $\alpha$ -receptor reveals mechanisms of cytokine engagement. *J. Biol. Chem.* **295**(24), 8285–8301.
25. Lokau J., Agthe M., Flynn C.M., Garbers C. (2017) Proteolytic control of interleukin-11 and interleukin-6 biology. *Biochim. Biophys. Acta Mol. Cell. Res.* **1864**(11 Pt B), 2105–2117.
26. Montero A.J., Estrov Z., Freireich E.J., Khouri I.F., Koller C.A., Kurzrock R. (2006) Phase II study of low dose interleukin 11 in patients with myelodysplastic syndrome. *Leuk. Lymphoma.* **47**, 2049–2054.
27. Hatae M., Ariyoshi Y., Fukuoka M., Furuse K., Noda K., Hirabayashi K., Yakushiji M., Ogawa M., Takaku F. (2005) An early phase II clinical study of YM 294 (rhIL 11) in patients with solid tumors and malignant lymphoma. *Gan To Kagaku Ryoho.* **32**, 489–496.
28. Hatae M., Noda K., Yamamoto K., Ozaki M., Hirabayashi K., Nishida T., Ohashi Y., Ogawa M., Takaku F. (2005) A clinical study of YM 294 (rhIL 11) in patients with gynecologic cancer. *Gan To Kagaku Ryoho.* **32**, 479–487.
29. Zhang Y., Taveggia C., Melendez-Vasquez C., Einheber S., Raine C.S., Salzer J.L., Brosnan C.F., John G.R. (2006) Interleukin-11 potentiates oligodendrocyte survival and maturation, and myelin formation. *J. Neurosci.* **26**(47), 12174–12185.
30. Zhang X., Tao Y., Chopra M., Dujmović-Basuroski I., Jin J., Tang Y., Drulović J., Marković-Pleše S. (2015) IL-11 induces Th17 cell responses in patients with early relapsing-remitting multiple sclerosis. *J. Immunol.* **194**(11), 5139–5149.
31. Trepicchio W.L., Wang L., Bozza M., Dorner A.J. (1997) IL-11 regulates macrophage effector function through the inhibition of nuclear factor- $\kappa$ B. *J. Immunol.* **159**(11), 5661–5670.
32. Leng S.X., Elias J.A. (1997) Interleukin-11 inhibits macrophage interleukin-12 production. *J. Immunol.* **159**(5), 2161–2168.
33. Sreenivasan L., Li L.V., Leclair P., Lim C.J. (2022) Targeting the gp130/STAT3 axis attenuates tumor microenvironment mediated chemoresistance in group 3 medulloblastoma cells. *Cells.* **11**, 381.
34. Airapetov M., Eresko S., Lebedev A., Bychkov E., Shabanov P. (2021) The role of Toll-like receptors in neurobiology of alcoholism. *BioSci. Trends.* **15**(2), 74–82.
35. Peterson P.K., Toborek M. (2014) *Neuroinflammation Neurodegeneration*. New York: Springer.

## Interleukin-11 in Pathology of the Nervous System

**M. I. Airapetov<sup>1, 2</sup>, S. O. Eresko<sup>2, 3, \*</sup>, P. D. Ignatova<sup>2</sup>, A. A. Lebedev<sup>2</sup>,  
E. R. Bychkov<sup>2</sup>, and P. D. Shabanov<sup>2, 4</sup>**

<sup>1</sup>*Department of Pharmacology, St. Petersburg State Pediatric Medical University, St. Petersburg, 194100 Russia*

<sup>2</sup>*Department of Neuropharmacology, Institute of Experimental Medicine, St. Petersburg, 197376 Russia*

<sup>3</sup>*Research and Training Center of Molecular and Cellular Technologies,*

*St. Petersburg State Chemical Pharmaceutical University, St. Petersburg, 197101 Russia*

<sup>4</sup>*Department of Pharmacology, Kirov Military Medical Academy, St. Petersburg, 194044 Russia*

*\*e-mail: eresko.sergei@yandex.ru*

The study of the role of cytokines in various pathological conditions of the body is one of the topical areas of modern biomedicine. Interleukin 11 (IL-11) was discovered in 1990 in fibrocyte-like bone marrow stromal cells, but there has been an increased interest in the cytokine in recent years. The cytokine plays a significant role in the central nervous system; local expression of this cytokine by nerve cells has been shown. Studies show the involvement of IL-11 in the mechanisms of development of a number of pathologies of the nervous system. The review summarizes information that shows the involvement of IL-11 in the mechanisms of development of brain pathologies. Probably, in the near future, this cytokine will be able to find clinical application in the correction of mechanisms that are involved in the formation of pathological conditions of the nervous system.

**Keywords:** IL-11, IL-6 family, gp130, brain, neuroinflammation, neurodegeneration

УДК 577.216.4

## РАЗНООБРАЗИЕ ФУНКЦИЙ ХЕЛИКАЗЫ MLE В РЕГУЛЯЦИИ ЭКСПРЕССИИ ГЕНОВ У ВЫСШИХ ЭУКАРИОТ

© 2023 г. Ю. В. Николенко<sup>а</sup>\*, С. Г. Георгиева<sup>а</sup>, Д. В. Копытова<sup>а</sup><sup>а</sup>Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991 Россия

\*e-mail: julia.v.nikolenko@gmail.com

Поступила в редакцию 20.06.2022 г.

После доработки 08.08.2022 г.

Принята к публикации 11.08.2022 г.

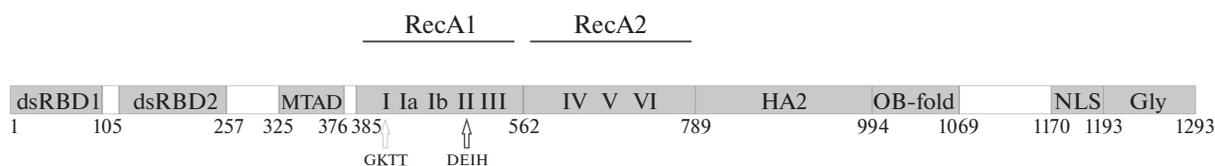
Белок MLE (Maleless) *Drosophila melanogaster* – консервативная хеликаза, вовлеченная в широкий спектр процессов регуляции экспрессии генов. У многих высших эукариот, включая человека, обнаружен ортолог MLE, получивший название DNX9. DNX9 участвует в таких процессах, как поддержание стабильности генома, репликация, транскрипция, сплайсинг, редактирование и транспорт клеточных и вирусных РНК, регуляция трансляции. Часть из этих функций изучена довольно подробно, но большая часть остается неохарактеризованной. Изучение функций ортолога MLE у млекопитающих *in vivo* ограничено тем, что потеря этого белка приводит к эмбриональной летальности. Хеликазу MLE у *D. melanogaster* первоначально обнаружили и долгое время изучали как участника процесса дозовой компенсации. Однако в последние годы появились данные о том, что в клетках *D. melanogaster* и в клетках млекопитающих хеликаза MLE участвует в одних и тех же процессах, при этом многие функции этого белка эволюционно консервативны. Кроме того, в экспериментах на *D. melanogaster* обнаружены такие новые важные функции MLE, как участие в гормон-зависимой регуляции транскрипции, взаимодействие с транскрипционным комплексом SAGA и другими кофакторами транскрипции и комплексами ремоделирования хроматина. В отличие от млекопитающих, мутации *mle* не приводят к гибели *D. melanogaster* на эмбриональной стадии, что позволяет изучать функции этого белка *in vivo* на протяжении всего онтогенеза у самок и до стадии куколки у самцов. Ортолог MLE у человека рассматривается как потенциальная мишень для противоопухолевой и противовирусной терапии. Поэтому изучение функций этой хеликазы на модельном организме *D. melanogaster* представляется важным и перспективным как в фундаментальном, так и в прикладном аспектах. В настоящем обзоре рассмотрено систематическое положение, доменное строение, а также консервативные и специфические функции хеликазы MLE *D. melanogaster*.

**Ключевые слова:** MLE, DNX9, РНК-хеликаза, *D. melanogaster*, экспрессия генов**DOI:** 10.31857/S0026898423010123, **EDN:** AYAXGF

### СИСТЕМАТИЧЕСКОЕ ПОЛОЖЕНИЕ MLE

Белок MLE относится к хеликазам, обширному классу ферментов, представленному во всех клеточных формах жизни и у многих вирусов [1–3]. На основании анализа аминокислотной последовательности, структуры и функций выделяют шесть суперсемейств хеликаз (SF1–6), которые, в свою очередь, делятся на семейства [4]. MLE входит в состав семейства хеликаз с DExH-боксом, входящего в суперсемейство SF2. Представители суперсемейства SF2 содержат консервативное каталитическое “ядро”, состоящее из двух глобулярных ResA-подобных доменов, сходных с доменами бактериальной рекомбиназы А, каждый из которых состоит из пяти β-структур, окруженных пятью α-спиралями. В последовательности ResA-подобных доменов выделяют от 7 до 12 наиболее консервативных аминокислот-

ных мотивов [1]. Еще совсем недавно семейство хеликаз с DExH-боксом объединяли в одну группу с DEAD-бокс-хеликазами и называли их хеликазами с DExD/H-боксом, однако со временем обнаружили принципиальные различия как в аминокислотной последовательности, так и в биохимических свойствах этих белков, что привело к выделению двух отдельных семейств в составе SF2: DEAD-бокс (альтернативное название DExD, или DDX) и DEAH/RHA (DExH, DNX). Свои названия эти два семейства получили по аминокислотной последовательности консервативного мотива II, представленного у этих белков последовательностями Asp-Glu-x-Asp и Asp-Glu-x-His, где x – остаток любой аминокислоты. Мотив II – главный сайт связывания и гидролиза NTP, который взаимодействует с их β- и γ-фосфатными группами, координируя ион Mg<sup>2+</sup>.



**Рис. 1.** Доменная структура белка MLE. Районы, соответствующие определенным доменам, обозначены серым цветом. Линкерные участки обозначены белым цветом. Представлены два домена связывания дцРНК – dsRBD1 и dsRBD2, минимальный домен трансактивации MTAD, каталитическое ядро, состоящее из доменов RecA1 и RecA2 и хеликаза-ассоциированного домена HA2, область связывания олигосахарида/олигонуклеотида OB-fold, сигнал ядерной локализации NLS, глицин-богатый С-концевой домен Gly. Консервативные мотивы в составе каталитического ядра обозначены римскими цифрами, мотив DExH представлен последовательностью DEIH. Подчеркнуты консервативные аминокислотные остатки в мотиве I: K413 и T415, замены которых приводят к мутациям, *mle<sup>GET</sup>* и *mle<sup>hapis</sup>* соответственно.

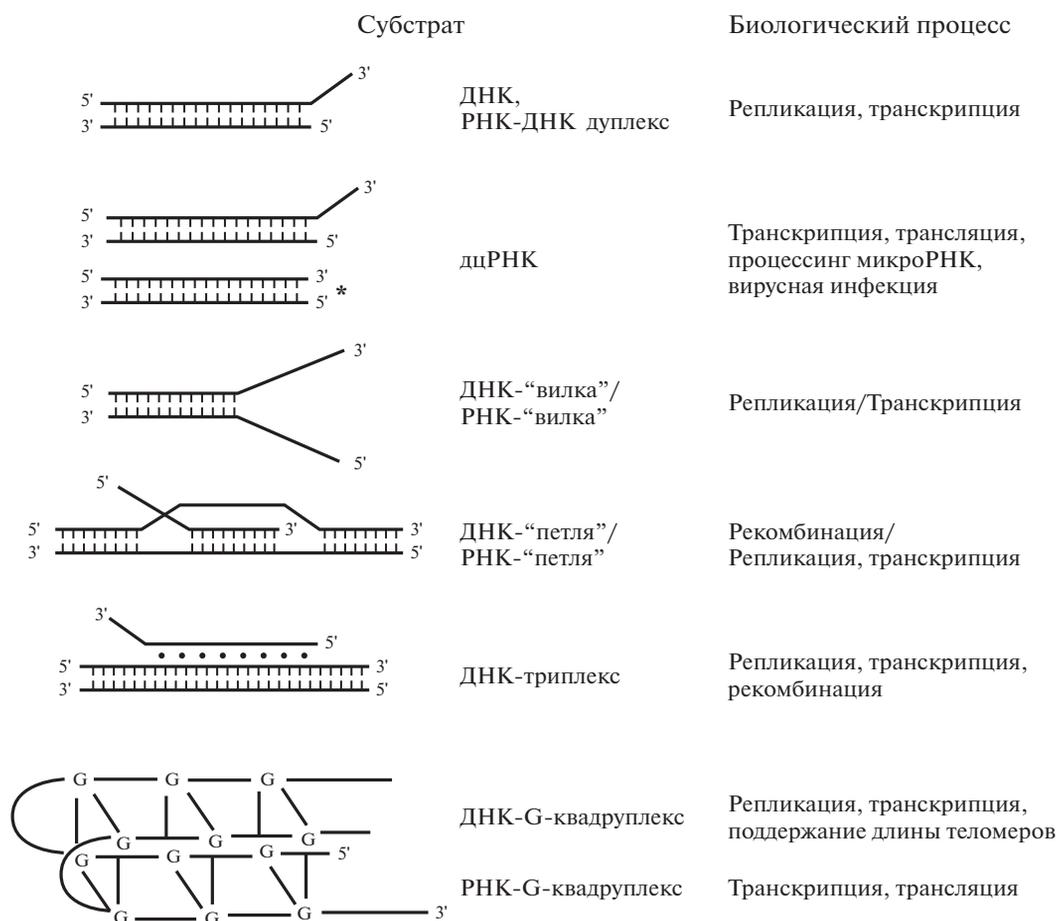
Несмотря на структурное сходство, выявлены и существенные различия в ключевых аминокислотных остатках консервативных мотивов белков семейств DDX и DHX. Как следствие, хеликазы DHX могут гидролизовать любые нуклеотиды, в то время как хеликазы DDX – только АТФ. Это свойство хеликаз связано с дополнительным консервативным Q-мотивом, расположенным непосредственно перед мотивом I в последовательности хеликаз DDX. Q-мотив образует контакт с остатком аденина, что и обеспечивает специфичность узнавания АТФ. У хеликаз DHX Q-мотив отсутствует. Другое отличие хеликаз DDX и DHX состоит в том, что хеликазы DDX осуществляют непроецессивное локальное расплетание молекулы нуклеиновой кислоты без транслокации. Хеликазы же семейства DHX с умеренной процессивностью расплетают нуклеиновые кислоты, перемещаясь вдоль молекулы в 3'→5'-направлении (есть исключение), часто этим хеликазам для связывания с субстратом нужен “свободный” одноцепочечный 3'-конец.

### СТРОЕНИЕ MLE, БИОХИМИЧЕСКАЯ АКТИВНОСТЬ *in vitro*

Хеликаза MLE консервативна у высших эукариот и все ее ортологи имеют одинаковое строение. Ортологи MLE найдены у человека и других млекопитающих (*Bos taurus*, *Mus musculus*), а также у других высших эукариот: *Caenorhabditis elegans*, *Arabidopsis thaliana* [5–7], но не обнаружены у птиц [8, 9]. В настоящее время все эти белки-ортологи имеют общее название DHX9. Далее мы будем использовать название MLE для белка *Drosophila melanogaster*, а DHX9 – для обозначения его ортологов у других организмов. Обнаружен высокий уровень гомологии DHX9 у разных видов млекопитающих составляет 90% и более. У белков *D. melanogaster* и человека идентичны 50% аминокислотных остатков, а 85% – подобны. Поликлональные антитела к MLE *D. melanogaster* распознают DHX9 в экстракте клеток HeLa [5].

Белки DHX9 человека и белки *C. elegans* и *A. thaliana* имеют 42 и 27% идентичных аминокислотных остатков соответственно [10]. Все эти белки имеют сходные размеры (1270 аминокислотных остатков у человека и 1293 – у *D. melanogaster*) и одинаковую доменную структуру [10–13].

Строение белка MLE представлено на рис. 1. В центральной части молекулы расположено характерное для хеликаз DHX каталитическое ядро, содержащее восемь консервативных мотивов, организованных в два хеликазных RecA-подобных домена. В первом домене располагаются мотивы I–III, а во втором – IV–VI. Связывание NTP и нуклеиновой кислоты происходит в углублении между доменами. Остатки глутаминовой и аспарагиновой кислоты в мотиве II, представленном последовательностью DEIH, как сказано выше, координируют катион Mg<sup>2+</sup> и взаимодействуют с β- и γ-фосфатными группами NTP. Также во взаимодействии с NTP и его гидролизе участвуют мотивы I и VI. Мутация *mle<sup>GET</sup>* (замена одного аминокислотного остатка K413E в мотиве I) приводит к полной потере АТФазной активности белка. Мотив III отвечает за хеликазную активность, т.е. за связь гидролиза NTP с расплетанием нуклеиновой кислоты, а мотивы Ia, Ib, IV и V – за связывание нуклеиновой кислоты. Сразу за RecA-подобными доменами локализован ассоциированный с хеликазой домен HA2. На N-конце белка находятся два домена связывания дцРНК (dsRBD1, dsRBD2), а затем небольшой (примерно 50 аминокислотных остатков) домен трансактивации MTAD [11]. За каталитическим ядром располагается район связывания олигосахарида/олигонуклеотида (OB-fold), затем следует сигнал ядерной локализации и глицин-богатая С-концевая область (Gly). Особенностью С-концевой области MLE является 9-кратный несовершенный повтор последовательности GGGYGNN [14], где тирозин занимает каждое седьмое положение в пределах повтора. Такое псевдогептадное расположение тирозина может указывать на формирование амфипатической спирали с гидрофобной



**Рис. 2.** Схематическое изображение образованных ДНК и РНК структур, которые способны расплетать *in vitro* MLE и ее ортолог у млекопитающих. Указаны биологические процессы, в которых образуются эти структуры. Звездочкой отмечены дуплексы РНК с “тупыми” концами, способность расплетать которые показана только для MLE *D. melanogaster*. Схема с модификациями, введенными нами, цитирована по [10, лицензия Creative Commons Attribution License].

поверхностью, которая может участвовать в межмолекулярных взаимодействиях [15]. Домены dsRBD2, HA2 и OB-fold необходимы для хеликазной активности MLE *in vitro*, в то время как dsRBD1 и С-концевой глицин-богатый домены для этого не важны. Однако эти домены необходимы для определения специфичности связывания MLE с нуклеиновыми кислотами и белками *in vivo*. В частности, dsRBD1 определяет связывание MLE с X-хромосомой самцов, а С-концевой домен обеспечивает ядерную локализацию [15].

Белок без dsRBD1 и С-концевого домена был кристаллизован в присутствии аналога АТФ и оцРНК [12]. Выявлены многочисленные взаимодействия доменов dsRBD2 и OB-fold с каталитическим ядром, образованным доменами ResA1, ResA2 и HA2. “Оборачиваясь” вокруг каталитического ядра, они формируют вместе с ним кольцевую структуру неправильной формы. Направляемая доменами dsRBD2 и OB-fold оцРНК проходит по каналу внутри кольцевой структуры.

Биохимическая активность MLE *D. melanogaster* и DHX9 человека подробно изучена *in vitro*. Показана способность DHX9 расплетать самые разнообразные субстраты (рис. 2), в том числе G-квадруплексы в ДНК и РНК [5, 10, 15, 16]. Экспериментально обнаружены две особенности MLE *in vitro*, не характерные для других хеликаз DHX, а именно, исключительно высокая аффинность к полиуридиновым последовательностям, а также зависящая от домена dsRBD2 способность расплетать дуплексы РНК с “тупыми” концами (без выступающего одноцепочечного 3'-конца) [12].

Таким образом, хеликаза MLE *in vitro* способна без участия N- и С-концевых доменов расплетать вторичные структуры в молекулах нуклеиновых кислот, перемещаясь в 3' → 5'-направлении. При этом *in vivo* домены на N- и С-концах молекулы MLE обеспечивают взаимодействие с разными белковыми факторами и определяют специфичность функций этой хеликазы.

## ФУНКЦИИ MLE В ДОЗОВОЙ КОМПЕНСАЦИИ

MLE экспрессируется на всех стадиях развития самцов и самок *D. melanogaster* и имеет ядерную локализацию. Исторически сложилось так, что в большей части научных исследований хеликазу MLE рассматривали как компонент комплекса дозовой компенсации, специфичного для *D. melanogaster*. За последние три десятилетия опубликовано множество работ, посвященных дозовой компенсации, и подробное рассмотрение этой темы выходит за рамки данного обзора. Здесь представлены только общие сведения о составе и функциях комплекса дозовой компенсации, более подробно будут рассмотрены другие функции MLE.

Комплекс дозовой компенсации у *D. melanogaster* (MSL – male specific lethal) – это РНП-комплекс, формирующийся только у самцов и необходимый для их жизнеспособности. Комплекс MSL необходим для выравнивания уровней экспрессии генов в одной X-хромосоме самцов и двух X-хромосомах самок. В состав комплекса помимо MLE входят еще четыре белка – MSL1, MSL2, MSL3 и MOF (males absent on the first), и одна из двух длинных некодирующих РНК, roX1 или roX2 (RNA on the X 1 или 2) [17–21]. MSL1, MSL2, MSL3 и MOF связаны белок-белковыми взаимодействиями и формируют гетерооктамер, а MLE входит в состав комплекса через взаимодействие с РНК roX.

Встраивание одной из РНК roX в MSL необходимо для его функционирования [22, 23]. РНК roX1 и roX2 отличаются по размеру и последовательности, но имеют общие консервативные мотивы вблизи 3'-конца: обогащенные уридином последовательности, названные *roX-box*. Эти последовательности необходимы для формирования вторичных структур стебель–петля. MLE связывается с одноцепочечной последовательностью петли и расплетает РНК-дуплекс спирального стебля, открывая тем самым сайты связывания белка MSL2 [18, 24–27]. Таким образом MLE интегрирует РНК roX в комплекс MSL [28–30].

Из перечисленных компонентов комплекса MSL только MSL2 и roX экспрессируются специфично у самцов и отсутствуют у самок *D. melanogaster* [31, 32]. У самцов функциональный комплекс MSL собирается на X-хромосоме в так называемых высокоаффинных участках (High Affinity Sites, HAS), а затем распространяется по всей длине X-хромосомы [2, 33]. В разных исследованиях выявлено от нескольких десятков до полсотни сотен высокоаффинных сайтов, самыми сильными из них являются гены *roX1* и *roX2*. Последовательности HAS содержат GA-богатый мотив длиной 21 п.н., названный MRE (MSL Recognition Element). Цистеин-богатый домен белка

MSL2 способен связываться с MRE, что частично определяет связывание комплекса MSL с X-хромосомой.

Анализ ряда мутаций MLE позволил установить, что замена K413E (мутация *mle<sup>GET</sup>*), нарушающая АТРАЗную и хеликазную активности белка, приводит к тому, что комплекс MSL, в состав которого входит такой белок, остается связанным с высокоаффинными сайтами и не распространяется по всей длине X-хромосомы. Это нарушает дозовую компенсацию и приводит к летальности самцов на стадии поздней личинки. Для более точного исследования функции MLE в распространении комплекса по X-хромосоме подобраны мутации, избирательно блокирующие хеликазную активность при сохранении АТРАЗной: аминокислотные замены в мотиве III и делеция мотива V. Комплекс MSL, содержащий такой дефектный белок MLE, сохраняет способность активировать экспрессию гена, расположенного в непосредственной близости к высокоаффинному сайту, но при этом не происходит распространения комплекса на большие расстояния, как и в случае с *mle<sup>GET</sup>*. Это означает, что для распространения комплекса MSL по всей длине X-хромосомы необходима хеликазная активность MLE [2, 33, 34].

Итак, хеликаза MLE необходима для включения РНК roX в комплекс MSL и распространения этого комплекса по всей длине X-хромосомы самцов, что, в свою очередь, необходимо для дозовой компенсации.

Классические представления о механизмах дозовой компенсации у *D. melanogaster* связывают уравнивание экспрессии одной дозы генов X-хромосомы у самцов и двух доз генов X-хромосомы у самок с двукратной активацией экспрессии X-хромосомы у самцов [17–19]. Действительно, MOF является ацетилтрансферазой гистонов, специфично ацетилирующей H4K16. Обогащение X-хромосомы самцов H4K16ас за счет активности MOF в составе комплекса MSL приводит к деконденсации хроматина и способствует активации транскрипции [35]. Показано, что усиление экспрессии X-сцепленных генов осуществляется как на стадии рекрутирования РНК-полимеразы II (PolII), так и на последующих стадиях инициации транскрипции и выхода PolII из состояния паузы [18, 19]. Однако в изучении феномена дозовой компенсации у *D. melanogaster* остается много вопросов. Исследователи сходятся на том, что в специфическое распознавание X-хромосомы самцов и достижение адекватных уровней экспрессии X-сцепленных и аутомсомных генов вовлечено много факторов, взаимодействующих с комплексом MSL или действующих независимо от него. Последнее заключение связано с тем, что у триплоидных по X-хро-

мосоме самок *D. melanogaster* (метасамки), несмотря на отсутствие у них функционального комплекса MSL, наблюдается дозовая компенсация, при которой X-сцепленные гены экспрессируются на том же уровне, что и у диплоидных самок, а экспрессия аутосомных генов снижается [36]. Не известно, каким образом обеспечивается точное двукратное усиление экспрессии генов X-хромосомы самцов. В ряде исследований показано, что комплекс MSL не обеспечивает полной компенсации, усиливая экспрессию не вдвое, а менее чем в 1.5 раза [17, 19, 37–39]. При этом повышение уровня H4K16ac может привести к активации генов X-хромосомы более чем в 2 раза. Согласно последним данным, в ограничение активации до двукратного уровня вовлечен белок Mtor, который связывается с комплексом MSL [40]. Кроме того, в дозовую компенсацию вовлечены факторы, участвующие в формировании гетерохроматина, ремоделировании хроматина и РНК-интерференции [17].

Существует также альтернативная гипотеза, которая представляет дозовую компенсацию у *D. melanogaster* как компенсацию глобального анеуплоидного эффекта, вызванного X-моносомией [36, 41–45]. Моносомия X-хромосомы может привести к увеличению количества не использованных транскрипционных активаторов у самцов по сравнению с самками, что, в свою очередь, может привести к повышению экспрессии генов других хромосом. Согласно этой гипотезе, усиленное ацетилирование X-хромосомы комплексом MSL позволяет секвестрировать лишние активаторы хроматина с аутосом, предотвращая тем самым их гиперактивацию. В этом случае результаты анализа экспрессии генов у MSL-мутантов интерпретируются как общее увеличение экспрессии аутосомных генов, а не снижение экспрессии X-сцепленных генов. Для проверки этой гипотезы необходимы исследования, в которых нормализация глобального транскрипционного анализа будет основана на внешнем эталоне [18].

Дисбаланс в геноме млекопитающих и человека при анеуплоидии имеет сильнейший физиологический эффект, чаще всего приводящий к эмбриональной летальности [46–48], анеуплоидия характерна также для раковых клеток, ее находят более чем в 90% солидных опухолей [49, 50]. Поскольку у млекопитающих DHX9 участвует в эмбриогенезе как самцов, так и самок, изначально предполагалось, что она не имеет отношения к дозовой компенсации. Однако поскольку DHX9 вовлечена в канцерогенез (что будет рассмотрено ниже), ее влияние на регуляцию анеуплоидного эффекта в клетках опухолей представляется вполне возможным. Эти вопросы представляют не только фундаментальный, но и прикладной интерес для медицины.

## ВЗАИМОДЕЙСТВИЕ MLE С РНК-ПОЛИМЕРАЗой II

Хеликаза MLE обладает способностью активировать транскрипцию, физически взаимодействуя с PolII [11, 51, 52]. За эту функцию отвечает минимальный домен трансактивации (MTAD) – небольшой консервативный домен (~50 аминокислотных остатков), предшествующий каталитическому ядру хеликазы. Особенно важны для взаимодействия с PolII остатки триптофана в консервативных положениях MTAD (у *D. melanogaster* это W328, W335 и W338). MLE с делецией участка 326–378, включающей всю последовательность MTAD, равно как и MLE с заменой трех консервативных остатков триптофана на остатки аланина, не взаимодействует с PolII. Эта функция MLE консервативна в эволюции, она показана также для DHX9 человека и нематоды *C. elegans* [11]. MTAD MLE может взаимодействовать как с гиперфосфорилированной (элонгирующей), так и (преимущественно) с гипофосфорилированной (инициаторной) формами PolII. Исследования *in vivo* показали, что MLE с делецией MTAD не спасает летальный фенотип самцов с нуль-аллелем *mle*, что указывает на необходимость MTAD для функционирования MLE в комплексе MSL. Интересно, что MLE с заменами консервативных остатков триптофана, как и MLE дикого типа, имеет ядерную локализацию, а MLE с делецией MTAD находится и в ядре, и в цитоплазме, несмотря на присутствие интактного сигнала ядерной локализации на C-конце белка. Природа этого явления на данный момент не ясна. В экспериментах с репортерными трансгенами показана способность MTAD активировать транскрипцию не только с промотора *roX*, известной мишени комплекса MSL, но и транскрипцию с промотора гена теплового шока *hsp70*, на которую MSL не влияет. Показано, что MLE *D. melanogaster*, как и ее ортолог у млекопитающих, может активировать транскрипцию независимо, а не как компонент комплекса дозовой компенсации [11].

Чтобы получить представление о возможной роли MLE в регуляции транскрипции *in vivo*, проведено картирование сайтов связывания MLE на политенных хромосомах из слюнных желез личинок *D. melanogaster* на стадии L3 у обоих полов [53]. Обнаружено, что паттерн распределения MLE в составе комплекса MSL на X-хромосоме самцов не зависит от стадии развития и, соответственно, от активного или неактивного статуса района связывания. В то же время отмечено, что MSL-независимое распределение MLE на аутосомах идентично у обоих полов, зависит от транскрипционного статуса района и напоминает паттерн распределения элонгирующей формы PolII. MLE связывается практически со всеми “пуффами развития”, экдизоновыми пуффами, типичными

для анализируемых стадий, а также со многими “интербэндами”. MLE присутствует в пuffed на X-хромосомах самок. После кратковременного теплового шока белок MLE можно обнаружить во многих пuffed, появляющихся при тепловом шоке. В частности, сильный сигнал наблюдался в районах 87C, 93D и 95D, но никогда в районе 87A. Два соседних района 87A и 87C, как известно, развиваются благодаря активности кластера генов *hsp70* [54]. Однако в отличие от района 87A, район 87C содержит не только кластер *hsp70*, но и  $\alpha$ -элемент, с которого синтезируется некодирующая РНК теплового шока (*hshRNA*), обладающая неизвестными на данный момент функциями [55]. Поскольку в условиях, не зависящих от теплового шока, MLE у обоих полов не связывается с районом 87C, предполагается, что *hshRNA*, которая принадлежит к тому же классу некодирующих РНК, что и гоX, может быть специфической мишенью MLE.

Таким образом, независимо от других белков комплекса MSL, MLE взаимодействует с наиболее заметными транскрипционно-активными районами хромосом и, вероятно, проявляет повышенную аффинность к транскриптам определенных генов [53].

Показано также, что DNX9 участвует в регуляции транскрипции генов человека, привлекая PolII непосредственно на промотор генов-мишеней. DNX9 специфически связывается с определенными последовательностями ДНК в промоторных областях этих генов [56–58]. Кроме того, DNX9 может активировать транскрипцию, связываясь одновременно с PolII (через MTAD) и с такими факторами транскрипции, как СВР/p300 и BRCA1 (через домены dsRBD) [59, 60]. В ряде случаев DNX9 может репрессировать транскрипцию. Связываясь с активатором транскрипции TonEBP, DNX9 ингибирует его активность посредством неизвестного механизма, привлекая, возможно, в область транскрипции другие белки [61].

### ФУНКЦИИ MLE В СПЛАЙСИНГЕ И РЕДАКТИРОВАНИИ пре-мРНК

Первую функцию MLE, не имеющую отношения к дозовой компенсации, выявили еще в 1991 году, но ее механизм долгое время оставался неизвестным [62]. Обнаружено, что мутация *mle<sup>napts</sup>* вызывает “паралитический” фенотип как у самок, так и у самцов при повышенной температуре. Этот фенотип обусловлен снижением экспрессии белка Na<sup>+</sup>-каналов, что приводит, в свою очередь, к блокаде потенциала действия. Дальнейшее изучение показало, что транскрипт гена *para*, кодирующего белок Na<sup>+</sup>-каналов, в норме подвергается редактированию, при котором аденин дезаминируется с образованием инозина

(который в дальнейшем при трансляции ведет себя как гуанин) [63, 64]. Для редактирования необходимо образование двухцепочечной структуры в РНК, включающей редактируемый экзон и следующий за ним интрон. На фоне мутации *mle<sup>napts</sup>*, которая приводит к замене T415S в мотиве I, повышение температуры приводит к тому, что более 80% транскриптов *para* представлены aberrантными мРНК, в которых пропущен редактируемый экзон. В то же время, нуль-мутация *mle* не имеет паралитического фенотипа. Считается, что в норме MLE дикого типа расплетает двухцепочечную структуру в районе редактируемого экзона. Повышение температуры вызывает изменение конформации мотива I с заменой T415S, что приводит к снижению каталитической активности MLE. При этом мутантный белок остается связанным с дцРНК, которую он не может расплести, а донорный сайт сплайсинга редактируемого экзона остается заблокированным. В этом случае сплайсинг происходит “в обход” заблокированного донорного сайта с пропуском экзона. Поскольку нуль-мутация *mle* такого эффекта не вызывает, можно предположить, что при полном отсутствии MLE расплетание вторичной структуры пре-мРНК гена *para* выполняет какая-то другая хеликаза [63].

Регуляция сопряженного с редактированием сплайсинга *para* – единственный изученный на данный момент пример участия MLE в регуляции сплайсинга пре-мРНК конкретного гена у *D. melanogaster*. Однако масс-спектрометрический анализ белков, которые MLE соосаждает из экстракта клеток S2 [65], показал, что самая многочисленная функциональная группа (19 белков) представлена белками, участвующими в сплайсинге. Нокдаун MSL не влияет на взаимодействие с 15 из этих белков, указывая на то, что MLE в сплайсинге участвует в форме, свободной от комплекса MSL. Кроме того, большая часть обнаруженных взаимодействий (16 из 19) была РНК-зависимой: при обработке РНКазой взаимодействия между белками исчезали. Эти данные согласуются с представлениями о том, что MLE в процессе сплайсинга осуществляет расплетание и ремоделирование вторичных структур, образующихся в пре-мРНК. Взаимодействие с тремя из обнаруженных белков изучили отдельно и подтвердили с помощью коиммунопреципитации и Вестерн-блот-анализа. Эти три белка, Per (Protein on ecdysone puffs, CG6143), Hrb87F (heterogeneous nuclear ribonucleoprotein at 87F, CG12749) и Hrb98DE (heterogeneous nuclear ribonucleoprotein at 98DE, CG9983), – гетерогенные ядерные рибонуклеопротеины, участвующие в процессе РНК и депонирования белка гетерохроматина (HP1a). Они коиммунопреципитируют с MLE РНК-зависимо и независимо от MSL-комплекса. Per и Hrb87F взаимодействуют с белком HP1a,

который участвует в сайленсинге генов и формировании гетерохроматина, а также в регуляции многочисленных генов эухроматина [66]. Белок Hrb98DE является близким гомологом белка Hrb87F (эти белки идентичны на 80%) и оба они гомологичны гяРНП A1 (hnRNP A1). Эти белки участвуют преимущественно в регуляции альтернативного сплайсинга мРНК и, по-видимому, взаимозаменяемы [65, 67, 68]. Показано, что DHX9 человека входит в состав пресплайсосомы [69], связывается как с мРНК, так и с пре-мРНК [70], взаимодействует с белком-регулятором сплайсинга MBNL1 [71], а также с белком SMN, компонентом мяРНП (snRNP) [72]. Кроме того, DHX9, функционируя согласованно с ADAR2, координирует редактирование и сплайсинг пре-мРНК рецептора глутамата [73]. Как и в случае гена *para*, редактирование и сплайсинг пре-мРНК рецептора глутамата являются конкурирующими процессами: ферменту ADAR2 для осуществления редактирования требуется транскрипт, содержащий вторичную структуру стебель–петля, которая секвестрирует 5'-сайт сплайсинга. Считается, что DHX9 расплетает этот двухцепочечный участок пре-мРНК после редактирования и тем самым делает возможным последующий правильный сплайсинг. Таким образом, участие MLE в регуляции сплайсинга консервативно в эволюции и не имеет отношения к дозовой компенсации.

В упомянутом выше масс-спектрометрическом исследовании [65], помимо взаимодействия MLE с белками сплайсосомы, выявлено еще несколько десятков взаимодействий MLE с белками, участвующими в метаболизме нуклеиновых кислот. В том числе, найдены взаимодействия с несколькими РНК-хеликазами, факторами, регулируемыми трансляцию, белками, ремоделирующими хроматин. Дальнейшее изучение функций MLE происходило в следующих направлениях: исследование конкретных функций MLE в РНК-интерференции [74], ремоделировании хроматина и регуляции экспрессии экдизонзависимых генов [65, 75], о чем пойдет речь далее.

### MLE И РНК-ИНТЕРФЕРЕНЦИЯ

В число белков, взаимодействующих с MLE, входят Ago2 и Dicer2 [65]. Обнаружено также, что помимо своих главных мишеней – РНК гоХ, MLE связывается в клетках S2 *D. melanogaster* с целым рядом пре-мРНК и мРНК, локализуясь преимущественно в интронах и в 3'-UTR (untranslated regions), образующих вторичные структуры [76]. Возникло предположение, что MLE участвует в расплетании котранскрипционно генерируемых шпилек РНК [74]. Для проверки этой гипотезы проанализированы линии *D. melanogaster*, полученные в рамках проекта Transgenic

RNAi Project (TRiP) и Венского исследовательского центра дрозофилы (VDRC). Эти линии несут распределенные по всему геному *D. melanogaster* индуцибельные трансгены, экспрессирующие шпилечные РНК, специфичные для определенных кодирующих генов [77–80]. На политечных хромосомах MLE специфически выявляется в местах транскрипции этих трансгенов, и связывание MLE не зависит от размера шпильки или ее геномного расположения. Параллельный анализ трансгенов *Notch* и *Egfr* (*epidermal growth factor receptor*) показал, что MLE необходима для функциональной интерференции РНК *in vivo*. Нуль-мутация *mle* в гомозиготном состоянии резко снижает эффективность интерференции РНК в трансгенных мухах, что приводит к спасению фенотипа индуцированных мутаций. По предположению авторов, связывание MLE со шпильками РНК может быть важным для их переноса в цитоплазму, и MLE может участвовать в формировании комплекса RISC. Альтернативная гипотеза заключается в том, что MLE может устранять неправильное сворачивание только что транскрибированных шпилек РНК, что позволяет им участвовать в интерференции РНК. Возможно, что MLE вмешивается в редактирующую активность фермента ADAR. Дезаминирование аденина до инозина может привести к снижению комплементарности и нестабильности дцРНК, что ограничивает синтез эффективных siРНК. MLE может физически блокировать связывание ADAR со шпильками РНК или же расплетать структуры, участвующие в рекрутировании ADAR. Все рассмотренные гипотезы не являются взаимоисключающими. В то же время, данные об участии DHX9 млекопитающих в интерференции РНК противоречивы [81–83].

### ВЗАИМОДЕЙСТВИЕ MLE С КОМПЛЕКСАМИ РЕМОДЕЛИРОВАНИЯ ХРОМАТИНА

Взаимодействие MLE с белками Mi-2, MEP-1 и rbb не зависит от MSL, а обработка иммунопреципитатов РНКазой ослабляет, но не разрушает полностью эти взаимодействия [65]. Таким образом, MLE можно рассматривать как функционального партнера комплексов NuRD и dMec. NuRD – это комплекс, который увеличивает плотность гистонов/нуклеосом на определенных последовательностях ДНК по всему геному и отвечает за изменения в организации нуклеосом в соседних локусах [84]. Как это часто бывает с комплексами ремоделирования хроматина, NuRD и dMec могут участвовать как в репрессии, так и в активации генов [85]. У *D. melanogaster* комплекс NuRD, выделенный из эмбрионов, содержит девять субъединиц. В состав комплекса входят такие белки, как деацетилаза гистонов

RPD3, белок с цинковым пальцем MEP-1 и ДНК-связывающая АТРаза CHD4/Mi-2, обладающая хеликазной активностью [85]. MEP-1 и Mi-2 также образуют отдельный комплекс dMEC [86]. Субъединица Mi-2, общая для обоих комплексов, необходима для экспрессии генов теплового шока [87], она играет важную роль в поддержании структуры хромосом, поскольку влияет на функцию когезина [88]. Участки связывания MLE и Mi-2 в геноме существенно перекрываются, при этом большая часть общих сайтов совпадает с энхансерами (54%), промоторами и сайтами стартов транскрипции (TSS) – 37%. Вопрос о функциональной значимости взаимодействий MLE с компонентами комплексов NuRD и dMEC остается открытым [65].

MLE взаимодействует также с субъединицами комплекса SWI/SNF (Brm, PB, Var170) [75]. Функциональная значимость этого взаимодействия изучена на примере регуляции транскрипции SWI/SNF-зависимого позднего гена экдизонового каскада *ftz-f1*. Ген *ftz-f1* кодирует ядерный рецептор, играющий важную роль в процессе онтогенеза *D. melanogaster*. Транскрипция этого гена в начале метаморфоза происходит с участием паузированной PolII, ограничена узким периодом времени и подвержена сложной многоступенчатой регуляции. Важную роль в адекватной экспрессии гена *ftz-f1* играет физическое сближение промотора и энхансера, расположенного в первом интроне. Привлечение SWI/SNF на промотор и энхансер гена важно для их взаимодействия. При нокдауне субъединицы SWI/SNF, привлекающей комплекс на ген, связь промотора с энхансером ослабевает. При этом в начале метаморфоза ген *ftz-f1* начинает экспрессироваться преждевременно, а его транскрипция остается на низком уровне [89–91]. Изучение связывания MLE с геном *ftz-f1* в процессе его активации показало, что профиль связывания MLE повторяет профиль связывания комплекса SWI/SNF. MLE сильнее всего связан с промотором и энхансером, и это связывание усиливается в процессе активации транскрипции. Нокдаун MLE в модельной системе в клетках S2 и мутация *mle* в гомозиготном состоянии *in vivo* вызывают эффект, сходный с описанным для SWI/SNF: происходит преждевременная, но при этом недостаточная активация гена *ftz-f1* [75]. Чтобы исключить возможное влияние комплекса MSL, в эксперименте использовали личинок и предкулолек исключительно женского пола. Итак, MLE совместно с комплексом SWI/SNF связывается с регуляторными элементами гена *ftz-f1* и влияет на его транскрипцию. Предположительно, MLE в процессе экспрессии данного гена расплетает сложные aberrантные структуры (РНК-вилки, РНК-петли, РНК-G-квадруплексы и др.), которые образуются в процессе транскрипции с участием паузированной PolII, а

также при сближении промотора и энхансера и одновременном прохождении транскрипции.

### УЧАСТИЕ MLE В РЕГУЛЯЦИИ ЭКСПРЕССИИ РАННИХ ГЕНОВ ЭКДИЗОНОВОГО КАСКАДА

Поскольку, как описано выше, MLE связывается с экдизоновыми пуффами на политенных хромосомах личинок *D. melanogaster*, а также взаимодействует с белком Per, который характеризуется таким же распределением, закономерно возник вопрос об участии MLE в экдизоновом генном каскаде. (Влияние MLE на транскрипцию гена позднего ответа *ftz-f1* описано выше [75].) Также изучено влияние MLE на экспрессию трех ранних генов этого каскада: *Eip74EF*, *BR-C* и *Eip75B* [65]. В этих экспериментах, чтобы избежать косвенных эффектов, связанных с функцией MLE в комплексе MSL, анализ проводили на личинках женского пола. У мутантных по *mle* самок содержание транскриптов генов *Eip74EF* и *BR-C* снижалось незначительно, но статистически значимо, тогда как уровень *Eip75B* оказался несколько повышенным. Этот результат предполагает вклад MLE в регуляцию экдизон-индуцируемой транскрипции. Возможно, в экдизоновых пуффах MLE, как и Per, связан с комплексом гЯРНП [92].

### ВЗАИМОДЕЙСТВИЕ MLE С ENY2 – КОМПОНЕНТОМ ТРАНСКРИПЦИОННОГО КОМПЛЕКСА SAGA

В двугибридных скринингах библиотек кДНК *D. melanogaster* и человека обнаружено, что MLE/DHX9 взаимодействует с консервативным белком ENY2 [93, 94]. ENY2 *D. melanogaster* входит в состав транскрипционного комплекса SAGA и ряда других белковых комплексов, участвует в организации транскрипционных доменов хроматина, активации и элонгации транскрипции, экспорте мРНК из ядра и регуляции пространственного расположения генов в ядре [95–100]. ENY2 – это небольшой белок с массой ~10 кДа. Белки dENY2 *D. melanogaster* и hENY2 человека содержат 56.4% идентичных и 37.3% подобных аминокислотных остатков. В обоих случаях во взаимодействии MLE с dENY2 и DHX9 с hENY2 участвует глицин-богатый С-концевой домен, консервативный в хеликазах человека и *D. melanogaster* (в районах, взаимодействующих с ENY2, идентичны 40% аминокислотных остатков). Обнаруженное у *D. melanogaster* взаимодействие подтверждено с помощью реакций коиммунопреципитации и генетических экспериментов. Анализ сайтов колокализации MLE и ENY2 на политенных хромосомах *D. melanogaster* показал, что они



**Рис. 3.** Схема участия MLE и ее ортолога в различных внутриклеточных процессах, связанных с метаболизмом нуклеиновых кислот. Стрелками со знаком вопроса обозначены процессы, участие MLE в которых не изучено, в то время как есть данные об участии DDX9 млекопитающих в этих процессах.

не совпадают с сайтами локализации комплекса MSL и обнаруживаются в сайтах активного хроматина на всех хромосомах самцов и самок [93].

Таким образом, показано, что MLE *D. melanogaster* может участвовать в широком спектре биологических процессов, посредством взаимодействия с ENY2-содержащими комплексами.

### ФУНКЦИИ DDX9 У МЛЕКОПИТАЮЩИХ

Помимо перечисленных функций, DDX9 млекопитающих участвует также в репликации ДНК и поддержании геномной стабильности [101], в регуляции трансляции [102, 103] и в транспорте РНК [104, 105]. Схематическое изображение участия MLE и ее ортолога у млекопитающих в процессах, ассоциированных с метаболизмом нуклеиновых кислот, представлено на рис. 3.

Очень важна связь DDX9 человека и других млекопитающих с канцерогенезом. Уровень экспрессии DDX9 повышается в клетках злокачественных опухолей нескольких типов, что рассматривается как плохой прогностический признак [106, 107]. Нокдаун DDX9 в панели культур раковых клеток человека приводит к гибели большей части из них [108]. В то же время, в других типах клеток DDX9 активирует транскрипцию генов онкосупрессоров, что предполагает двойную роль этой хеликазы в канцерогенезе [56, 109].

Ряд вирусов, в том числе ВИЧ-1 [110], вирус гепатита С [111], цитомегаловирус [112], аденовирусы [113] и др., используют DDX9 для репликации, транскрипции, трансляции своих РНК. Одновременно с этим, DDX9 участвует в противовирусном иммунном ответе [114, 115].

DDX9 человека вовлечен также в патогенез аутоиммунных заболеваний: этот белок является аутоантигеном у 6–23% пациентов с системной красной волчанкой. Количество больных, в сыворотке которых выявляются аутоантитела к DDX9, существенно варьирует в разных популяциях и на разных стадиях развития заболевания. На ранних стадиях аутоантитела к DDX9 выявляются с более высокой частотой и могут быть использованы в клинической практике в качестве диагностического маркера системной красной волчанки [116, 117].

### ЗАКЛЮЧЕНИЕ

Итак, поскольку MLE у *D. melanogaster* и DDX9 у млекопитающих способны взаимодействовать с различными ядерными и цитоплазматическими белками и связываться с разными формами нуклеиновых кислот, эти хеликазы считаются одними из ключевых регуляторов экспрессии генов. Посредством разных доменов MLE и DDX9 способны взаимодействовать с разными белками и участвовать в работе разных белковых комплексов, что и определяет многообразие функций этих хеликаз. Многие функции MLE и ее ортолога консервативны в эволюции. Некоторые функции избыточны, их важность становится очевидной только в сочетании с потерей функции других хеликаз. Так, отсутствие DDX9 в геномах домашних кур и уток делает их более чувствительными к вирусным инфекциям по сравнению с млекопитающими. При этом особенно это выражено у кур, у которых отсутствует также хеликаза RIG-1 (DDX58) [8, 9].

Несмотря на интенсивные исследования, остается много неясного в том, как функционируют MLE и ее ортологи *in vivo*, как определяется

и изменяется специфичность связывания субстратов, процессивность и т.д. в клетках разных тканей, в процессах онтогенеза, в патологических процессах. Преимущество *D. melanogaster* в качестве модельного объекта для изучения функций MLE состоит в том, что потеря функции MLE не приводит к эмбриональной летальности у *D. melanogaster*, что позволяет исследовать функции этого белка на дальнейших стадиях развития, а также установить, какие белки могут компенсировать отсутствие MLE в клетках разных тканей, частично выполняя ее функции.

Поскольку потенциально ортолог MLE представляет собой мишень для противоопухолевой и противовирусной терапии и вовлечен в развитие аутоиммунных заболеваний, дальнейшее изучение функций этого белка может иметь большую практическую ценность для медицины.

Работа поддержана Российским научным фондом (№ 22-14-00270).

В работе не использовали людей и животных в качестве объектов исследования.

Авторы сообщают об отсутствии конфликта интересов.

#### СПИСОК ЛИТЕРАТУРЫ

1. Fairman-Williams M.E., Guenther U.-P., Jankowsky E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324.
2. Lee C.-G. (1997) The NTPase/helicase activities of *Drosophila* maleless, an essential factor in dosage compensation. *EMBO J.* **16**, 2671–2681.
3. Gorbalenya A.E., Koonin E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**, 419–429.
4. Singleton M.R., Dillingham M.S., Wigley D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50.
5. Lee C.G., Hurwitz J. (1993) Human RNA helicase A is homologous to the maleless protein of *Drosophila*. *J. Biol. Chem.* **268**, 16822–16830.
6. Walstrom K.M., Schmidt D., Bean C.J., Kelly W.G. (2005) RNA helicase A is important for germline transcriptional control, proliferation, and meiosis in *C. elegans*. *Mech. Dev.* **122**, 707–720.
7. Wei W., Twell D., Lindsey K. (1997) A novel nucleic acid helicase gene identified by promoter trapping in *Arabidopsis*. *Plant J.* **11**, 1307–1314.
8. Sato H., Oshiumi H., Takaki H., Hikono H., Seya T. (2015) Evolution of the DEAD box helicase family in chicken: chickens have no DHX9 ortholog. *Microbiol. Immunol.* **59**, 633–640.
9. Barber M.R.W., Aldridge J.R., Webster R.G., Magor K.E. (2010) Association of RIG-I with innate immunity of ducks to influenza. *Proc. Natl. Acad. Sci. USA.* **107**, 5913–5918.
10. Lee T., Pelletier J. (2016) The biology of DHX9 and its potential as a therapeutic target. *Oncotarget.* **7**, 42716–42739.
11. Aratani S., Kageyama Y., Nakamura A., Fujita H., Fujii R., Nishioka K., Nakajima T. (2008) MLE activates transcription via the minimal transactivation domain in *Drosophila*. *Int. J. Mol. Med.* **21**, 469–476.
12. Prabu J.R., Müller M., Thomae A.W., Schüssler S., Bonneau F., Becker P.B., Conti E. (2015) Structure of the RNA helicase MLE reveals the molecular mechanisms for uridine specificity and RNA–ATP coupling. *Mol. Cell.* **60**, 487–499.
13. Xing L., Zhao X., Niu M., Kleiman L. (2014) Helicase associated 2 domain is essential for helicase activity of RNA helicase A. *Biochim. Biophys. Acta (BBA) – Proteins Proteomics.* **1844**, 1757–1764.
14. Kuroda M.I., Kernan M.J., Kreber R., Ganetzky B., Baker B.S. (1991) The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila*. *Cell.* **66**, 935–947.
15. Izzo A., Regnard C., Morales V., Kremmer E., Becker P.B. (2008) Structure-function analysis of the RNA helicase maleless. *Nucl. Acids Res.* **36**, 950–962.
16. Robinson J., Raguseo F., Nuccio S.P., Liano D., Di Antonio M. (2021) DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucl. Acids Res.* **49**, 8419–8431.
17. Makki R., Meller V.H. (2021) When down is up: heterochromatin, nuclear organization and X upregulation. *Cells.* **10**, 3416.
18. Samata M., Akhtar A. (2018) Dosage compensation of the X chromosome: a complex epigenetic assignment involving chromatin regulators and long noncoding RNAs. *Annu. Rev. Biochem.* **87**, 323–350.
19. Kuroda M.I., Hilfiker A., Lucchesi J.C. (2016) Dosage compensation in *Drosophila* – a model for the coordinate regulation of transcription. *Genetics.* **204**, 435–450.
20. Keller C.I., Akhtar A. (2015) The MSL complex: juggling RNA–protein interactions for dosage compensation and beyond. *Curr. Opin. Genet. Develop.* **31**, 1–11.
21. Georgiev P., Chlamydas S., Akhtar A. (2011) *Drosophila* dosage compensation. *Fly (Austin).* **5**, 147–154.
22. Franke A., Baker B.S. (1999) The Rox1 and Rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol. Cell.* **4**, 117–122.
23. Meller V.H. (2002) The *RoX* genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J.* **21**, 1084–1091.
24. Park S.-W., Kuroda M.I., Park Y. (2008) Regulation of histone H4 Lys16 acetylation by predicted alternative secondary structures in RoX noncoding RNAs. *Mol. Cell. Biol.* **28**, 4952–4962.
25. Park S.-W., Kang Y.I., Sypula J.G., Choi J., Oh H., Park Y. (2007) An evolutionarily conserved domain of RoX2 RNA is sufficient for induction of H4-Lys16 acetylation on the *Drosophila* X chromosome. *Genetics.* **177**, 1429–1437.
26. Stuckenholtz C., Meller V.H., Kuroda M.I. (2003) Functional redundancy within roX1, a noncoding

- RNA involved in dosage compensation in *Drosophila melanogaster*. *Genetics*. **164**, 1003–1014.
27. Kelley R.L., Lee O.-K., Shim Y.-K. (2008) Transcription rate of noncoding roX1 RNA controls local spreading of the *Drosophila* MSL chromatin remodeling complex. *Mech. Dev.* **125**, 1009–1019.
  28. Ilik I.A., Quinn J., Georgiev P., Tavares-Cadete F., Maticzka D., Toscano S., Wan Y., Spitale R., Luscombe N., Backofen R., Chang H., Akhtar A. (2013) Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell*. **51**, 156–173.
  29. Militti C., Maenner S., Becker P.B., Gebauer F. (2014) UNR facilitates the interaction of MLE with the lncRNA roX2 during *Drosophila* dosage compensation. *Nat. Commun.* **5**, 4762.
  30. Maenner S., Müller M., Fröhlich J., Langer D., Becker P.B. (2013) ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol. Cell*. **51**, 174–184.
  31. Bashaw G.J., Baker B.S. (1997) The regulation of the *Drosophila msl-2* gene reveals a function for sex-lethal in translational control. *Cell*. **89**, 789–798.
  32. Kelley R.L., Wang J., Bell L., Kuroda M.I. (1997) Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature*. **387**, 195–199.
  33. Morra R., Smith E.R., Yokoyama R., Lucchesi J.C. (2008) The MLE subunit of the *Drosophila* MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. *Mol. Cell. Biol.* **28**, 958–966.
  34. Morra R., Yokoyama R., Ling H., Lucchesi J.C. (2011) Role of the ATPase/helicase maleless (MLE) in the assembly, targeting, spreading and function of the male-specific lethal (MSL) complex of *Drosophila*. *Epigenet. Chromatin*. **4**, 6.
  35. Akhtar A., Becker P.B. (2000) Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol. Cell*. **5**, 367–375.
  36. Sun L., Johnson A.F., Donohue R.C., Li J., Cheng J., Birchler J.A. (2013) Dosage compensation and inverse effects in triple X metafemales of *Drosophila*. *Proc. Natl. Acad. Sci. USA*. **110**, 7383–7388.
  37. Deng X., Meller V.H. (2006) roX RNAs are required for increased expression of X-linked genes in *Drosophila melanogaster* males. *Genetics*. **174**, 1859–1866.
  38. Straub T., Gilfillan G.D., Maier V.K., Becker P.B. (2005) The *Drosophila* MSL complex activates the transcription of target genes. *Genes Dev.* **19**, 2284–2288.
  39. Hamada F.N., Park P.J., Gordadze P.R., Kuroda M.I. (2005) Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev.* **19**, 2289–2294.
  40. Aleman J.R., Kuhn T.M., Pascual-Garcia P., Gospovic J., Lan Y., Bonasio R., Little S.C., Capelson M. (2021) Correct dosage of X chromosome transcription is controlled by a nuclear pore component. *Cell Rep.* **35**, 109236.
  41. Bhadra U., Gandhi S.G., Palaparthi R., Balyan M.K., Pal-Bhadra M. (2016) *Drosophila* maleless gene counteracts X global aneuploid effects in males. *FEBS J.* **283**, 3457–3470.
  42. Sun L., Fernandez H.R., Donohue R.C., Li J., Cheng J., Birchler J.A. (2013) Male-specific lethal complex in *Drosophila* counteracts histone acetylation and does not mediate dosage compensation. *Proc. Natl. Acad. Sci. USA*. **110**, E808–E817.
  43. Sun L., Johnson A.F., Li J., Lambdin A.S., Cheng J., Birchler J.A. (2013) Differential effect of aneuploidy on the X chromosome and genes with sex-biased expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA*. **110**, 16514–16519.
  44. Zhang S., Qi H., Huang C., Yuan L., Zhang L., Wang R., Tian Y., Sun L. (2021) Interaction of male specific lethal complex and genomic imbalance on global gene expression in *Drosophila*. *Sci. Rep.* **11**, 19679.
  45. Birchler J.A. (2016) Parallel universes for models of X chromosome dosage compensation in *Drosophila*: a review. *Cytogenet. Genome Res.* **148**, 52–67.
  46. Nawata H., Kashino G., Tano K., Daino K., Shimada Y., Kugoh H., Oshimura M., Watanabe M. (2011) Dysregulation of gene expression in the artificial human trisomy cells of chromosome 8 associated with transformed cell phenotypes. *PLoS One*. **6**, e25319.
  47. Torres E.M., Williams B.R., Amon A. (2008) Aneuploidy: cells losing their balance. *Genetics*. **179**, 737–746.
  48. Jallepalli P.V., Pellman D. (2007) Aneuploidy in the balance. *Science*. **317**, 904–905.
  49. Williams B.R., Amon A. (2009) Aneuploidy: cancer's fatal flaw? *Cancer Res.* **69**, 5289–5291.
  50. Taylor A.M., Shih J., Ha G., Gao G.F., Zhang X., Berger A.C., Schumacher S.E., Wang C., Hu H., Liu J., Lazar A.J., Cancer Genome Atlas Research Network; Cherniack A.D., Beroukhi R., Meyerson M. (2018) Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell*. **33**, 676–689, e3.
  51. Aratani S., Fujii R., Fujita H., Fukamizu A., Nakajima T. (2003) Aromatic residues are required for RNA helicase A mediated transactivation. *Int. J. Mol. Med.* **12**, 175–180.
  52. Aratani S., Fujii R., Oishi T., Fujita H., Amano T., Ohshima T., Hagiwara M., Fukamizu A., Nakajima T. (2001) Dual roles of RNA helicase A in CREB-dependent transcription. *Mol. Cell. Biol.* **21**, 4460–4469.
  53. Kotlikova I.V., Demakova O.V., Semeshin V.F., Shloima V.V., Boldyreva L.V., Kuroda M.I., Zhimulev I.F. (2006) The *Drosophila* dosage compensation complex binds to polytene chromosomes independently of developmental changes in transcription. *Genetics*. **172**, 963–974.
  54. Ish-Horowicz D., Pinchin S.M., Gausz J., Gyurkovics H., Bencze G., Goldschmidt-Clermont M., Holden J.J. (1979) Deletion mapping of two *D. melanogaster* loci that code for the 70,000 dalton heat-induced protein. *Cell*. **17**, 565–571.
  55. Sharma A., Lakhota S.C. (1995) *In situ* quantification of hsp70 and alpha-beta transcripts at 87A and 87C loci in relation to *tohsr-omega* gene activity in polytene cells of *Drosophila melanogaster*. *Chromosome Res.* **3**, 386–393.

56. Myöhänen S., Baylin S.B. (2001) Sequence-specific DNA binding activity of RNA helicase A to the p16 promoter. *J. Biol. Chem.* **276**, 1634–1642.
57. Zhong X., Safa A.R. (2004) RNA helicase A in the MEF1 transcription factor complex up-regulates the *MDR1* gene in multidrug-resistant cancer cells. *J. Biol. Chem.* **279**, 17134–17141.
58. Huo L., Wang Y.-N., Xia W., Hsu S.-C., Lai C.-C., Li L.-Y., Chang W.-C., Wang Y., Hsu M.-C., Yu Y.-L., Huang T.-H., Ding Q., Chen C.-H., Tsai C.-H., Hung M.-C. (2010) RNA helicase A is a DNA-binding partner for EGFR-mediated transcriptional activation in the nucleus. *Proc. Natl. Acad. Sci. USA.* **107**, 16125–16130.
59. Nakajima T., Uchida C., Anderson S.F., Lee C.-G., Hurwitz J., Parvin J.D., Montminy M. (1997) RNA helicase A mediates association of CBP with RNA polymerase II. *Cell.* **90**, 1107–1112.
60. Anderson S.F., Schlegel B.P., Nakajima T., Wolpin E.S., Parvin J.D. (1998) BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. *Nat. Genet.* **19**, 254–256.
61. Colla E., Lee S.D., Sheen M.R., Woo S.K., Kwon H.M. (2006) TonEBP is inhibited by RNA helicase A via interaction involving the E'F loop. *Biochem. J.* **393**, 411–419.
62. Kernan M.J., Kuroda M.I., Kreber R., Baker B.S., Ganetzky B. (1991) naps, a mutation affecting sodium channel activity in *Drosophila*, is an allele of *mle* a regulator of X chromosome transcription. *Cell.* **66**, 949–959.
63. Reenan R.A., Hanrahan C.J., Ganetzky B. (2000) The *mle*(naps) RNA helicase mutation in *Drosophila* results in a splicing catastrophe of the para Na<sup>+</sup> channel transcript in a region of RNA editing. *Neuron.* **25**, 139–149.
64. Hanrahan C.J., Palladino M.J., Ganetzky B., Reenan R.A. (2000) RNA editing of the *Drosophila para* Na<sup>+</sup> channel transcript: evolutionary conservation and developmental regulation. *Genetics.* **155**, 1149–1160.
65. Cugusi S., Kallappagoudar S., Ling H., Lucchesi J.C. (2015) The *Drosophila* helicase maleless (MLE) is implicated in functions distinct from its role in dosage compensation. *Mol. Cell. Proteomics.* **14**, 1478–1488.
66. Piacentini L., Fanti L., Negri R., del Vescovo V., Fatica A., Altieri F., Pimpinelli S. (2009) Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in *Drosophila*. *PLoS Genet.* **5**, e1000670.
67. Herold N., Will C.L., Wolf E., Kastner B., Urlaub H., Lührmann R. (2009) Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. *Mol. Cell. Biol.* **29**, 281–301.
68. Borah S., Wong A.C., Steitz J.A. (2009) *Drosophila* hnRNPA1 homologs Hrp36/Hrp38 enhance U2-type versus U12-type splicing to regulate alternative splicing of the *prospero* twintron. *Proc. Natl. Acad. Sci. USA.* **106**, 2577–2582.
69. Hartmuth K., Urlaub H., Vornlocher H.-P., Will C.L., Gentzel M., Wilm M., Lührmann R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. USA.* **99**, 16719–16724.
70. Zhang S., Herrmann C., Grosse F. (1999) Pre-mRNA and mRNA binding of human nuclear DNA helicase II (RNA helicase A). *J. Cell Sci.* **112**, 1055–1064.
71. Paul S., Dansithong W., Jog S.P., Holt I., Mittal S., Brook J.D., Morris G.E., Comai L., Reddy S. (2011) Expanded CUG repeats dysregulate RNA splicing by altering the stoichiometry of the muscleblind 1 complex. *J. Biol. Chem.* **286**, 38427–38438.
72. Terns M.P., Terns R.M. (2001) Macromolecular complexes: SMN – the master assembler. *Curr. Biol.* **11**, R862–R864.
73. Bratt E., Öhman M. (2003) Coordination of editing and splicing of glutamate receptor pre-mRNA. *RNA.* **9**, 309–318.
74. Cugusi S., Li Y., Jin P., Lucchesi J.C. (2016) The *Drosophila* helicase MLE targets hairpin structures in genomic transcripts. *PLoS Genet.* **12**, e1005761.
75. Николенко Ю.В., Куршакова М.М., Краснов А.Н., Георгиева С.Г. (2021) Хеликаза MLE – новый участник регуляции транскрипции гена *ftz-f1*, кодирующего ядерный рецептор у высших эукариот. *Докл. Акад. Наук. Науки о жизни.* **496**, 48–51.
76. Ilik I.A., Maticzka D., Georgiev P., Gutierrez N.M., Backofen R., Akhtar A. (2017) A mutually exclusive stem-loop arrangement in roX2 RNA is essential for X-chromosome regulation in *Drosophila*. *Genes Dev.* **31**, 1973–1987.
77. Ni J.Q., Markstein M., Binari R., Pfeiffer B., Liu L.-P., Villalta C., Booker M., Perkins L., Perrimon N. (2008) Vector and parameters for targeted transgenic RNA interference in *Drosophila melanogaster*. *Nat. Meth.* **5**, 49–51.
78. Ni J.Q., Liu L.P., Binari R., Hardy R., Shim H.S., Cavallaro A., Booker M., Pfeiffer B.D., Markstein M., Wang H., Villalta C., Lavery T., Perkins L., Perrimon N. (2009) A *Drosophila* resource of transgenic RNAi lines for neurogenetics. *Genetics.* **182**, 1089–1100.
79. Ni J.Q., Zhou R., Czech B., Liu L.P., Holderbaum L., Yang-Zhou D., Shim H.S., Tao R., Handler D., Karpowicz P., Binari R., Booker M., Brennecke J., Perkins L., Hannon G., Perrimon N. (2011) A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nat. Meth.* **8**, 405–407.
80. Dietzl G., Chen D., Schnorrer F., Su K., Barinova Y., Fellner M., Gasser B., Kinsey K., Oettel S., Scheiblauer S., Couto A., Marra V., Keleman K., Dickson B. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature.* **448**, 151–156.
81. Fu Q., Yuan Y.A. (2013) Structural insights into RISC assembly facilitated by dsRNA-binding domains of human RNA helicase A (DHX9). *Nucl. Acids Res.* **41**, 12–23.
82. Robb G.B., Rana T.M. (2007) RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol. Cell.* **26**, 523–537.
83. Liang X., Croke S.T. (2013) RNA helicase A is not required for RISC activity. *Biochim. Biophys. Acta (BBA) – Gene Regul. Mech.* **1829**, 1092–1101.

84. Moshkin Y.M., Chalkley G., Kan T., Reddy B., Ozgur Z., van Ijcken W., Dekkers D., Demmers J., Travers A., Verrijzer C.P. (2012) Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. *Mol. Cell. Biol.* **32**, 675–688.
85. Reddy B.A., Bajpe P.K., Bassett A., Moshkin Y.M., Kozhevnikova E., Bezstarosti K., Demmers J.A.A., Travers A.A., Verrijzer C.P. (2010) *Drosophila* transcription factor Tramtrack69 binds MEP1 to recruit the chromatin remodeler NuRD. *Mol. Cell. Biol.* **30**, 5234–5244.
86. Kunert N., Wagner E., Murawska M., Klinker H., Kremmer E., Brehm A. (2009) dMec: a novel Mi-2 chromatin remodelling complex involved in transcriptional repression. *EMBO J.* **28**, 533–544.
87. Murawska M., Hassler M., Renkawitz-Pohl R., Ladurner A., Brehm A. (2011) Stress-induced PARP activation mediates recruitment of *Drosophila* Mi-2 to promote heat shock gene expression. *PLoS Genet.* **7**, e1002206.
88. Fasulo B., Deuring R., Murawska M., Gause M., Dorigi K.M., Schaaf C.A., Dorsett D., Brehm A., Tamkun J.W. (2012) The *Drosophila* Mi-2 chromatin-remodeling factor regulates higher-order chromatin structure and cohesin dynamics *in vivo*. *PLoS Genet.* **8**, e1002878.
89. Николенко Ю.В., Краснов А.Н., Воробьева Н.Е. Ремоделирующий хроматин комплекс SWI/SNF влияет на пространственную организацию локуса гена *ftz-fl*. (2019) *Генетика.* **55**, 156–164.
90. Николенко Ю.В., Краснов А.Н., Мазина М.Ю., Георгиева С.Г., Воробьева Н.Е. (2017) Изучение свойств нового экдизонзависимого энхансера. *Докл. Акад. Наук.* **474**, 756–759.
91. Vorobyeva N.E., Nikolenko J.V., Nabirochkina E.N., Krasnov A.N., Shidlovskii Y.V., Georgieva S.G. (2012) SAYP and Brahma are important for ‘repressive’ and ‘transient’ Pol II pausing. *Nucl. Acids Res.* **40**, 7319–7331.
92. Amero S.A., Matunis M.J., Matunis E.L., Hockensmith J.W., Raychaudhuri G., Beyer A.L. (1993) A unique ribonucleoprotein complex assembles preferentially on ecdysone-responsive sites in *Drosophila melanogaster*. *Mol. Cell. Biol.* **13**, 5323–5330.
93. Николенко Ю.В., Куршакова М.М., Краснов А.Н. (2019) Мультифункциональный белок ENY2 взаимодействует с РНК-хеликазой MLE. *Докл. Акад. Наук.* **489**, 637–640.
94. Georgieva S., Nabirochkina E., Dilworth F.J., Eickhoff H., Becker P., Tora L., Georgiev P., Soldatov A. (2001) The novel transcription factor e(y)2 interacts with TAF<sub>II</sub> 40 and potentiates transcription activation on chromatin templates. *Mol. Cell. Biol.* **21**, 5223–5231.
95. Gurskiy D., Orlova A., Vorobyeva N., Nabirochkina E., Krasnov A., Shidlovskii Y., Georgieva S., Kopytova D. (2012) The DUBm subunit Sgf11 is required for mRNA export and interacts with Cbp80 in *Drosophila*. *Nucl. Acids Res.* **40**, 10689–10700.
96. Николенко Ю.В., Вдовина Ю.А., Фефелова Е.И., Глухова А.А., Набирочкина Е.Н., Копытова Д.В. (2021) Деубиквитирующий (DUB) модуль SAGA участвует в Pol III-зависимой транскрипции. *Молекуляр. биология.* **55**, 500–509.
97. Kopytova D.V., Krasnov A.N., Orlova A.V., Gurskiy D.Ya., Nabirochkina E.N., Georgieva S.G., Shidlovskii Y.V. (2010) ENY2: Couple, triple ... more? *Cell Cycle.* **9**, 479–481.
98. Popova V.V., Orlova A.V., Kurshakova M.M., Nikolenko J.V., Nabirochkina E.N., Georgieva S.G., Kopytova D.V. (2018) The role of SAGA coactivator complex in snRNA transcription. *Cell Cycle.* **17**, 1859–1870.
99. Kopytova D.V., Orlova A.V., Krasnov A.N., Gurskiy D.Ya., Nikolenko J.V., Nabirochkina E.N., Shidlovskii Y.V., Georgieva S.G. (2010) Multifunctional factor ENY2 is associated with the THO complex and promotes its recruitment onto nascent mRNA. *Genes Dev.* **24**, 86–96.
100. Фурсова Н.А., Николенко Ю.В., Сошникова Н.В., Мазина М.Ю., Воробьева Н.Е., Краснов А.Н. (2018) Белок CG9890 с доменами цинковых пальцев – новый компонент ENY2-содержащих комплексов дрозофилы. *Acta Naturae.* **10**, 110–114.
101. Zhou K. (2003) RNA helicase A interacts with dsDNA and topoisomerase II alpha. *Nucl. Acids Res.* **31**, 2253–2260.
102. Hartman T.R., Qian S., Bolinger C., Fernandez S., Schoenberg D.R., Boris-Lawrie K. (2006) RNA helicase A is necessary for translation of selected messenger RNAs. *Nat. Struct. Mol. Biol.* **13**, 509–516.
103. Ranji A., Shkriabai N., Kvaratskhelia M., Musier-Forsyth K., Boris-Lawrie K. (2011) Features of double-stranded RNA-binding domains of RNA helicase A are necessary for selective recognition and translation of complex mRNAs. *J. Biol. Chem.* **286**, 5328–5337.
104. Zhang S., Buder K., Burkhardt C., Schlott B., Görlach M., Grosse F. (2002) Nuclear DNA helicase II/RNA helicase A binds to filamentous actin. *J. Biol. Chem.* **277**, 843–853.
105. Tang H., Wong-Staal F. (2000) Specific interaction between RNA helicase A and tap, two cellular proteins that bind to the constitutive transport element of type D retrovirus. *J. Biol. Chem.* **275**, 32694–32700.
106. Wei X., Pacyna-Gengelbach M., Schlüns K., An Q., Gao Y., Cheng S., Petersen I. (2004) Analysis of the RNA helicase A gene in human lung cancer. *Oncol. Rep.* **11**, 253–258.
107. Sun Z., Wang L., Eckloff B., Deng B., Wang Y., Wampfler J., Jang J., Wieben E., Jen J., You M., Yang P. (2014) Conserved recurrent gene mutations correlate with pathway deregulation and clinical outcomes of lung adenocarcinoma in never-smokers. *BMC Med. Genomics.* **7**, 486.
108. Lee T., Paquet M., Larsson O., Pelletier J. (2016) Tumor cell survival dependence on the DHX9 DEXH-box helicase. *Oncogene.* **35**, 5093–5105.
109. Chen Z.X., Wallis K., Fell S., Sobrado V., Hemmer M., Ramsköld D., Hellman U., Sandberg R., Kenchappa R., Martinson T., Johnsen J., Kogner P., Schlisio S. (2014) RNA helicase A is a downstream mediator of KIF1B $\beta$  tumor-suppressor function in neuroblastoma. *Cancer Discovery.* **4**, 434–451.

110. Xing L., Niu M., Zhao X., Kleiman L. (2013) Roles of the linker region of RNA helicase A in HIV-1 RNA metabolism. *PLoS One*. **8**, e78596.
111. He Q.S., Tang H., Zhang J., Truong K., Wong-Staal F., Zhou D. (2008) Comparisons of RNAi approaches for validation of human RNA helicase A as an essential factor in hepatitis C virus replication. *J. Virol. Meth.* **154**, 216–219.
112. Lenarcic E.M., Ziehr B.J., Moorman N.J. (2015) An unbiased proteomics approach to identify human cytomegalovirus RNA-associated proteins. *Virology*. **481**, 13–23.
113. Liao H.-J., Kobayashi R., Mathews M.B. (1998) Activities of adenovirus virus-associated RNAs: purification and characterization of RNA binding proteins. *Proc. Natl. Acad. Sci. USA*. **95**, 8514–8519.
114. Fuchsová B., Novák P., Kafková J., Hozák P. (2002) Nuclear DNA helicase II is recruited to IFN- $\alpha$ -activated transcription sites at PML nuclear bodies. *J. Cell Biol.* **158**, 463–473.
115. Sadler A.J., Latchoumanin O., Hawkes D., Mak J., Williams B.R.G. (2009) An antiviral response directed by PKR phosphorylation of the RNA helicase A. *PLoS Pathogens*. **5**, e1000311.
116. Yamasaki Y., Narain S., Yoshida H., Hernandez L., Barker T., Hahn P., Sobel E., Segal M., Richards H., Chan E., Reeves W., Satoh M. (2007) Autoantibodies to RNA helicase A: a new serologic marker of early lupus. *Arthritis Rheumatism*. **56**, 596–604.
117. Vazquez-Del Mercado M., Palafox-Sanchez C., Munoz-Valle J., Orozco-Barocio G., Oregon-Romero E., Navarro-Hernandez R., Salazar-Paramo M., Armendariz-Borunda J., Gamez-Nava J., Gonzalez-Lopez L., Chan J., Chan E., Satoh M. (2010) High prevalence of autoantibodies to RNA helicase A in Mexican patients with systemic lupus erythematosus. *Arthritis Res. Therapy*. **12**, R6

## The Diversity of MLE Helicase Functions in the Regulation of Gene Expression in Higher Eukaryotes

J. V. Nikolenko<sup>1</sup>, \*, S. G. Georgieva<sup>1</sup>, and D. V. Kopytova<sup>1</sup>

<sup>1</sup>Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia

\*e-mail: julia.v.nikolenko@gmail.com

MLE (Maleless) protein of *D. melanogaster* is a conserved helicase involved in a wide range of gene expression regulation processes. MLE ortholog, named DHX9, has been found in many higher eukaryotes, including humans. It is involved in such diverse processes as maintenance of genome stability, replication, transcription, splicing, editing and transport of cellular and viral RNAs, and translation regulation. Some of these functions have been studied in detail to date, but the most of them remain uncharacterized. The study of the functions of MLE ortholog in mammals *in vivo* is limited by the fact that the loss of function of this protein is lethal at the embryonic stage. Helicase MLE in *D. melanogaster* was originally discovered and studied for a long time as a participant in the dosage compensation process. However, in recent years, evidence has emerged that in *D. melanogaster* helicase MLE is involved in cellular processes similar to those in which it participates in mammals, and that many functions of this protein are conserved in evolution. In addition, in experiments on *D. melanogaster* new important functions of MLE were discovered, such as participation in hormone-dependent regulation of transcription, interaction with the SAGA transcription complex and other transcription cofactors and chromatin remodeling complexes. In contrast to mammals, in *D. melanogaster*, MLE mutations do not lead to death at the embryonic stage and allow the functions of this protein to be studied *in vivo* throughout ontogenesis in females and up to the pupal stage in males. The MLE ortholog in humans is a potential target for anticancer and antiviral therapy. Therefore, continued study of the functions of this helicase in the *D. melanogaster* model organism seems important and promising in both fundamental and practical aspects. In this review, the systematic position, domain structure, and conserved and specific functions of MLE helicase in *D. melanogaster* are discussed.

**Keywords:** MLE, DHX9, RNA-helicase, *D. melanogaster*, regulation of gene expression

УДК 57.052;57.056;575.827.2

## РЕГУЛЯТОРНЫЙ ПОТЕНЦИАЛ SNP-МАРКЕРОВ В ГЕНАХ, КОДИРУЮЩИХ БЕЛКИ СИСТЕМ РЕПАРАЦИИ ДНК<sup>1</sup>

© 2023 г. Н. П. Бабушкина<sup>а</sup>\*, А. Н. Кучер<sup>а</sup>

<sup>а</sup>Научно-исследовательский институт медицинской генетики,  
Томский национальный исследовательский медицинский центр Российской академии наук,  
Томск, 634050 Россия

\*e-mail: nad.babushkina@medgenetics.ru

Поступила в редакцию 11.05.2022 г.

После доработки 16.08.2022 г.

Принята к публикации 16.08.2022 г.

Выявление широчайшего спектра локализованных в некодирующих участках генома однонуклеотидных полиморфизмов (SNP), ассоциированных с заболеваниями человека и патогенетически значимыми признаками, остро поставило вопрос по идентификации механизмов, объясняющих эти связи. Ранее нами выявлен ряд ассоциаций полиморфных вариантов генов, кодирующих белки репарации ДНК, с многофакторными заболеваниями. Для выяснения возможных механизмов, лежащих в их основе, нами проведена подробная аннотация регуляторного потенциала изучаемых маркеров с использованием ряда on-line ресурсов (GTXPortal, VannoPortal, Ensemble, RegulomeDB, Polympact, UCSC, GnomAD, ENCODE, GeneHancer, EpiMap Epigenomics 2021, HaploReg, GWAS4D, JASPAR, ORegAnno, DisGeNet, OMIM). В статье охарактеризован регуляторный потенциал следующих полиморфных вариантов: rs560191 (в гене *TP53BP1*), rs1805800 и rs709816 (*NBN*), rs473297 (*MRE11*), rs189037 и rs1801516 (*ATM*), rs1799977 (*MLH1*), rs1805321 (*PMS2*), rs20579 (*LIG1*). Приведена как общая характеристика изученных маркеров, так и информация по их влиянию на экспрессию “своего” и корегулируемых генов, на аффинность связывания факторов транскрипции. Рассмотрены опубликованные данные по адаптогенному и патологическому потенциалу этих SNP и о колокализированных с ними модификациях гистонов. Потенциальная вовлеченность на различных уровнях в регуляцию функционирования не только генов, в состав которых входят исследованные маркеры, но и близлежащих генов может объяснять ассоциированность изученных SNP с заболеваниями и их клиническими фенотипами.

**Ключевые слова:** SNP, ассоциации, регуляция экспрессии, сплайсинг, транскрипционные факторы, гистоновый код, патогенность, консервативность

**DOI:** 10.31857/S0026898423010032, **EDN:** AXUPQD

### ВВЕДЕНИЕ

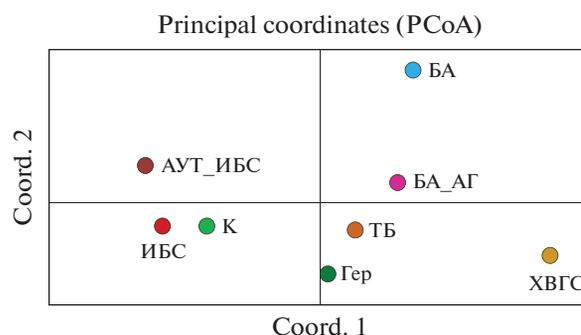
Смещение акцентов в ассоциативных исследованиях от анализа генов-кандидатов к полногеномным ассоциативным исследованиям (GWAS) привело к открытию большого числа новых маркеров, ассоциированных с заболеваниями и количественными (в том числе и патогенетически значимыми) признаками. Оказалось, что среди ассоциированных однонуклеотидных полимор-

физмов (SNP) большая часть локализована в некодирующих участках генов и межгенных регионах [1]. В то же время ассоциированные с заболеваниями генетические варианты, выявленные при проведении GWAS, необязательно “указывают” на наличие функциональной значимости для развития патологии именно этих вариантов, не всегда понятно, как эти варианты изменяют функциональное состояние клетки и в итоге работу организма в целом, определяя риск развития патологии. В этой связи в последние годы особенно активно стали развиваться биоинформатические подходы, позволяющие оценить функциональную значимость полиморфных вариантов не напрямую, а через существующие блоки сцепления, эпистатические взаимодействия с различными генами, “взаимоотношения” различных уровней регуляции.

<sup>1</sup> Дополнительная информация для этой статьи доступна по DOI 10.31857/S0026898423010032.

Сокращения: enhD (enhancer-like distal element) – энхансерподобный дистальный элемент; enhP (enhancer-like proximal element) – энхансерподобный проксимальный элемент; GWAS (Genome-Wide Association Studies) – полногеномные ассоциативные исследования; SNP (single nucleotide polymorphism) – однонуклеотидный полиморфизм(ы); TF (transcription factor) – транскрипционный фактор; ИБС – ишемическая болезнь сердца.

К числу относительно новых генов, структурная вариабельность которых может вносить вклад в формирование предрасположенности к различным многофакторным заболеваниям, относятся гены, кодирующие белки систем репарации ДНК. Помимо вовлеченности в канцерогенез (чему посвящены многочисленные исследования), накапливаются данные о значимости этих генов при развитии заболеваний сердечно-сосудистой, пищеварительной, мочеполовой, скелетно-мышечной, гемопозитической и других систем (см. обзор [2]). Ранее нами изучена вовлеченность 9 SNP в генах, кодирующих белки различных репарационных систем: *TP53BP1* (rs560191), *NBN* (rs1805800, rs709816), *MRE11* (rs473297), *ATM* (rs189037, rs1801516), *MLH1* (rs1799977), *PMS2* (rs1805321), *LIG1* (rs20579), – в развитие многофакторной патологии различной этиологии. Мы проанализировали около 1.5 тыс. образцов ДНК из банка ДНК НИИ медицинской генетики Томского НИМЦ РАН, полученных от пациентов с различными заболеваниями: изолированная аллергическая бронхиальная астма (БА) и смешанная БА в сочетании с артериальной гипертензией (БА\_АГ), хронический вирусный гепатит С (ХВГС), туберкулез (ТБ), сердечно-сосудистые заболевания (ишемическая болезнь сердца (ИБС) и аутопсийный материал умерших в результате сердечно-сосудистого события (АУТ\_ИБС)); а также от жителей г. Томска (популяционная выборка – К) и долгожителей (Гер) [3–8]. При незначительной подразделенности изученной популяции на субгруппы в соответствии с основными диагнозами пациентов (попарные  $F_{ST} < 1\%$ ) выявлен неплохой дифференцирующий потенциал маркеров. При визуализации матрицы генетических расстояний видно (рис. 1), что первая координата отделяет контроль (К) и группы пациентов с ИБС и умерших в раннем возрасте от сердечно-сосудистых катастроф (АУТ\_ИБС) от групп пациентов с другими заболеваниями. Это может свидетельствовать о малом вкладе изученных маркеров генов, кодирующих белки систем репарации ДНК, в развитие сердечно-сосудистой патологии, что согласуется с результатами ассоциативного анализа: прямых ассоциаций с патологией не выявлено, ассоциации зарегистрированы только с патогенетически значимыми признаками [3]. Вторая координата от кластера сердечно-сосудистой патологии и контроля отделяет группу умерших от сердечно-сосудистого заболевания, из чего можно предполагать вклад изученных маркеров в раннее развитие или в быструю прогрессию таких заболеваний и, как следствие, в ранние неблагоприятные исходы. От большого кластера прочих патологий вторая координата отделяет БА – изолированную форму и сочетанную с артериальной гипертензией, – что предполагает наличие вклада изученных маркеров в аллергическую компонен-



**Рис. 1.** Расположение изученных групп патологий в поле главных координат на основании генетических дистанций, рассчитанных по маркерам, локализованным в генах, кодирующих белки систем репарации ДНК. Расчет генетических дистанций по методу Nei и их визуализация выполнены с помощью приложения GenAIEx 6.503 [9]. ИБС – ишемическая болезнь сердца; АУТ\_ИБС – аутопсийный материал умерших в результате сердечно-сосудистого события; ТБ – туберкулез; ХВГС – хронический вирусный гепатит С; БА – бронхиальная астма; БА\_АГ – бронхиальная астма в сочетании с артериальной гипертензией; Гер – выборка долгожителей; К – средневозрастная популяционная выборка.

ту этих заболеваний. И это вполне возможно, так как белки репарационных систем вовлечены в V(D)J-рекомбинацию и переключение синтеза классов иммуноглобулинов. Результаты ассоциативного анализа, в свою очередь, также подтверждают этот вывод [4, 5, 8].

Таким образом, вклад изученных маркеров в развитие многофакторной патологии очевиден, хотя не все полученные ассоциации можно логично объяснить с точки зрения патогенеза анализируемых заболеваний. Для выяснения возможных механизмов фенотипической реализации установленных ассоциаций нами проведена подробная аннотация регуляторного потенциала изучаемых маркеров. Аннотация выполнена с использованием таких открытых ресурсов, как GTXPortal, VannoPortal, Ensemble, RegulomeDB, Polym pact, UCSC, GnomAD, ENCODE, GeneHancer, EpiMap Epigenomics 2021, HaploReg, GWAS4D, JASPAR, ORegAnno, DisGeNet и OMIM.

### ОБЩАЯ ХАРАКТЕРИСТИКА ИССЛЕДОВАННЫХ МАРКЕРОВ

Маркеры для анализа изначально выбирали с учетом их вероятной функциональной значимости, соответственно они располагались либо в кодирующей (rs560191, rs709816, rs1801516, rs1799977, rs1805321), либо в промоторной (rs473297, rs1805800, rs189037, rs20579) части генов (табл. 1), что предполагает влияние замены на изменения либо структуры белка, либо уровня экспрессии. Согласно рангу по RegulomeDB, наибольшим ре-

**Таблица 1.** Исследованные маркеры: локализация и функциональный класс

SNP ID: нуклеотидная замена	Хромосомная локализация	Ген/локализация в гене/замена <sup>a</sup>	RegulomeDB, <sup>b</sup> ранг/шкала	MAF <sup>c</sup> (min–max)
rs560191: G>C	15q15.3	<i>TP53BP1</i> /ex 9/Asp358Glu	1f/0.554	0.526 (0.231–0.973)
rs473297: T>G	11q21	<i>MRE11</i> /5'UTR <sup>d</sup> (in1/ <i>ANCRD</i> )	1f/0.223	0.539 (0.346–0.652)
rs1805800: C>T	8q21.3	<i>NBN</i> /5'UTR	4/0.609	0.353 (0.167–0.500)
rs709816: G>A	8q21.3	<i>NBN</i> /ex10/Asp399	6/0.288	0.609 (0.319–0.925)
rs189037: G>A	11q22.3	<i>ATM</i> /5'UTR	2a/0.98	0.467 (0.138–0.703)
rs1801516: G>A	11q22.3	<i>ATM</i> /ex37/Asp1853Asn	7/0.184	0.067 (0.0004–0.237)
rs1799977: A>G	3p22.2	<i>MLH1</i> /ex8/Ile219Val	5/0.611	0.170 (0.006–0.360)
rs1805321: G>A	7p22.1	<i>PMS2</i> /ex11/Pro470Ser	5/0.135	0.358 (0.206–0.505)
rs20579: G>A	19q13.33	<i>LIG1</i> /5'UTR	5/0.304	0.173 (0.059–0.343)

<sup>a</sup> Замена аминокислотного остатка в указанной позиции кодируемого белка; in – интрон, ex – экзон.

<sup>b</sup> Функциональный класс по классификации RegulomeDB (<https://regulomedb.org/regulome-search/>): 1f – вариант экспрессионного локуса количественных признаков (eQTL) в мотиве связывания TF или в регионе гиперчувствительности к ДНКазе; 2a – вариант локализован в мотиве связывания TF и изменяет эти мотивы, а также в регионе гиперчувствительности к ДНКазе; 4 – вариант локализован в мотиве связывания TF и в регионе гиперчувствительности к ДНКазе; 5 – вариант локализован в мотиве связывания TF или в регионе гиперчувствительности к ДНКазе; 6 – нарушение мотивов связывания TF; 7 – другое.

<sup>c</sup> Среднепопуляционная частота альтернативного аллеля.

<sup>d</sup> 5'UTR – 5'-нетранслируемая область гена.

гуляторным потенциалом обладают rs560191 и rs473297, которые представляют собой eQTL-варианты, а также расположены в регионах связывания транскрипционных факторов (TF) или в регионах чувствительности к ДНКазе. Наименьший регуляторный потенциал предполагается для rs709816 (изменение мотивов связывания TF) и rs1801516 (ранг – “Другое”).

Вариант rs560191 в гене *TP53BP1* представляет собой миссенс-замену G/C в экзоне 9, приводящую к замене Asp358Glu в кодируемом белке (табл. 1). Средняя частота аллеля C составляет 0.526, варьируя от 0.213 у финнов до 0.973 у афроамериканцев. У африканцев и афроамериканцев в том же геномном положении зарегистрирована также замена G/T, частота аллеля T – около  $6 \times 10^{-5}$  (<http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>). Данный SNP локализован между двумя дистальными энхансерподобными регуляторными элементами: E1757875/enhD и E1757876/enhD (<https://genome.ucsc.edu/>, <https://www.encodeproject.org/>).

В гене *NBN* изучено две нуклеотидные замены: rs1805800 и rs709816 (табл. 1). rs1805800 представ-

ляет собой замену C/T в 5'UTR гена *NBN*. Среднепопуляционная частота аллеля T rs1805800 составляет 0.353, варьируя в пределах от 0.167 (африканские популяции Карибов и Барбадоса) до 0.500 у индийцев-гуджарати (<http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>). Данный маркер находится на расстоянии 395 п.н. от промотора гена *NBN*; входит в состав промотора GN08J089980 и проксимального энхансерподобного элемента E2646296/enhP (<https://genome.ucsc.edu/>, <https://www.encodeproject.org/>, <https://www.genecards.org/>). Замена A/G (rs709816) в экзоне 10 гена *NBN* синонимичная (Asp399). В референсной последовательности описан аллель A, хотя фактически это производный аллель. Среднепопуляционная частота аллеля G составляет 0.609, наименьшая частота зарегистрирована у жителей Англии и Шотландии (0.319), наибольшая – у населения Гамбии (0.925) (<http://www.mulinlab.org/vportal/index.html>, <http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>). Эта сайленс-замена локализована менее чем в 5000 п.н. от дистального энхансерподобного элемента E2646270/enhD и энхансера GN06J089948

(<https://genome.ucsc.edu/>, <https://www.encodeproject.org/>, <https://www.genecards.org/>).

Два SNP изучено также в гене *ATM* (табл. 1). rs189037 представляет собой замену G/A в 5'UTR гена, локализованную в CpG-островке, в 29 п.н. от промотора гена *ATM*, внутри промотора GH11J108219, между промоторподобными элементами E1567620/prom (в 73 п.н.) и E1567621/prom (в 128 п.н.) (<https://genome.ucsc.edu/>, <https://www.encodeproject.org/>, <https://www.genecards.org/>). Среднепопуляционная частота составляет 0.467, варьируя от 0.138 у гамбийцев до 0.703 у населения Пакистана (<http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>). rs1801516 – миссенс-вариант G/A в 37 экзоне, приводящий к замене Asp1853Asn. Этот вариант локализован в мультирегионе взаимодействия *KDELC2/GH11J108219*, расположен на расстоянии 893 п.н. от дистального энхансерподобного элемента E1567660/enhD (<https://genome.ucsc.edu/>, <https://www.encodeproject.org/>, <https://www.genecards.org/>). Среднепопуляционная частота аллеля A составляет 0.0669, варьируя от 0.0004 у японцев до 0.2367 у финнов (<http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>).

rs473297 в гене *MRE11* (табл. 1) представляет собой замену T/G в 5'UTR, локализованную в 79 п.н. от промотора GH11J094492, в 550 п.н. от проксимального энхансерподобного элемента E1561900/enhP, в регионе взаимодействия *MRE11/GH11J094492* (<https://genome.ucsc.edu/>, <https://www.genecards.org/>). Среднепопуляционная частота аллеля G составляет 0.539, варьируя от 0.346 у британцев и шотландцев до 0.652 у ишан из Нигерии. В этой же точке генома у жителей Восточной Азии описана замена T/A, частота аллеля A составляет менее  $3 \times 10^{-4}$  (<http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>).

В экзоне 8 гена *MLH1* изучена несинонимичная замена A/G – rs1799977 (табл. 1), – ведущая к замене Ile219Val в кодируемом белке. Этот вариант локализован в мультирегионе взаимодействия *LRRFIP2/GH03J036988* (<https://genome.ucsc.edu/>, <https://www.genecards.org/>). Среднепопуляционная частота аллеля G составляет 0.170, варьируя от 0.006 (менде, Сьерра-Леоне) до 0.360 (тосканцы, Италия). В редких случаях в этой точке генома регистрируют замену A/T (описана в базе TOPMed) с частотой аллеля T  $1.5 \times 10^{-5}$  (<http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>).

rs1805321 (табл. 1) в гене *PMS2* представляет собой нуклеотидную замену G/A в экзоне 11, приводящую к замене Pro470Ser в кодируемом белке. Этот SNP локализован в 1512 п.н. от дистального энхансерподобного элемента E2531915/enhD (<https://genome.ucsc.edu/>, <https://www.encodeproject.org/>). Среднепопуляционная частота замены – 0.358, варьирует в пределах от 0.206 (у менде,

Сьерра-Леоне) до 0.505 (у британцев Англии и Шотландии) (<http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>).

Замена G/A (rs20579) расположена в 5'UTR гена *LIG1* (табл. 1), в некодирующей части экзона 2, она локализована в мультирегионах взаимодействия *LIG1/GH19J048121* и *PLA2G4C/GH19J048253*; в 158 п.н. от замены заканчивается проксимальный энхансерподобный элемент E1959701/enhP (<https://genome.ucsc.edu/>, <https://www.encodeproject.org/>). Среднепопуляционная частота аллеля A составляет 0.173, варьирует от 0.0594 (у японцев) до 0.343 (у йоруба и ишан в Нигерии) (<http://www.ensembl.org/index.html>, <https://gnomad.broadinstitute.org/>).

Таким образом, все анализируемые нуклеотидные замены локализованы в регионах (или рядом с ними), непосредственно регулирующих транскрипцию генов (промоторов, энхансеров и подобных им элементов), и, следовательно, могут обладать регуляторным потенциалом.

#### МОДИФИКАЦИИ ГИСТОНОВ, КОЛОКАЛИЗОВАННЫЕ С АНАЛИЗИРУЕМЫМИ МАРКЕРАМИ

Известно, что в определении функционального состояния отдельно взятых фрагментов генома важнейшую роль играют посттрансляционные модификации гистонов (“гистоновый код”) [10]. Для анализа гистоновых модификаций в регионах локализации исследуемых SNP были использованы данные репозитория EpiMap Epigenomics 2021, представляющего собой интегральный ресурс, обобщающий эпигеномные карты по 869 биообразцам, отнесенным к 33 категориям тканей (<http://compbio.mit.edu/epimap/#chromatin-states>) [11]. Под биообразцами при этом понимаются как клетки различных тканей (в норме на разных стадиях развития и при патологии), так и различные клеточные линии и их дериваты.

Для кодирующих регионов генов, в которых локализовано пять привлеченных к рассмотрению полиморфных вариантов, отмечается небольшое количество гистоновых модификаций, характерных для активно транскрибируемого хроматина. Так, для всех пяти локусов выявлено триметилирование лизина в позиции 36 гистона H3 (H3K36me3), типичное для открытого хроматина и способствующее элонгации транскрипции (табл. 2) (<http://www.mulinlab.org/vportal/index.html>, <http://compbio.mit.edu/epimap/#chromatin-states>). Другие эпигенетические метки (также характеризующие активную транскрипцию) встречаются редко, в отдельных типах клеток. Так, с регионом гиперчувствительности к ДНКазе колокализованы четыре из исследованных замен в кодирующих участках генов: rs560191 (в быстрорастущем клоне линии LNCaP, звездчатых



клетках печени, гепатоцитах, лимфобластоидной клеточной линии GM19239, линии фибробластов легких AG08396), rs709816 (в эмбриональных клетках желудка), rs1801516 (в гематопоэтических стволовых и В-клетках при миелоидном лейкозе), rs560191 (в быстрорастущем клоне линии LNCaP, звездчатых клетках печени, гепатоцитах, лимфобластоидной клеточной линии GM19239, линии фибробластов легких AG08396), rs1799977 (в клетках аденокарциномы простаты) (<http://www.mulinlab.org/vportal/index.html>, <http://compbio.mit.edu/epimap/#chromatin-states>). Кроме того, rs560191 в гене *TP53BP* колокализован с регионом связывания РНК-полимеразы II (в клетках влагалища), с ацетилированным лизином в позиции 23 гистона H3 – H3K23ac (в клетках трофобласта, дериватах H1-hESC), с H3K79me1 (в мезенхимальных и мезодермальных стволовых клетках, дериватах H1-hESC; в эмбриональных стволовых клетках линии H9; в эмбриональных фибробластах легкого линии IMR-90), с H3K27ac и H3K9ac (в гепатоцитах), H4K20me1 (в эмбриональных клетках, подобных стволовым, – линия H1-hESC) (<http://www.mulinlab.org/vportal/index.html>, <http://compbio.mit.edu/epimap/#chromatin-states>). Нуклеотидная последовательность, содержащая rs709816 в гене *NBN*, колокализована с H3K27ac (большеберцовые нервы), с H3K79me2 (нейробластома) (<http://www.mulinlab.org/vportal/index.html>, <http://compbio.mit.edu/epimap/#chromatin-states>). rs1801516 в гене *ATM* колокализован с H4K20me1 (колоректальная аденокарцинома) (табл. 2) (<http://www.mulinlab.org/vportal/index.html>, <http://compbio.mit.edu/epimap/#chromatin-states>).

Иной паттерн модификаций гистонов колокализуется с маркерами в 5'UTR генов (табл. 2). Каждый из этих SNP колокализован с пятью модификациями хроматина: регионом гиперчувствительности к ДНКазе, H3K27ac и H3K4me2 (характеризующими энхансерные последовательности), H3K4me3 и H3K9ac (способствующими активации транскрипции). Три маркера колокализованы с модификацией H3K4me1 (также характерной для энхансерных последовательностей): rs473297 в 141, rs1805800 в 255 биообразцах и rs20579 в эпителиальных клетках меланомы и нейроэпителиоме. С модификацией, поддерживающей активацию транскрипции и элонгацию, H3K79me2, колокализованы следующие маркеры: rs473297 в 732, rs20579 в 235 и rs189037 в 6 биообразцах (в различных субпопуляциях Т-лимфоцитов, а также в Т-лимфоцитах при остром лимфобластном лейкозе и в лимфобластоидной клеточной линии). Модификация H3K36me3, поддерживающая элонгацию транскрипции, колокализуется с rs20579 (в 48 образцах) и rs189037 (в Т-лимфоцитах, гладкомышечных клетках двенадцатиперстной кишки, фибробластах крайней плоти).

Два локуса (rs1805800 и rs189037) находятся в открытом хроматине с сайтами связывания таких белков, как CTCF, SMC3, EP300, H2AFZ. rs189037, кроме этого, колокализован с сайтом связывания РНК-полимеразы II и RAD21. Известно, что транскрипционный репрессор CTCF связывается с инсуляторами генов [12]. Кроме того, у млекопитающих CTCF совместно с когезиновым комплексом (в который входят в том числе SMC3 и RAD21) и гистоном H2AFZ принимает участие в организации хромосом в топологически ассоциированные домены [13, 14]. Следовательно, нуклеотидные последовательности, в которых находятся rs1805800 и rs189037, имеют важное топологическое значение и любые изменения в их структуре могут способствовать развитию патологических состояний.

Два маркера колокализованы с H3K79me1 (активация элонгации): rs20579 (в эмбриональных фибробластах легкого, дериватах линии H1-hESC – мезенхимальных и мезодермальных стволовых клетках), rs189037 (в клетках трофобласта – дериваты из H1-hESC). С монометилированным лизином в положении 20 гистона H4 (H4K20me1), также поддерживающим активацию транскрипции [15], колокализованы rs473297 (в двух клеточных линиях Т-лимфоцитов при остром лимфобластном лейкозе) и rs20579 (в Т-лимфоцитах при остром лимфобластном лейкозе и в эндотелиальных клетках пупочной вены новорожденного).

rs189037 в гене *ATM* в различных эмбриональных тканях/дериватах колокализован с целым рядом гистоновых модификаций, также поддерживающих активацию транскрипции: H2AK9ac, H2BK120ac, H2BK12ac, H2BK15ac, H3K23ac, H3K4ac, H2BK5ac, H3K18ac, H3K56ac, H4K12ac, H4K8ac и H4K91ac. Такое множество модификаций в единичных эмбриональных клеточных линиях может свидетельствовать о важной роли данного локуса на ранних этапах развития.

Все перечисленные выше модификации гистонов запускают/усиливают/поддерживают транскрипцию. И лишь две из колокализованных с анализируемыми маркерами модификации приводят к транскрипционному сайленсингу – это H3K27me3 [16] и H3K23me2 [17]. Среди изученных маркеров с H3K27me3 колокализован rs189037 в клетках зародышевого матрикса; а с H3K23me2 – rs20579 в эмбриональных стволовых клетках линии H1-hESC. Вероятно, более тонкая регуляция функций этих локусов в эмбриональном развитии может играть важную роль.

Таким образом, с учетом характера модификаций гистонов, колокализованных с привлеченными к анализу SNP генов систем репарации ДНК, можно заключить, что, во-первых, все эти регионы в большинстве изученных тканей активно транскрибируются; во-вторых, два региона

(колокализованные с rs1805800 и rs189037) участвуют в организации топологических доменов, причем один из них (колокализованный с rs180937), вероятно, играет важную роль в эмбриональном развитии.

### ЗАВИСИМОСТЬ ЭКСПРЕССИОННОГО СТАТУСА ГЕНОВ ОТ АНАЛИЗИРУЕМЫХ ПОЛИМОРФНЫХ ВАРИАНТОВ

Репарация ДНК — один из базовых процессов живой клетки, а экспрессия генов, белковые продукты которых задействованы в его реализации и контроле, идет во всех типах тканей на всех стадиях онтогенеза. Однако, в отличие от генов “домашнего хозяйства”, уровень экспрессии генов, кодирующих белки систем репарации ДНК варьирует в зависимости от потребности клетки; и чем шире сфера компетенции этих генов, тем выше наблюдаемый уровень экспрессии. Для всех рассматриваемых здесь генов минимальный уровень экспрессии составлял от 1–2 TPM (Transcripts Per Million — транскриптов на миллион): *ATM* в скелетных мышцах, *MRE11* в тканях головного мозга, *TP53BP1* в цельной крови, *PMS2* в миокарде, *LIG1* в левом желудочке — до 5–6 TPM (*NBN* в коре почек, *MLH1* в цельной крови) (<https://gtexportal.org/home/>). Интересно, что максимальный уровень этих транскриптов в большинстве случаев детектировали в лимфоцитах, стимулированных вирусом Эпштейна–Барр (*ATM* — 21, *MRE11* — 22, *NBN* — 67, *MLH1* — 48, *PMS2* — 22, *LIG1* — 51 TPM), и только для гена *TP53BP1* максимальный уровень экспрессии регистрировали в гипофизе — 50 TPM (<https://gtexportal.org/home/>). Все вышеперечисленные маркеры (табл. 1) относятся к cis-eQTL-вариантам (то есть могут влиять на экспрессию “своего” и расположенных рядом генов) и большая часть — к sQTL-вариантами (вливают на сплайсинг).

В подавляющем большинстве случаев изменение экспрессии подчиняется линейной зависимости: либо  $RR > RP > PP$ , либо  $RR < RP < PP$  (где R — референсный аллель, P — производный аллель), — поэтому далее мы преимущественно обсудим различия между экспрессией гомозиготных генотипов, подразумевая, что гетерозиготы занимают промежуточное положение. Те случаи, когда наблюдаются отклонения от указанной тенденции, будут рассмотрены особо.

#### Влияние rs560191 в гене *TP53BP1* на экспрессию и сплайсинг

rs560191 (*TP53BP1*) находится в пределах большого блока корегулируемых генов (рис. 2). Уровень сцепления, близкий к единице, выявлен для SNP, расположенных вокруг rs560191 и охватывающих регион больше 87 млн.п.н. Этот блок сцеп-

ления включает маркеры, которые представляют собой eQTL для 20 генов и sQTL для 11 генов (<http://www.mulinlab.org/vportal/index.html>).

Для большинства генов, регулируемых rs560191 (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_gus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_gus.zip)), характерно однонаправленное изменение экспрессии в различных тканях в зависимости от генотипа. Так, для *AC011330.13* (4 ткани), *AC011330.5* (40 тканей), *ADAL* (37 тканей), *CATSPER2* (1 ткань), *CCNDBP1* (4 ткани), *CKMT1A* (4 ткани), *CKMT1B* (1 ткань), *PDIA3* (3 ткани), *TP53BP1* (15 тканей), *TTBK2* (2 ткани) у гомозиготных носителей анцестрального аллеля (GG) уровень экспрессии выше, чем у гомозигот по производному аллелю (CC), а для гетерозигот характерны промежуточные значения. В то же время обратная ситуация ( $GG < CC$ ) наблюдается для генов *CATSPER2P1* (14 тканей), *MAPIA* (2 ткани), *RNU6-554P* (1 ткань), *STRC* (27 тканей), *STRCP1* (25 тканей), *TGM7* (3 ткани), *TUBGCP4* (2 ткани). Для трех генов (*LCMT2*, *TGM5*, *ZSCAN29*) отмечается межтканевая вариабельность характера экспрессии. Так, максимальный уровень экспрессии гена *LCMT2* в слизистой оболочке пищевода наблюдается у гомозигот CC, в то время как в четырех других тканях (большеберцовые нервы, культуры фибробластов, семенники, гладкомышечные клетки пищевода) — у гомозигот GG. В трех тканях (в слизистой оболочке и гладкомышечных клетках пищевода, ободочной кишке) характер экспрессии гена *TGM5* изменяется в направлении  $GG > GC > CC$ , а в трансформированных вирусом Эпштейна–Барр лимфоцитах — в обратном направлении ( $GG < GC < CC$ ). Экспрессия гена *ZSCAN29* выше у гомозигот по предковому аллелю в 25 исследованных тканях, кроме ткани мозга (хвостатое и прилежащее ядро базальных ганглиев, фронтальная кора, передняя поясная кора), в которых экспрессия этого гена наиболее выражена у носителей производного аллеля в гомозиготном состоянии (<https://gtexportal.org/home/>).

Показано влияние генотипов rs560191 на сплайсинг 11 генов, из них 10 совпадают с генами, для которых этот полиморфный вариант является eQTL, еще один ген — *PPIP5K1* — не входит в этот список. Данный регион характеризуется сложной регуляцией транскрипции. В частности, происходит образование общего транскрипта при считывании генов *CATSPER2P1*, *AC011330.5*, *CATSPER2*. Для них показаны общие эффекты сплайсинга: производный аллель в дозозависимой манере увеличивает эффективность вырезания интрона 11 гена *CATSPER2*, что оказывает влияние на сплайсинг мРНК генов *AC011330.5* (в 20 тканях) и *CATSPER2* (в 37 тканях). Эффективность вырезания интрона 10 гена *CATSPER2* из общего транскрипта снижается в скелетных мышцах, шейном

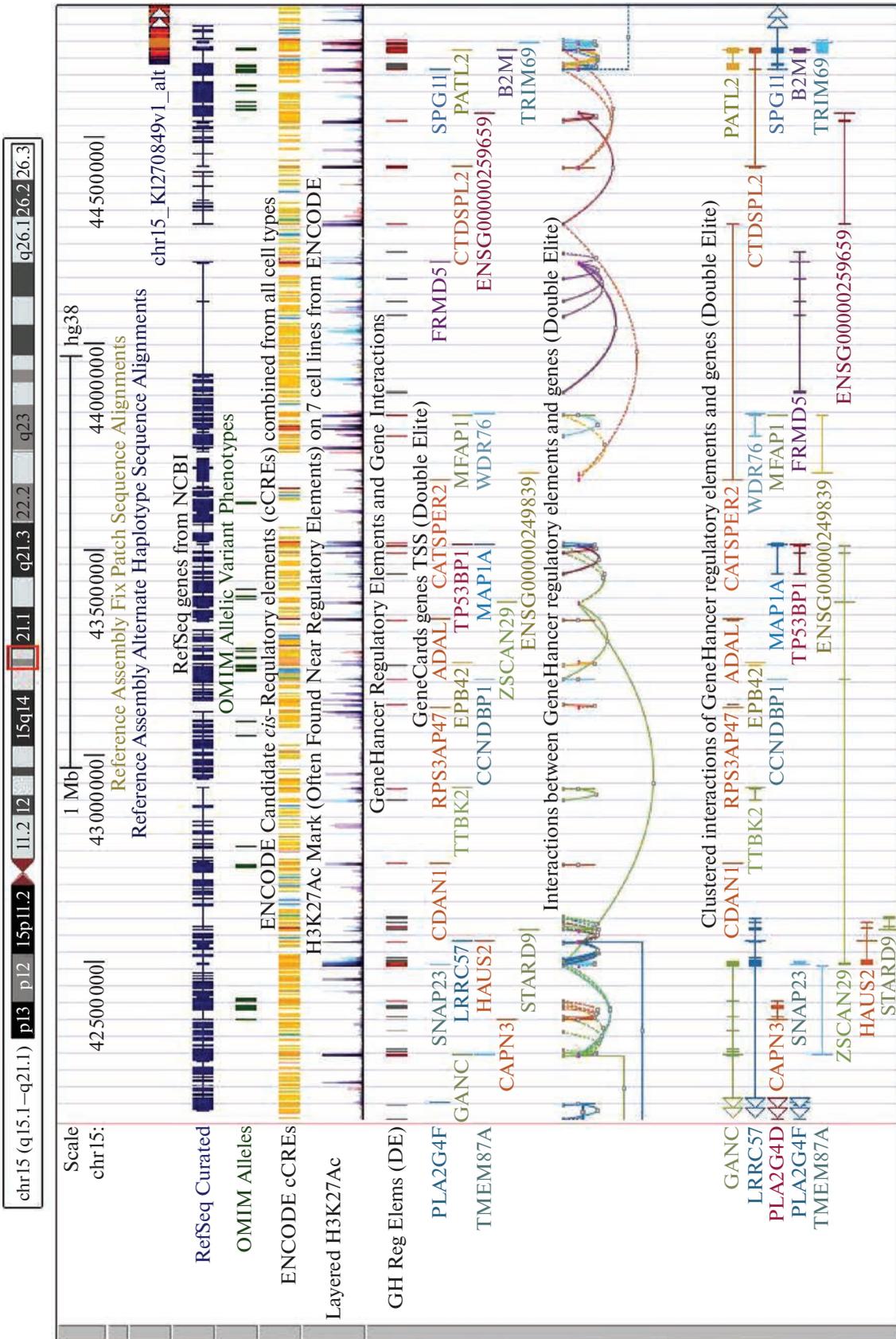


Рис. 2. Взаимодействия между регуляторными элементами и генами (согласно GeneHancer) в регионе локализации гена *TP53BP1*. Разными цветами выделены отдельные гены и их регуляторные взаимодействия (воспроизведено из онлайн-браузера UCSC: <https://genome.ucsc.edu/>).

отделе спинного мозга, мозжечке и черной субстанции головного мозга (оказывая влияние на сплайсинг мРНК гена *CATSPER2*) и повышается в семенниках (влияя на сплайсинг *CATSPER2P1* и *CATSPER2*). Повышение эффективности вырезания области, включающей экзон 1—интрон 4 гена *CATSPER2*, влияет на сплайсинг мРНК гена *AC011330.5* (в мозжечке). Кроме этого, носительство производного аллеля rs560191 приводит к снижению эффективности вырезания отдельных интронов в генах *AC011330.5*, *CATSPER2*, *PP1P5K1*, *STRCP1*, *TGM5*, *TGM7*, *TP53BP1* и *ADAL* (показано в одной—трех тканях для каждого транскрипта).

Повышение эффективности вырезания показано для экзона 2 (с прилегающими интронами) гена *ADAL* (в 9 тканях), экзона 19 (с прилегающими интронами) гена *STRC* (в мозжечке), интрона 2 гена *ZSCAN29* (в сигмовидной кишке) (<https://gtexportal.org/home/>).

Соотношение изоформ белка, синтезирующихся в результате альтернативного сплайсинга, зависит в том числе от того, насколько эффективно будет происходить вырезание тех или иных последовательностей из пре-мРНК. В случае, когда зарегистрирована зависимость и уровня экспрессии, и сплайсинга от одного и того же варианта одного и того же гена в одной и той же ткани, можно проанализировать, каким образом меняется синтез зрелой мРНК в зависимости от анализируемой нуклеотидной замены. Можно предположить, что при однонаправленном изменении, например при усилении экспрессии в направлении  $RR < RP < PP$  и повышении эффективности вырезания какого-либо фрагмента мРНК, общее увеличение продукции гена происходит, главным образом, за счет того/тех вариантов, в которых вырезаемый фрагмент отсутствует. И напротив, если эффективность снижается (в соответствии с примером выше, это  $RR > RP > PP$ ), то усиление экспрессии происходит за счет относительного увеличения частоты тех вариантов, в которых этот фрагмент остается.

Анализ данных по влиянию rs560191 на экспрессию и сплайсинг показывает, что повышение эффективности вырезания интрона 11 гена *CATSPER2* сопряжено со снижением уровня транскрипции генов *CATSPER2* (в слизистой оболочке желудка) и *AC011330.5* (в 21 ткани). Снижение экспрессии *AC011330.5* в мозжечке идет на фоне увеличения эффективности вырезания экзона 1—интрона 4 гена *CATSPER2*, а в семенниках — интрона 25 гена *AC011330.5*. Снижение экспрессии гена *ADAL* происходит на фоне повышения эффективности вырезания экзона 2 с прилегающими интронами (в 7 тканях) и понижения эффективности вырезания интрона 2 (в гипофизе). Общее снижение экспрессии генов *TGM5* и *TP53BP1* сопряжено со снижением эффективности вырезания их интро-

нов (9 и 23 соответственно). Снижение эффективности вырезания интрона 11 в гене *TGM7* и интрона 24 гена *STRCP1*, а также повышение эффективности вырезания экзона 19 с прилегающими интронами в гене *STRC* сопряжены с увеличением уровня экспрессии соответствующих генов в отдельных тканях (<https://gtexportal.org/home/>).

#### *Влияние rs1805800 и rs709816 в гене NBN на экспрессию и сплайсинг*

Изученные маркеры в гене *NBN* находятся на расстоянии 29 628 п.н.: rs1805800 локализован в регионе рядом с 5'UTR гена *NBN*, rs709816 — в экзоне 10. Вместе с тем, они достаточно тесно сцеплены. Так, для популяции г. Томска значения показателей сцепления: коэффициента неравновесия по сцеплению ( $D'$ ), логарифма отношения шансов ( $LOD$ -score) и коэффициента корреляции Пирсона ( $r^2$ ) — равны соответственно 0.942, 85.11 и 0.712; по данным проекта “1000 Genomes” у европеоидов  $D' = 0.995$ ,  $r^2 = 0.846$ . Исходя из данных проекта “1000 Genomes” по сцеплению маркеров, SNP, расположенные в пределах региона размером 142 т.п.н., включающего полностью ген *NBN*, сцеплены достаточно тесно ( $D'$  в пределах 0.945–1.00) и образуют единый регуляторный блок eQTL- и sQTL-вариантов (<http://www.mullinlab.org/vportal/index.html>, <https://genome.ucsc.edu/>).

По данным ресурса GTExPortal (<https://gtexportal.org/home/>), указанные нуклеотидные замены относятся к eQTL-вариантам, влияющим на экспрессию как “своего”, так и близкорасположенных генов: *CALB1*, *DECR1*, *OSGIN2* — и, кроме того, на уровень экспрессии гена *RP11-662G23.1*, локализованного более чем в 821 т.п.н. от промотора гена *NBN* (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_rus.zip)) (<https://gtexportal.org/home/>, <https://genome.ucsc.edu/>).

Для большинства изученных тканей характерно однонаправленное изменение экспрессии ряда генов в зависимости от генотипов по rs1805800. Так, при увеличении дозы альтернативного аллеля (Т) возрастает уровень экспрессии гена *CALB1* (в 6 из 6 изученных тканей), *NBN* (в 15 из 18 тканей), *OSGIN2* (в трех тканях из четырех изученных). Из общей тенденции есть несколько исключений. Для *RP11-662G23.1* в единственной изученной ткани (мышечной ткани пищевода) отмечается сходная в отношении гомозиготных генотипов тенденция (ТТ>СС), однако минимальный уровень экспрессии зарегистрирован у носителей гетерозиготного генотипа. Самая высокая экспрессия наблюдается у гомозигот по референсному аллелю в четырех тканях из пяти изученных для гена *DECR1* (за исключением скелет-

ных мышц, в которых наблюдается обратная зависимость). Для гена *NBN* к исключениям относятся цельная кровь и ткань пищеводно-желудочного соединения, в которых наиболее высокий уровень экспрессии детектируют у гомозигот с референсным генотипом (CC), а также кора головного мозга, где, при общей тенденции CC<TT, уровень экспрессии максимален у гетерозигот. Для ткани пищеводно-желудочного соединения показаны также отличия в экспрессии гена *OSGIN2*: у гомозигот по производному аллелю уровень экспрессии ниже, чем по анцестральному (<https://gtexportal.org/home/>).

В зависимости от генотипов по маркеру rs709816 в экзоне 10 гена *NBN* меняется уровень экспрессии тех же генов, хотя характер изменений несколько иной. У гомозигот по альтернативному аллелю (AA) экспрессия гена *NBN* выше в 14 из 14 изученных тканей (как показано и для rs1805800), однако более низкий уровень экспрессии зарегистрирован для генов *CALB1* (в 7 изученных тканях) и *RP11-662G23.1* (в мышечной ткани пищевода). Ген *DECRI* экспрессируется интенсивнее у носителей производного аллеля в четырех изученных тканях, еще в двух (в скелетных мышцах и ушке предсердия) зависимость экспрессии обратная. Уровень экспрессии гена *OSGIN2* в мышечной ткани пищевода выше у гомозигот по альтернативному аллелю, тогда как в цельной крови и ткани пищеводно-желудочного соединения – у гомозигот по референсному аллелю (<https://gtexportal.org/home/>).

Анализируемые маркеры в гене *NBN* относятся к sQTL-вариантам для генов *DECRI* и *NBN*. При процессинге мРНК гена *DECRI* эффективность вырезания интрона 1 выше у носителей производных аллелей обоих изученных маркеров: для генотипа TT rs1805800 в легких и коже, не подвергающейся солнечной экспозиции; для генотипа GG rs709816 – в легких, коже, не подвергающейся солнечной экспозиции, и тканях молочной железы (<https://gtexportal.org/home/>).

Для гена *NBN* отмечено тканезависимое изменение соотношений вариантов сплайсинга в зависимости от генотипов по изученным маркерам. У носителей гомозиготных генотипов обоих исследованных маркеров повышается эффективность вырезания интрона 4 в семенниках, скелетных мышцах, большеберцовых артериях, ткани пищеводно-желудочного соединения, ткани молочной железы и снижается в большеберцовых нервах и в лимфоцитах, трансформированных вирусом Эпштейна–Барр. Аналогично снижается эффективность вырезания интрона 2 в культурах фибробластов и в подкожной жировой клетчатке (<https://gtexportal.org/home/>).

В базах данных для пяти тканей приведена информация по изменениям и уровням экспрессии

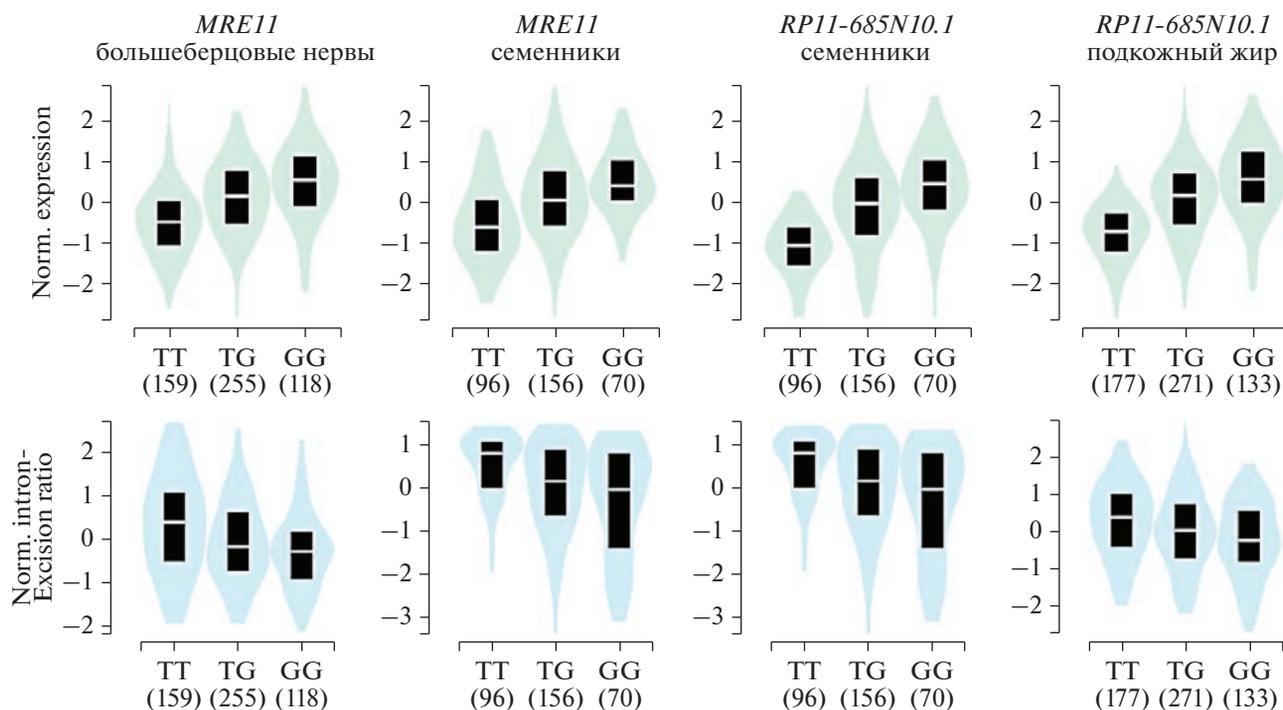
гена *NBN*, а также вариантам его сплайсинга (в зависимости от изученных SNP). В отношении обоих маркеров (rs1805800 и rs709816) выявленная изменчивость носит одинаковый характер: в трех тканях носительство альтернативных аллелей приводит к увеличению экспрессии гена в целом при увеличении эффективности вырезания интрона 4 (в семенниках, скелетных мышцах, большеберцовых артериях); в двух тканях общее увеличение уровня экспрессии происходит на фоне снижения эффективности вырезания интрона 4 (в большеберцовых нервах) и интрона 2 (в подкожной жировой клетчатке) (<https://gtexportal.org/home/>).

#### *Влияние rs189037 и rs1801516 в гене ATM на экспрессию и сплайсинг*

Исследованные маркеры в гене *ATM* в популяции г. Томска находятся в неравновесии по сцеплению. Так, показатели сцепления для rs189037 и rs1801516 в гене *ATM* следующие:  $D' = 0.869$ ,  $LOD = 11.44$ ,  $r^2 = 0.109$ . В научных публикациях и базах данных (см., например, Ensemble: <http://www.ensembl.org/index.html>) их сцепление не анализируется ввиду значительной удаленности друг от друга – изученные маркеры rs189037 (промоторный регион) и rs1801516 (экзон 37) находятся на расстоянии 81.5 т.п.н. Тем не менее для каждого из анализируемых SNP идентифицирован обширный регион, маркеры в котором тесно сцеплены и являются eQTL для одних и тех же генов. Для rs189037 размер этого региона не менее 167 т.п.н., для rs1801516 – не менее 192 т.п.н. (<http://www.mulinlab.org/vportal/index.html>).

rs189037 представляет собой eQTL-вариант для “своего” и близлежащих генов: *ACAT1*, *NPAT*, *C11orf65*, *KDELC2* (*POGLUT3*) (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_rus.zip)). В большинстве случаев альтернативный аллель (A) приводит к снижению уровня экспрессии корегулируемых генов. Исключения отмечены в отдельных тканях для генов *NPAT* (в цельной крови) и *ATM* (в щитовидной железе), а также для *C11orf65* в коре надпочечников (единственная ткань, для которой в настоящее время показано влияние rs189037 на уровень экспрессии рассматриваемого гена). Кроме того, для гена *ATM* в трех тканях (ушке предсердия, сальнике и слизистой оболочке пищевода) наименьший уровень экспрессии зарегистрирован у гетерозигот (при сохранении общей тенденции: GG>AA) (<https://gtexportal.org/home/>).

rs1801516 также выступает как eQTL-вариант для “своего” и близлежащих генов: *ACAT1*, *NPAT*, *C11orf65* (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_rus.zip)). У гомозигот по аль-



**Рис. 3.** Изменение уровня экспрессии (вверху) и эффективности сплайсинга (внизу) генов *MRE11* (в большеберцовых нервах и семенниках) и *RP11-685N10.1* (в семенниках и подкожной жировой ткани) в зависимости от rs473297 в гене *MRE11*. Ось X – генотипы по rs473297, ось Y – нормализованные уровни экспрессии (вверху), нормализованный уровень вырезания интрона (по данным GTXPortal: <https://gtexportal.org/home/>).

тернативному аллелю (AA) регистрируют пониженные уровни экспрессии генов *ACAT1* (в щитовидной железе, скелетных мышцах, легких, культурах фибробластов) и *ATM* (мышечная ткань пищевода) и повышенные для генов *NPAT* (большеберцовые нервы и аорта) и *C11orf65* (в гипофизе, щитовидной железе и коже, подвергающейся солнечной экспозиции) (<https://gtexportal.org/home/>).

К sQTL относится только rs189037, причем только для гена *ATM*: в культуре фибробластов происходит дозозависимое снижение эффективности вырезания интрона 40, в то время как в легких, поджелудочной железе и сальнике эффективность вырезания повышается, как и в отношении интрона 26 в большеберцовых нервах (<https://gtexportal.org/home/>).

Для двух тканей можно соотнести eQTL- и sQTL-влияние rs189037 на синтез зрелой мРНК белка ATM. Общее снижение уровня экспрессии гена *ATM* в культуре фибробластов происходит параллельно с увеличением эффективности вырезания интрона 40, а в клетках сальника – на фоне снижения эффективности вырезания этого интрона (<https://gtexportal.org/home/>).

#### Влияние rs473297 в гене *MRE11* на экспрессию и сплайсинг

rs473297 находится в большом блоке сцепления (не менее 123 т.п.н.), в пределах которого D'

между анализируемым маркером и другими eQTL-SNP составляет 0.95–1.00. rs473297 служит eQTL-вариантом для генов *MRE11* (49 тканей), *GPR83* (4 ткани), *IZUMO1R* (1 ткань) и *RP11-685N10.1* (37 тканей) (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_rus.zip)). Во всех изученных тканях наличие альтернативного варианта приводит к возрастанию уровня экспрессии генов *IZUMO1R*, *MRE11*, *RP11-685N10.1* и к снижению экспрессии гена *GPR83* (<https://gtexportal.org/home/>). Влияние на уровень сплайсинга показано для генов *MRE11* (семенники и большеберцовые нервы) и *RP11-685N10.1* (8 тканей). Наличие замены приводит к снижению эффективности вырезания интрона 7, внутри которого локализован ген *RP11-685N10.1*. В большеберцовых нервах происходит также снижение эффективности вырезания интрона 1 мРНК гена *MRE11* (рис. 3) (<https://gtexportal.org/home/>).

#### Влияние rs1799977 в гене *MLH1* на экспрессию и сплайсинг

Для rs1799977 в гене *MLH1* показано полное сцепление (и участие в регуляции одних и тех же генов) с маркерами, находящимися от него на значительном удалении – более 78 т.п.н. в направлении к 3'-концу гена. Тесное сцепление ( $D' = 0.912$ ) показано и с маркером, лежащим на

расстоянии более 80 т.п.н. по направлению к 5'-концу (интронный вариант гена *TRANK1*); в то же время этот маркер фактически относится к другому регуляторному блоку (<http://www.mulinlab.org/vportal/index.html>, <https://genome.ucsc.edu/>). По данным проектов ENCODE и GeneHancer, весь этот регион обогащен регуляторными последовательностями, наблюдается значительная ко-регуляция колокализованных генов (<https://genome.ucsc.edu/>, <https://www.genecards.org/>). rs1799977 – eQTL-вариант для десяти и sQTL для трех генов (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_rus.zip)). Для пяти генов есть информация об изменениях уровней экспрессии в зависимости от данной замены в одной ткани. Показано, что у гомозигот по вариантному аллелю уровни экспрессии *MLH1* (в скелетных мышцах), *EPM2AIP1* (в щитовидной железе), *RP11-259K5.2* (в коже, подвергавшейся солнечной экспозиции) повышены, а *ITGA9* (в цельной крови) и *PRADCIP1* (в большеберцовых нервах) снижены. Для гена *RP11-129K12.1* показано снижение уровня экспрессии в трех исследованных тканях (в подкожной жировой клетчатке, сальнике, коже, подвергающейся солнечной экспозиции). Для остальных генов зарегистрирован тканезависимый характер изменения экспрессии. Так, при наличии производного аллеля (G) выявлен более низкий уровень экспрессии генов *RP11-285J16.1* и *UBE2FP1* (в коже, подвергающейся солнечной экспозиции), а также *GOLGA4* (в слизистой оболочке пищевода) и *LRRFIP2* (в коже вне зависимости от солнечной экспозиции и в цельной крови). Напротив, у носителей генотипа GG выше уровень экспрессии генов *RP11-285J16.1* и *UBE2FP1* (в щитовидной железе), *GOLGA4* (в скелетных мышцах и коже, подвергающейся солнечной экспозиции), *LRRFIP2* (в 14 тканях) (<https://gtexportal.org/home/>).

Влияние rs1799977 на сплайсинг показано для трех генов. В подавляющем большинстве случаев замена приводит к снижению эффективности вырезания интронов: интрон 12 в гене *MLH1* (в большеберцовых артериях), интрон 1 в *LRRFIP2* (в 31 ткани), интрон 22 в *GOLGA4* (в слизистой оболочке пищевода). Повышение эффективности вырезания показано только для интрона 2 гена *GOLGA4* (в скелетных мышцах) и интрона 1 гена *LRRFIP2* (в левом желудочке) (<https://gtexportal.org/home/>). Несмотря на сходный тип “сплайсингового ответа” на альтернативный аллель, в результате тканеспецифичного изменения характера экспрессии *LRRFIP2* имеет место тканеспецифичное изменение соотношений сплайсинговых вариантов этого гена. Так, на фоне снижения эффективности вырезания интрона 1 из мРНК гена *LRRFIP2* в трех тканях (кожа вне зависимости от солнечной экспозиции и цельная кровь) проис-

ходит снижение экспрессии этого гена, а в 11 тканях – усиление (<https://gtexportal.org/home/>).

#### Влияние rs1805321 в гене *PMS2* на экспрессию и сплайсинг

rs1805321 в гене *PMS2* находится в блоке сцепления eQTL-SNP, охватывающем не менее 76 т.п.н. Эта замена – eQTL-вариант для 7 генов (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_rus.zip)). В подавляющем большинстве случаев альтернативный аллель (A) приводит к возрастанию уровня экспрессии. Такой характер изменчивости показан для генов *ANKRD61* (в 6 тканях), *CCZ1B* (в 11 тканях), *EIF2AK1* (в 17 тканях), *PMS2* (в 28 тканях), *SNORA42* (в трех тканях). В ряде тканей уровень экспрессии этих генов носит несколько иной характер: при сохранении общей тенденции (GG<AA) максимальный уровень наблюдается у гетерозигот. Такая зависимость зарегистрирована для гена *ANKRD61* в большеберцовых нервах, для *EIF2AK1* в адипоцитах сальника, для *PMS2* в коже, не подвергавшейся солнечной экспозиции, и в базальных ганглиях головного мозга; для *SNORA42* во фронтальной коре головного мозга. И, напротив, показано снижение экспрессии при наличии производного аллеля для гена *CCZ1* (в 8 тканях). Для гена *RAC1* влияние данной замены выявлено в одной ткани (щитовидная железа), уровень максимален у гомозигот по референсному аллелю, но минимален у гетерозигот (<https://gtexportal.org/home/>).

Влияние rs1805321 на сплайсинг показано для одного гена – *AIMP2* – в двух тканях: эффективность вырезания интрона 1 при наличии производного аллеля снижается в скелетных мышцах и возрастает в коже, подвергающейся солнечной экспозиции (<https://gtexportal.org/home/>).

#### Влияние rs20579 в гене *LIG1* на экспрессию и сплайсинг

rs20579 в гене *LIG1* находится внутри региона сцепления eQTL-SNP, охватывающего около 28 т.п.н. Это eQTL-вариант для 6 генов (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_rus.zip)). Производный аллель приводит к возрастанию уровня экспрессии генов *PLA2G4C* (в 22 тканях), *PLA2G4C-AS1* (в легких), *CTC-453G23.5* (в коже, подвергавшейся солнечной экспозиции). Понижение экспрессии наблюдается для генов *LIG1* (в большеберцовых нервах) и *AC022154.7* (в скелетных мышцах). Тканезависимое изменение уровня экспрессии зарегистрировано для гена *CARD8*: у носителей производного аллеля в путамене (базальные ганглии головного мозга) наблюдается повышение уров-

ня экспрессии, а в прилежащем ядре (базальные ганглии головного мозга), напротив, снижение (<https://gtexportal.org/home/>). Влияния rs20579 на сплайсинг к настоящему времени не показано.

Таким образом, все рассматриваемые маркеры изменяют транскрипцию как своего, так и близлежащих генов, будучи для них *cis*-QTL-вариантами. Тканезависимый характер изменений может определять различия степени поражения тканей и органов при развитии патологического фенотипа; при этом поражения органов-мишеней могут быть обусловлены изменениями (как уровня экспрессии в целом, так и альтернативного сплайсинга) спектра корегулируемых генов. Например, ген *ACAT1*, корегулируемый и локализованный с геном *ATM* (табл. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Бабушкина\\_gus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Бабушкина_gus.zip)), кодирует митохондриальную ацетил-КоА-ацетилтрансферазу – один из ключевых ферментов этерификации холестерина. Следовательно, вовлеченность этого фермента в развитие атеросклероза и его патогенетически значимых признаков логически обоснована; блокаторы этого фермента используют для лечения атеросклероза [18]. Одним из возможных объяснений полученных нами ранее ассоциаций гена *ATM* с липидными показателями у больных ИБС [7] может быть именно изменение уровня экспрессии *ACAT1* (и уже его влияние на развитие патологии) в результате нуклеотидных замен в гене *ATM*.

#### ИЗМЕНЕНИЯ МОТИВОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ, ПРОИСХОДЯЩИЕ В РЕЗУЛЬТАТЕ АНАЛИЗИРУЕМЫХ НУКЛЕОТИДНЫХ ЗАМЕН

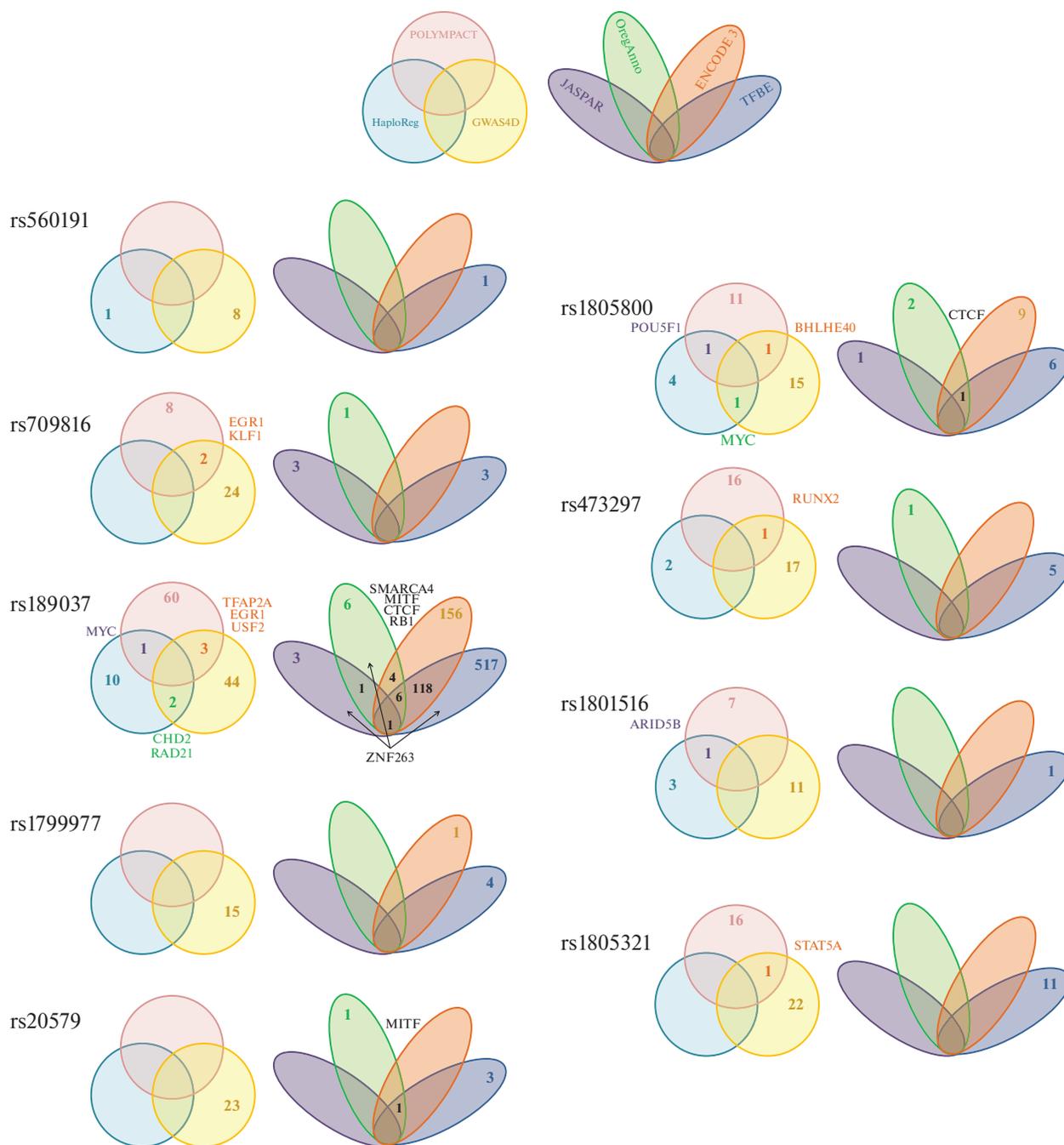
В результате изменения нуклеотидной последовательности могут меняться сайты связывания TF, что позволяет определять статус нуклеотидных замен в качестве *cis*-eQTL-SNP. В настоящее время накоплена обширная информация о взаимодействии хроматина и различных TF. Имеющиеся данные получены как экспериментальным путем (методами иммунопреципитации хроматина, например ChIP-Seq), так и биоинформатическими методами. Существует множество инструментов для биоинформатического анализа, позволяющих выявить в нуклеотидной последовательности вероятные мотивы связывания TF и предсказать, каким образом будет меняться их аффинность в результате всевозможных структурных изменений. Однако следует учитывать, что различные инструменты анализа не всегда дают согласующуюся информацию.

Мы проанализировали изменения мотивов связывания TF с помощью ресурсов HaploReg (<https://pubs.broadinstitute.org/mammals/haploreg/>

[haploreg.php](http://haploreg.php)), Polympact (<https://regulomedb.org/regulome-search/>), GWAS4D ([http://mulinlab.tmu.edu.cn/gwas4d/gwas4d/gwas4d/gwas4d\\_server](http://mulinlab.tmu.edu.cn/gwas4d/gwas4d/gwas4d/gwas4d_server)) (рис. 4, круговые диаграммы Венна). Наименьшее число изменяющихся мотивов предсказывает HaploReg (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>). Согласно полученным данным, в результате проанализированных вариантов меняются сайты связывания TF в генах *ATM* (rs189037, rs1801516), *TP53BP1* (rs560191), *NBN* (rs1805800). Ресурс Polympact не определяет изменений мотивов в результате замены в гене *TP53BP1*, но показывает их наличие в генах *ATM*, *NBN*, *MRE11*, *PMS2* (рис. 4). Согласно GWAS4D, все изученные нуклеотидные замены влияют на аффинность связывания TF с их сайтами. Так, rs560191 в гене *TP53BP1* изменяет только 8 сайтов, а rs189037 в гене *ATM* – 33; всего для 9 изученных SNP меняется 168 мотивов для 99 TF (<https://bcglab.cibio.unitn.it/polympact/>, <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>, [http://mulinlab.tmu.edu.cn/gwas4d/gwas4d/gwas4d/gwas4d\\_server](http://mulinlab.tmu.edu.cn/gwas4d/gwas4d/gwas4d/gwas4d_server)).

Для перекрывающихся мотивов все алгоритмы дают согласующуюся информацию по изменению аффинности связывания TF при изменении нуклеотидной последовательности, хотя перекрывание полученных списков в целом довольно слабое. Так, и HaploReg, и GWAS4D предсказывают изменения аффинности сайтов связывания MYC (rs1805800 в гене *NBN*), а также CHD2 и RAD21 (rs189037 в гене *ATM*). HaploReg и Polympact дают согласующиеся результаты по отношению к POU5F1 (rs1805800, ген *NBN*), MYC (rs189037, ген *ATM*) и ARID5B (rs1801516, ген *ATM*). Наибольшее число совпадающих результатов получено между данными GWAS4D и Polympact: rs1805800 (*NBN*) приводит к изменению сайта связывания BHLHE40; rs709816 (*NBN*) – EGR1 и KLF1; rs473297 (*MRE11*) – RUNX2; rs189037 (*ATM*) – USF2, EGR1, TFAP2A; rs1805321 (*PMS2*) – STAT5A (<https://regulomedb.org/regulome-search/>, <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>, [http://mulinlab.tmu.edu.cn/gwas4d/gwas4d/gwas4d/gwas4d\\_server](http://mulinlab.tmu.edu.cn/gwas4d/gwas4d/gwas4d/gwas4d_server)) (рис. 4). Большинство из этих TF экспрессируются с разным уровнем в широком спектре тканей, за исключением TFAP2A, который экспрессируется в тканях пищеварительной системы (малая слюнная железа, слизистая оболочка пищевода), мочевыводящей системы (кора и мозговое вещество почки, мочевой пузырь), репродуктивной системы (семенники, влагалище), коже (вне зависимости от солнечной экспозиции) (<https://genome.ucsc.edu/>, <https://www.encodeproject.org/>).

Чтобы проанализировать, связывание каких TF подтверждено экспериментально, мы привлекли данные ChIP-Seq-анализа из ресурсов JASPAR (<https://jaspar.genereg.net/>) [19], ORegAnno (<http://www.oreganno.org/>) [20], ENCODE ([МОЛЕКУЛЯРНАЯ БИОЛОГИЯ том 57 № 1 2023](https://www.</a></p>
</div>
<div data-bbox=)



**Рис. 4.** Влияние изученных SNP-маркеров на сайты связывания транскрипционных факторов. Круговые диаграммы – теоретически рассчитанные мотивы связывания TF с изменяющейся в результате нуклеотидной замены аффинностью (по данным HaploReg, Polympact, GWAS4D). Овальные диаграммы – экспериментально подтвержденные ДНК-белковые взаимодействия в регионах локализации нуклеотидной замены (по данным, приведенным в JASPAR, ORegAnno, ENCODE, VannoPortal). Цифрами указано число мотивов в разных базах данных; названия совпадающих по данным различных ресурсов TF приведены рядом.

encodeproject.org/) [21], а также VannoPortal (<http://www.mulinlab.org/vportal/index.html>), на котором представлены объединенные данные CistromeDB 20181120, DeepBlueR V1.0, GTRD 2020-06, EpiMap 2021-01-11, включающие информацию о связывании TF и других взаимодействующих с

хроматином белков [22–24] (рис. 4, овальные диаграммы Венна).

На основании данных по иммунопреципитации хроматина можно сделать вывод, что регионы локализации анализируемых SNP в кодирующей последовательности минимально связаны с TF.

Так, по данным JASPAR, ORegAnno и ENCODE, не обнаружено сайтов связывания TF для rs560191 в гене *TP53BP1*, rs1801516 в гене *ATM*, rs1805321 в гене *PMS2*. Однако по данным VannoPortal, для этих нуклеотидных последовательностей показано связывание в отдельных тканях/клеточных линиях с TF либо иными белками: модификаторами хроматина, ферментативными комплексами и другими, также влияющими на уровень экспрессии белков. Например, в регионе локализации rs560191 выявлено связывание лизинной метилтрансферазы KMT2A (в клеточной линии RS4 острого лимфобластного лейкоза) и POLR2A – субъединицы А РНК-полимеразы II (в клетках влагалища); с регионом локализации rs1801516 в гене *ATM* связывается BRD4 (в клетках крови); в области гена *PMS2*, содержащей rs1805321, обнаружено связывание 13 белков (каждый в 1–3 типах клеток). В регионах еще двух SNP в кодирующей последовательности и одного маркера в 5'UTR факты связывания с единичными TF зарегистрированы в нескольких базах данных, но эти результаты не перекрываются. Так, с регионом локализации rs1799977 в гене *MLH1* по данным ENCODE связывается CTCF (в клетках аденомы паращитовидной железы), по данным VannoPortal – FOXA1 (в клетках простаты), SPIB (в костном мозге), GATA4 и FOXA2 (в коже). С регионом локализации rs709816 в гене *NBN* связывается ряд белков: ZSCAN4, SEBPD, SEBPA (JASPAR); E2F1 (ORegAnno); AR, MYC, POLR2A, BRD4 (VannoPortal). Область локализации rs473297 в гене *MRE11* совпадает с сайтом связывания SMARCA4 (ORegAnno); ARID1A, FOXA2, GATA3, GATA4, GATA6 (VannoPortal) (<http://www.mulinlab.org/vportal/index.html>, <https://www.encodeproject.org/>, <https://jaspar.genereg.net/>, <http://www.oreganno.org/>).

Только для трех изученных маркеров (5'UTR генов *NBN*, *LIG1*, *ATM*) списки связываемых с областью их локализации белков частично пересекаются (рис. 4). Для rs1805800 выявлено связывание нуклеотидной последовательности с TFEB (JASPAR); SMARCA4 и CTCF (ORegAnno); CTCF (в 150 тканях), H2AZ, EP300, SNAI2, SMC3, NR3C1 (VannoPortal); CTCF (в 187 тканях) и еще 8 TF (в 1–5 тканях) (ENCODE). Таким образом, в трех из проанализированных ресурсов содержится информация о связывании этой нуклеотидной последовательности с транскрипционным репрессором CTCF. В регионе rs20579 (в гене *LIG1*) зарегистрировано связывание MITF (ORegAnno); H2AZ, MITF, FOXP1 (VannoPortal). В двух базах в этом регионе зарегистрировано связывание MITF.

Наибольшее число ДНК-белковых взаимодействий зарегистрировано для региона, в котором находится rs189037 (ген *ATM*). В ресурсе JASPAR для этой области показано взаимодействие с факторами ZNF263, SP5 и ZNF148. Согласно ORegAnno,

ДНК в этом регионе ассоциирована с SMARCA4, RBL2, RB1, ZNF263, CTCF и MITF. С этой областью взаимодействует 156 TF по данным ENCODE, 517 – по данным VannoPortal; для двух последних ресурсов 118 белков общие. В списках сразу трех ресурсов встречаются SMARCA4, RB1, CTCF, MITF (ORegAnno, ENCODE, VannoPortal) и ZNF263 (ORegAnno, JASPAR, VannoPortal); из них к TF, связываемым в широком спектре тканей, относится только CTCF (рис. 4).

Как можно заметить, для двух промоторных регионов (rs1805800 и rs189037) характерно связывание с CTCF. Это CCCTC-связывающий фактор, играющий важную мультифункциональную роль в ремоделировании хроматина [12, 25]. Хорошо известна роль CTCF в связывании инсульторных последовательностей, поэтому можно предположить, что именно эта функция CTCF наиболее значима в данных локусах.

В то время как результаты, полученные с помощью методов биоинформатического анализа, указывают на наличие возможности связывания с ДНК определенных TF, методы иммунопреципитации хроматина показывают реально существующие связи, но в отдельных клетках/тканях на определенной стадии развития, в определенном физиологическом состоянии и с определенной нуклеотидной последовательностью. В связи с этим отсутствие экспериментального подтверждения теоретических расчетов нельзя рассматривать как доказательство отсутствия такого взаимодействия в принципе – просто оно не зарегистрировано в проанализированных биообразцах. Мы сопоставили результаты биоинформатического анализа и данные ChIP-seq для исследованных маркеров. Только для rs189037 выявлено совпадение TF и других белков хроматина, аффинность которых к сайтам связывания теоретически изменяется в результате нуклеотидной замены и связывание которых с этим регионом имеет экспериментальное подтверждение. Для аннотированных в базе JASPAR TF не выявлено изменений аффинности к мотивам связывания, а для ряда TF, аннотированных в трех других базах, такие мотивы идентифицированы (табл. 3).

Экспериментально выявляемые регионы ДНК-белковых взаимодействий достаточно протяжены (среди проанализированных их длина составила 152–1975 п.н.). Длина рассчитанных теоретически мотивов значительно короче (8–20 п.н.). Для всех потенциально изменяющихся мотивов подтверждена их локализация внутри региона взаимодействия (на рис. 5 в качестве примера приведены данные для мотивов четырех TF). Таким образом, нуклеотидная замена в ДНК может приводить к изменению аффинности связывания соответствующих TF и тем самым определять эффективность их функционирования. Кроме того,

**Таблица 3.** Экспериментально подтвержденные ДНК-белковые взаимодействия, для которых теоретически предсказано изменение аффинности связывания транскрипционных факторов с их сайтами в результате нуклеотидной замены rs189037

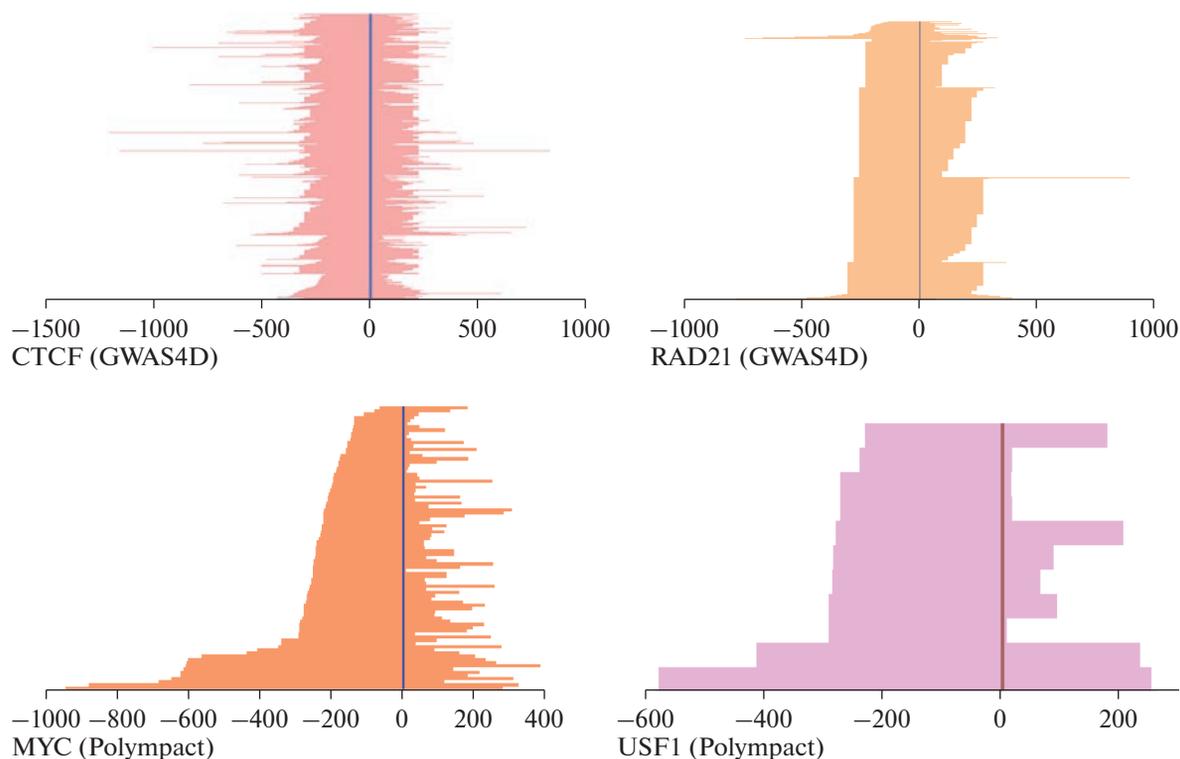
Биоинформатический ресурс	Ресурсы с данными иммунопреципитации хроматина <sup>a</sup>			
	JASPAR	ORegAnno	ENCODE	VannoPortal
HaploReg	0	0	4 MYC ↑ RAD21 ~	5 MYC ↑ RAD21 ~ ELF1 ↓ ZNF143 ↓
GWAS4D	0	1 CTFC ↑	17 CTCF ↑ RAD21 ↑	26 CTCF ↑ RAD21 ↑ SMC3 ↑ GABPA ↑ NRF1 ↑ HDAC2 ↓ SIN3A ↑ CDK8 ↑ USF1 ↑
Polymract	0	0	9 MYC ↑	22 MYC ↑ REST ↑ E2F1 ↓ (7 сайтов) TRIM28 ↓ USF1 ↑ (6 сайтов)

<sup>a</sup> Указано число TF, общих для разных ресурсов; приведены названия только тех TF, наличие которых в данной точке зарегистрировано более чем в 10 тканях. Стрелками ↑ и ↓ указано соответственно увеличение и снижение аффинности TF к мотиву при наличии альтернативного аллеля; знак “~” указывает на незначительное изменение (по биоинформатическим данным). Отдельные мотивы в пределах одного региона в различных ресурсах могут быть сдвинуты вправо или влево относительно друг друга.

при наличии альтернативного аллеля биоинформатические ресурсы предсказывают появление *de novo* сайтов связывания TF, предполагающих появление новых регуляторных контуров. Появление таких сайтов предсказано для всех изученных SNP. Непонятно, реализуется ли эта возможность, но с учетом тканеспецифичности широкого спектра TF можно определить, какие ткани могут быть мишенями в этом случае.

Таким образом, разнообразие изменяемых мотивов связывания TF может объяснить роль исследованных SNP в качестве QTL-вариантов. Так, например, с регионом локализации rs1189037 (в гене *ATM*) связывается фактор USF1 (<http://www.mulinlab.org/vportal/index.html>), регулирующий экспрессию широкого спектра генов метаболизма липидов и глюкозы [26, 27]. В литературе описана мутация в гене *USF1*, приводящая к семейной гиперлипидемии [28], из чего можно сделать вывод о его важной роли в детерминации уровня

липидов. rs1189037 меняет аффинность связывания USF1. Так, согласно ресурсу GWAS4D, при наличии альтернативного аллеля появляется новый, ранее не существовавший мотив связывания этого TF с локализацией 11:108093828–108093837; согласно ресурсу Polymract, увеличивается аффинность пяти вероятных мотивов и появляется один новый (с точкой старта в 11:108093826). Таким образом, при наличии альтернативного аллеля средство TF к последовательности ДНК усиливается, что может влиять на эффективность его функционирования. Следовательно, этот путь может быть еще одним механизмом, объясняющим ассоциированность rs1189037 с липидными показателями у больных ИБС [7]. Вместе с тем вышеприведенная информация указывает на то, что рассматриваемые маркеры при ассоциативных исследованиях могут отражать участие в патологическом процессе не только “своего” гена, но и тех, в регуляции функционирования которых они принимают участие.



**Рис. 5.** Локализация изменяемых мотивов связывания транскрипционных факторов в регионах иммунопреципитации соответствующих белков в различных тканях (составлено по данным Polym pact и VannoPortal). Вертикальная полоса представляет собой теоретически рассчитанный изменяемый мотив, цветные горизонтальные полосы – экспериментально выявленные регионы ДНК-белковых взаимодействий в различных тканях. Ось  $X$  представляет собой длину мотивов связывания TF (в нуклеотидах); за нулевую отметку принято положение теоретически рассчитанного изменяемого мотива. Ось  $Y$  – ткани, в которых с помощью иммунопреципитации хроматина показано связывание этих TF (число тканей составляет 1457, 912, 89 и 11 для CTCF, RAD21, MYC и USF1 соответственно).

### АДАПТОГЕННЫЙ И ПАТОЛОГИЧЕСКИЙ ПОТЕНЦИАЛ АНАЛИЗИРУЕМЫХ ПОЛИМОРФНЫХ ВАРИАНТОВ

Межвидовые сравнения свидетельствуют о высоком уровне консервативности четырех из девяти проанализированных локусов. В первую очередь это касается SNP в кодирующем регионе гена *ATM* – rs1801516, для которого консервативность у позвоночных определена с помощью всех анализируемых алгоритмов (табл. 4). Немного ниже уровень консервативности rs1799977 – замены в кодирующем регионе гена *MLH1*: методом PhyloP (измеряет эволюционную консервативность отдельных последовательностей ДНК) выявлена высокая консервативность у млекопитающих и позвоночных в целом, но не у приматов. Методами PhastCons, GERP<sup>++</sup> консервативность этого локуса на разных эволюционных уровнях подтверждена (табл. 4). Можно предположить, что данный регион в ходе эволюционной истории стал изменяться именно у приматов.

Несколько менее выражена консервативность промоторной последовательности гена *ATM* – для rs189037 устойчивое сохранение последова-

тельности выявлено с помощью анализов PhyloP и GERP<sup>++</sup>, но не PhastCons (определяющего вероятность того, что каждый нуклеотид принадлежит консервативному элементу) (табл. 4). Напротив, для rs560191 в гене *TP53BP1* консервативность определена методом PhastCons на различных уровнях, но PhyloP подтверждает ее наличие только у приматов, а оценки GERP<sup>++</sup> – наличие высокой гомологии у разных видов (табл. 4). Статистика GerpN (оценивающая сохранение последовательности у видов) определяет как консервативные наибольшее число проанализированных локусов (rs560191, rs709816, rs189037, rs1801516, rs1799977, rs1805321), причем только этот метод считает “вероятно консервативными” rs709816 (в гене *NBN*) и rs1805321 (в гене *PMS2*) (табл. 4). Согласно всем видам анализа, к консервативным не относятся промоторные регионы генов *NBN* (rs1805800), *MRE11* (rs473297) и *LIG1* (rs20579).

Не все полученные оценки хорошо согласуются между собой. Так, показано, что 7 из 9 изученных SNP оказывают выраженное влияние на приспособленность (согласно оценкам пригодности FitCons), в том числе и “высоконеутральный” rs20579. В то же время локус rs473297, находящийся-

**Таблица 4.** Эволюционная роль изученных полиморфных вариантов генов, кодирующих белки систем репарации ДНК

Способ оценки <sup>a</sup>	Score/Phred Score <sup>b</sup>								
	rs560191	rs1805800	rs709816	rs189037	rs1801516	rs473297	rs1799977	rs1805321	rs20579
verPhyloP	0.38/5.87	-0.20/1.38	-0.16/1.50	1.66/14.75	4.82/27.54	-0.63/0.64	2.59/19.80	-0.01/2.003	-1.36/0.18
mamPhyloP	0.39/6.28	-0.21/1.45	-0.18/1.51	1.19/12.37	2.75/28.23	-0.64/0.69	2.21/20.12	0.39/6.23	-1.66/0.12
priPhyloP	0.65/14.49	-0.27/1.17	-0.28/1.14	0.65/16.10	0.59/13.11	-2.45/0.03	0.53/8.76	-0.15/1.40	-0.90/0.35
verPastCons	0.99/16.76	0.001/1.50	0.01/3.94	0.04/5.42	1/17.61	0/0	1/17.61	0/0	0.002/2.21
mamPastCons	0.99/17.75	0.001/1.57	0.001/1.57	0.05/6.40	1/18.75	0/0	1/18.75	0.01/3.51	0.001/1.57
priPastCons	0.98/19.14	0.03/3.21	0.002/0.51	0.04/3.65	0.97/17.83	0.01/1.54	0.10/23.77	0.004/0.89	0.10/5.27
GerpN	5.9/25.29	2.35/3.82	5.45/17.31	4.97/13.76	5.52/18.00	3.06/5.51	5.76/21.23	5.84/22.77	4.12/9.69
GerpS	1.95/9.59	-0.79/1.21	-1.6/0.74	4.04/17.01	5.52/23.88	-2.58/0.44	4.61/19.16	0.57/5.32	-7.72/0.02
FitCons	0.70/23.33	0.17/15.34	0.72/27.76	0.09/8.40	0.72/27.76	0.09/8.42	0.72/27.76	0.70/23.34	0.36/18.87
bStatistic	<b>6/0.01</b>	526/0.60	528/0.60	<b>130/0.06</b>	<b>115/0.05</b>	635/0.96	<b>154/0.08</b>	725/1.44	743/1.56

<sup>a</sup> Показатели PhyloP и PastCons оценивают эволюционную консервативность на основании межвидовых сравнений, за исключением человека (приставки обозначают классификационные ранги: ver – позвоночные, mam – млекопитающие, pri – приматы). Балл PhyloP соответствует  $-\lg p$ -value нулевой гипотезы нейтральной эволюции; положительные значения (до 3) указывают на очищающий отбор, а отрицательные значения (до -14) свидетельствуют об ускоренной эволюции консервативных (и, следовательно, вероятно функциональных) элементов [29, 30]. Показатели PhastCons оценивают вероятность того, что локус содержится в консервативном элементе, варьируют от 0 до 1 и отражают вероятность отрицательного отбора [30–32]. Оценки GerpN и GerpS основаны на анализе отдельных нуклеотидов: высокое значение GerpN указывает на высокую гомологию локуса у разных видов; положительные значения GerpS – на дефицит замен, отрицательные – на их избыточность – и оценивают уровень “нейтральности” локуса [30, 33]. FitCons оценивает вероятность того, что нуклеотид в данном положении влияет на приспособленность (на основании метода INSIGHT); показатель варьирует от 0 до 1 [34]. bStatistic – оценка фонового отбора, указывающая на ожидаемую долю нейтрального разнообразия, присутствующего на соответствующем участке: значения, близкие к 0, означают почти полное удаление разнообразия в результате отбора; значения, близкие к 1, указывают на незначительный эффект отбора на разнообразие.

<sup>b</sup> Score – оценка, полученная для каждого варианта с помощью использованного алгоритма; Phred Score – показатель качества полученной оценки (значение  $\geq 20$  соответствует вероятности ошибки 1%). Цифры в светло-серой заливке отображают вероятно консервативные точки генома; в темно-серой заливке – высокое влияние на приспособленность; светло-серая заливка с жирным шрифтом указывает на сильный фоновый отбор, темно-серая с белым шрифтом – локус находится под положительным отбором, черная – локус высоконеитрален.

ся под действием позитивного отбора (то есть замена в этом локусе, вероятно, имеет некоторые адаптивные преимущества), оценки FitCons определяют как не оказывающий влияния на приспособленность. Интересно, что действие фонового отбора зарегистрировано для четырех локусов: обеих замен в гене *ATM* и несинонимичных замен в генах *TP53BP1* и *MLH1* (табл. 4). Фоновый отбор показывает воздействие не на сами анализируемые локусы, а на тесно сцепленные с ними, то есть консервативность этих регионов может быть связана не столько (или не исключительно) с важной ролью именно анализируемых последовательностей, но и быть следствием сцепления. В пользу этого свидетельствует тот факт, что почти во всех генах, в которых локализованы рассматриваемые SNP, описаны мутации, приводящие к моногенным заболеваниям (за исключением *TP53BP1*). Это такие патологии, как атаксия-телеангиэктазия и сходное с ней заболевание (Ataxia-telangiectasia-like disorder 1 – заболевание, подобное атаксии-телеангиэктазии, типа 1), вы-

зываемые мутациями в генах *ATM* и *MRE11* соответственно; синдром хромосомных поломок Ниймеген (мутации в гене *NBN*); синдром Линча и синдром Туркота (причина которых – мутации и в *MLH1*, и в *PMS2*), синдром Мюир–Торре (мутации в гене *MLH1*); аутосомно-рецессивный иммунодефицит-96 (мутации в *LIG1*) (<https://www.ncbi.nlm.nih.gov/omim/?term>).

С позиций “менделевского кода” полиморфные варианты в генах менделевских заболеваний могут оказаться значимыми для многофакторных патологий, в которых задействованы те же биохимические пути. Замечено, что по результатам полногеномных ассоциативных исследований наиболее значимые ассоциации нередко выявляют с маркерами, локализованными вблизи генов менделевских заболеваний, причем наблюдается некоторое перекрытие фенотипических проявлений изучаемых моногенных и сложнаследуемых заболеваний [35]. Считается, что до четверти генов, мутации в которых приводят к менделевским заболеваниям, ассо-

**Таблица 5.** Оценка патогенности несинонимичных нуклеотидных замен<sup>a</sup>

Инструмент	rs560191	rs1801516	rs1799977	rs1805321
SIFT	Tolerated	Tolerated	Tolerated	Tolerated
SIFT4G	Tolerated	Tolerated	Tolerated	Tolerated
Polyphen2_HDIV		Benign	Benign	Benign
Polyphen2_HVAR	Benign	Benign	Benign	Benign
LRT	Neutral	Deleterious	Deleterious	Neutral
MutationTaster	Polymorphism automatic	Polymorphism automatic	Polymorphism automatic	Polymorphism automatic
MutationAssessor		Low	Low	Low
FATHMM	Tolerated			Tolerated
MetaSVM	Tolerated	Tolerated	Tolerated	Tolerated
MetaLR	Tolerated	Tolerated	Tolerated	Tolerated
PrimateAI	Tolerated	Tolerated	Tolerated	Tolerated
DEOGEN2	Tolerated	Tolerated	Tolerated	Tolerated
BayesDel_addAF	Tolerated	Tolerated	Tolerated	Tolerated
BayesDel_noAF	Tolerated	Tolerated	Tolerated	Tolerated
ClinPred	Tolerated	Tolerated	Tolerated	Tolerated
LIST-S2	Tolerated	Tolerated	Damaging	Tolerated

<sup>a</sup> Представлены данные по статусу патогенности миссенс-вариантов, полученные с помощью различных алгоритмов оценки, аккумулированные в dbNSFP (составлено по данным VannoPortal, <http://www.mulinlab.org/vportal/index.html>). В зависимости от алгоритмов и решаемых с их помощью задач статус патогенности оценивали в разных терминах. В таблице сохранена исходная терминология: к благоприятным вариантам относятся Tolerated, Benign, Neutral, Polymorphism automatic, Low; к неблагоприятным – Deleterious, Damaging.

цировано и с многофакторной патологией [36]. С этой точки зрения изучаемые маркеры перспективны для ассоциативного анализа широкого спектра патологий.

Оценки патогенности исследованных SNP неоднозначны. С одной стороны, данные по оценкам патогенности несинонимичных вариантов, полученные на основании широкого спектра наиболее часто используемых алгоритмов, подтверждают отсутствие патологически значимых нарушений белковых молекул вследствие указанных нуклеотидных замен (табл. 5). С другой стороны, геномная оценка патогенности (по regBase [37]) указывает на вероятность вовлеченности в патологические процессы всех анализируемых вариантов (табл. 6). В наибольшей степени это касается SNP в генах *ATM* (rs189037 в 5'UTR, rs1801516 в экзоне 37) и *MLH1* (rs1799977 в экзоне 8). Можно предположить, что геномные оценки патогенности отображают вероятность вовлеченности маркера в развитие многофакторной патологии. Оценки онкогенности также указывают на вероятность “драйверного” эффекта (likely cancer driver) данных нуклеотидных замен для развития онкопатологии (табл. 6).

Следует отметить, что вовлеченность данных маркеров в развитие многофакторной патологии в литературе обсуждается мало. Так, в базе DisGeNet

(<https://www.disgenet.org/>) отсутствует информация относительно rs1805800, rs473297, rs1805321 в контексте привлечения их к ассоциативному анализу при любых патологиях. Только с точки зрения вовлеченности в развитие онкопатологии и/или ее осложнений изучают rs560191, rs709816, rs1801516 и rs20579. Преимущественно при онкопатологии изучают также rs189037 и rs1799977, хотя анализировали вовлеченность rs1799977 в развитие язвенного колита [38, 39], rs189037 – в развитие сердечно-сосудистой патологии, азооспермии, когнитивных нарушений [40–43]. В отношении всех рассматриваемых маркеров результаты ассоциативных исследований противоречивы, в некоторых случаях обсуждается вероятность этноспецифических особенностей ассоциаций (<https://www.disgenet.org/>).

Таким образом, всесторонний анализ изучаемых локусов дает большой объем дополнительной информации, которая может быть полезной для объяснения результатов ассоциативных исследований и более адекватной оценки патологического и адаптогенного потенциала изучаемых локусов. Результаты проведенного анализа указывают на высокий регуляторный потенциал рассматриваемых локусов. Так, показано, что большинство из них находится в составе крупных блоков сцепления QTL-SNP, оказывающих влияние на уровни экспрессии широкого спектра генов

**Таблица 6.** Геномная оценка патогенности и онкогенности нуклеотидных замен<sup>a</sup>

Инструмент	rs560191	rs1805800	rs709816	rs189037	rs1801516	rs473297	rs1799977	rs1805321	rs20579
	Геномная оценка патогенности								
regBase_PAT									
CADD									
DANN									
FATHMM-MKL									
Eigen									
Eigen_PC									
GenoCanyon									
ReMM									
LINSIGHT									
fitCons									
CDTS									
	Геномная оценка онкогенности								
regBase_CAN									
FunSeq2									
CScape									

<sup>a</sup> Суммированы данные по геномной оценке патогенности и онкогенности, полученные с помощью различных часто используемых алгоритмов оценки некодирующих вариантов, заархивированных в базе данных regBase (RegBase\_PAT и RegBase\_CAN) [24, 37]. Составлено на основании информации в VannoPortal (<http://www.mulinlab.org/vportal/index.html>). Светло-серые блоки – оценка “вероятно патогенный”, темно-серые – “вероятный фактор развития рака”.

(например, rs560191 является eQTL для 20 генов), а влияние этих замен может сказаться не только на изменении генной экспрессии в целом, но и на сплайсинге. Выявлен возможный эффект варианта rs189037 на усиление топологической роли включающего его хромосомного региона. Показано, что рассматриваемые маркеры при ассоциативных исследованиях могут отражать участие в патологическом процессе не только “своего” гена, но и тех генов, в регуляции которых они принимают участие.

Работа выполнена при частичном финансировании Госзадания Министерства науки и высшего образования (№ 122020300041-7).

Этические нормы соблюдены. Обзор написан с использованием открытых публикаций.

Авторы заявляют об отсутствии конфликта интересов.

## СПИСОК ЛИТЕРАТУРЫ

- Zhang F., Lupski J.R. (2015) Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**(R1), R102–R110. <https://doi.org/10.1093/hmg/ddv259>
- Бабушкина Н.П., Постригань А.Е., Кучер А.Н. (2021) Вовлеченность генов белков BRCA1-ассоциированного комплекса наблюдения за геномом (BASC) в развитие многофакторной патологии. *Молекуляр. биология.* **55**(2), 318–337. <https://doi.org/10.31857/S0026898421020038>
- Бабушкина Н.П., Постригань А.Е., Кучер А.Н. (2018) Вовлеченность генов систем репарации ДНК в развитие сердечно-сосудистой патологии. Сб.: *Молекулярно-биологические технологии в медицинской практике*. Ред. А.Б. Масленников. Новосибирск: Академиздат, с. 48–62.
- Бабушкина Н.П., Постригань А.Е., Хитринская Е.Ю., Кучер А.Н. (2019) Средовые эффекты на ассоциации генов белков систем репарации ДНК с бронхиальной астмой. *VII Съезд Вавиловского общества генетиков и селекционеров (ВОГиС) (2019)*, Санкт-Петербург, Россия. Сборник тезисов, с. 788.
- Бабушкина Н.П., Постригань А.Е., Хитринская Е.Ю., Кучер А.Н. (2019) Вовлеченность полиморфных вариантов генов систем репарации ДНК в развитие многофакторных заболеваний. Сб.: *Генетика человека и патология: актуальные проблемы клинической и молекулярной цитогенетики*. Ред. В.А. Степанов. Томск: Литературное бюро, с. 5–6.
- Бабушкина Н.П., Постригань А.Е., Кучер А.Н. (2020) Гены белков репарации ДНК и продолжительности жизни. *Медицинская генетика.* **19**(5), 99–100. <https://doi.org/10.25557/2073-7998.2020.05.99-100>

7. Бабушкина Н.П., Постригань А.Е., Кучер А.Н., Кужелева Е.А., Гарганеева А.А. (2020) Ассоциации полиморфизма генов систем репарации ДНК с показателями липидного обмена. *Кардиология 2020 – новые вызовы и новые решения, Казань. Сборник тезисов*, с. 811.
8. Постригань А.Е., Бабушкина Н.П., Кучер А.Н. (2020) Вовлеченность полиморфизма гена *NBN* в формирование предрасположенности к дистропным заболеваниям. *Медицинская генетика*. **19**(8), 98–99.  
<https://doi.org/10.25557/2073-7998.2020.08.98-99>
9. Peakall R., Smouse P.E. (2012) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*. **28**(19), 2537–2539.  
<https://doi.org/10.1093/bioinformatics/bts460>
10. Taverna S.D., Li H., Ruthenburg A.J., Allis C.D., Patel D.J. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**(11), 1025–1040.  
<https://doi.org/10.1038/nsmb1338>
11. Hoon D.S.B., Rahimzadeh N., Bustos M.A. (2021) EpiMap: fine-tuning integrative epigenomics maps to understand complex human regulatory genomic circuitry. *Signal. Transduct. Target Ther.* **6**(1), 179.  
<https://doi.org/10.1038/s41392-021-00620-5>
12. Fu Y., Sinha M., Peterson C.L., Weng Z. (2008) The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4**(7), e1000138.  
<https://doi.org/10.1371/journal.pgen.1000138>
13. Rubio E.D., Reiss D.J., Welch P.L., Disteche C.M., Filippova G.N., Baliga N.S., Aebersold R., Ranish J.A., Krumm A. (2008) CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. USA*. **105**(24), 8309–8314.  
<https://doi.org/10.1073/pnas.0801273105>
14. Mishiro T., Ishihara K., Hino S., Tsutsumi S., Aburatani H., Shirahige K., Kinoshita Y., Nakao M. (2009) Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* **28**(9), 1234–1245.  
<https://doi.org/10.1038/emboj.2009.81>
15. Shoaib M., Chen Q., Shi X., Nair N., Prasanna C., Yang R., Walter D., Frederiksen K.S., Einarsson H., Svensson J.P., Liu C.F., Ekwall K., Lerdrup M., Nordenskiöld L., Sørensen C.S. (2021) Histone H4 lysine 20 mono-methylation directly facilitates chromatin openness and promotes transcription of housekeeping genes. *Nat. Commun.* **12**(1), 4800.  
<https://doi.org/10.1038/s41467-021-25051-2>
16. Hansen K.H., Bracken A.P., Pasini D., Dietrich N., Gehani S.S., Monrad A., Rappsilber J., Lerdrup M., Helin K. (2008) A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**(11), 1291–1300.  
<https://doi.org/10.1038/ncb1787>
17. Vandamme J., Sidoli S., Mariani L., Friis C., Christensen J., Helin K., Jensen O.N., Salcini A.E. (2015) H3K23me2 is a new heterochromatic mark in *Caenorhabditis elegans*. *Nucleic Acids Res.* **43**(20), 9694–9710.  
<https://doi.org/10.1093/nar/gkv1063>
18. Yang W., Bai Y., Xiong Y., Zhang J., Chen S., Zheng X., Meng X., Li L., Wang J., Xu C., Yan C., Wang L., Chang C.C., Chang T.Y., Zhang T., Zhou P., Song B.L., Liu W., Sun S.C., Liu X., Li B.L., Xu C. (2016) Potentiating the antitumour response of CD8<sup>+</sup> T cells by modulating cholesterol metabolism. *Nature*. **531**(7596), 651–655.  
<https://doi.org/10.1038/nature17412>
19. Fornes O., Castro-Mondragon J.A., Khan A., van der Lee R., Zhang X., Richmond P.A., Modi B.P., Correard S., Gheorghe M., Baranasic D., Santana-Garcia W., Tan G., Cheneby J., Ballester B., Parcy F., Sandelin A., Lenhard B., Wasserman W.W., Mathelier A. (2020) JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**(D1), D87–D92.  
<https://doi.org/10.1093/nar/gkz1001>
20. Lesurf R., Cotto K.C., Wang G., Griffith M., Kasaian K., Jones S.J., Montgomery S.B., Griffith O.L.; Open Regulatory Annotation Consortium. (2016) ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res.* **44**(D1), D126–D132.  
<https://doi.org/10.1093/nar/gkv1203>
21. Kazachenka A., Bertozzi T.M., Sjöberg-Herrera M.K., Walker N., Gardner J., Gunning R., Pahita E., Adams S., Adams D., Ferguson-Smith A.C. (2018) Identification, characterization, and heritability of murine metastable epialleles: implications for non-genetic inheritance. *Cell*. **175**(5), 1259–1271.e13.  
<https://doi.org/10.1016/j.cell.2018.09.043>
22. Li M.J., Wang L.Y., Xia Z., Sham P.C., Wang J. (2013) GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic Acids Res.* **41**(Web Server issue), W150–W158.  
<https://doi.org/10.1093/nar/gkt456>
23. Huang D., Yi X., Zhang S., Zheng Z., Wang P., Xuan C., Sham P.C., Wang J., Li M.J. (2018) GWAS4D: multi-dimensional analysis of context-specific regulatory variant for human complex diseases and traits. *Nucleic Acids Res.* **46**(W1), W114–W120.  
<https://doi.org/10.1093/nar/gky407>
24. Huang D., Zhou Y., Yi X., Fan X., Wang J., Yao H., Sham P.C., Hao J., Chen K., Li M.J. (2022) VannoPortal: multiscale functional annotation of human genetic variants for interrogating molecular mechanism of traits and diseases. *Nucleic Acids Res.* **50**(D1), D1408–D1416.  
<https://doi.org/10.1093/nar/gkab853>
25. Lee R., Kang M.K., Kim Y.J., Yang B., Shim H., Kim S., Kim K., Yang C.M., Min B.G., Jung W.J., Lee E.C., Joo J.S., Park G., Cho W.K., Kim H.P. (2022) CTCF-mediated chromatin looping provides a topological framework for the formation of phase-separated transcriptional condensates. *Nucleic Acids Res.* **50**(1),

- 207–226.  
<https://doi.org/10.1093/nar/gkab1242>
26. Putt W., Palmén J., Nicaud V., Tregouet D.A., Tahridaizadeh N., Flavell D.M., Humphries S.E., Talmud P.J.; EARSII group. (2004) Variation in *USF1* shows haplotype effects, gene : gene and gene : environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II. *Hum. Mol. Genet.* **13**(15), 1587–1597.  
<https://doi.org/10.1093/hmg/ddh168>
  27. Laurila P.P., Naukkarinen J., Kristiansson K., Ripatti S., Kauttu T., Silander K., Salomaa V., Perola M., Karhunen P.J., Barter P.J., Ehnholm C., Peltonen L. (2010) Genetic association and interaction analysis of *USF1* and *APOA5* on lipid levels and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **30**(2), 346–352.  
<https://doi.org/10.1161/ATVBAHA.109.188912>
  28. Taghizadeh E., Mirzaei F., Jalilian N., Ghayour Mobarhan M., Ferns G.A., Pasdar A. (2020) A novel mutation in *USF1* gene is associated with familial combined hyperlipidemia. *IUBMB Life.* **72**(4), 616–623.  
<https://doi.org/10.1002/iub.2186>
  29. Pollard K.S., Hubisz M.J., Rosenbloom K.R., Siepel A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**(1), 110–121.  
<https://doi.org/10.1101/gr.097857.109>
  30. Caron B., Luo Y., Rausell A. (2019) NCBoost classifies pathogenic non-coding variants in Mendelian diseases through supervised learning on purifying selection signals in humans. *Genome Biol.* **20**(1), 32.  
<https://doi.org/10.1186/s13059-019-1634-2>
  31. Siepel A. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**(8), 1034–1035.  
<https://doi.org/10.1101/gr.3715005>
  32. Hubisz M.J., Pollard K.S., Siepel A. (2011) PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**(1), 41–51.  
<https://doi.org/10.1093/bib/bbq072>
  33. Zerbino D.R., Achuthan P., Akanni W., Amode M.R., Barrell D., Bhai J., Billis K., Cummins C., Gall A., Giron C.G., Gil L., Gordon L., Haggerty L., Haskell E., Hourlier T., Izuogu O.G., Janacek S.H., Juettemann T., To J.K., Laird M.R., Lavidas I., Liu Z., Loveland J.E., Maurel T., McLaren W., Moore B., Mudge J., Murphy D.N., Newman V., Nuhn M., Ogeh D., Ong C.K., Parker A., Patricio M., Riat H.S., Schuilenburg H., Sheppard D., Sparrow H., Taylor K., Thormann A., Vullo A., Walts B., Zadissa A., Frankish A., Hunt S.E., Kostadima M., Langridge N., Martin F.J., Muffato M., Perry E., Ruffier M., Staines D.M., Trevanion S.J., Aken B.L., Cunningham F., Yates A., Flicek P. (2018) Ensembl 2018. *Nucleic Acids Res.* **46**(D1), D754–D761.  
<https://doi.org/10.1093/nar/gkx1098>
  34. Gulko B., Hubisz M.J., Gronau I., Siepel A. (2015) A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat. Genet.* **47**(3), 276–283.  
<https://doi.org/10.1038/ng.3196>
  35. Freund M.K., Burch K.S., Shi H., Mancuso N., Kichaev G., Garske K.M., Pan D.Z., Miao Z., Mohlke K.L., Laakso M., Pajukanta P., Pasaniuc B., Arbolada V.A. (2018) Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. *Am. J. Hum. Genet.* **103**, 535–552.  
<https://doi.org/10.1016/j.ajhg.2018.08.017>
  36. Spataro N., Rodriguez J. A., Navarro A., Bosch E. (2017) Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology. *Hum. Mol. Genet.* **26**, 489–500.  
<https://doi.org/10.1093/hmg/ddw405>
  37. Zhang S., He Y., Liu H., Zhai H., Huang D., Yi X., Dong X., Wang Z., Zhao K., Zhou Y., Wang J., Yao H., Xu H., Yang Z., Sham P.C., Chen K., Li M.J. (2019) regBase: whole genome base-wise aggregation and functional prediction for human non-coding regulatory variants. *Nucleic Acids Res.* **47**(21), e134.  
<https://doi.org/10.1093/nar/gkz774>
  38. Plotz G., Raedle J., Spina A., Welsch C., Stallmach A., Zeuzem S., Schmidt C. (2008) Evaluation of the MLH1 I219V alteration in DNA mismatch repair activity and ulcerative colitis. *Inflamm. Bowel Dis.* **14**(5), 605–611.  
<https://doi.org/10.1002/ibd.20358>
  39. Vietri M.T., Riegler G., De Paola M., Simeone S., Boggia M., Improta A., Parisi M., Molinari A.M., Cioffi M. (2009) I219V polymorphism in *hMLH1* gene in patients affected with ulcerative colitis. *Genet. Test. Mol. Biomarkers.* **13**(2), 193–197.  
<https://doi.org/10.1089/gtmb.2008.0088>
  40. Li S., Zhang L., Chen T., Tian B., Deng X., Zhao Z., Yuan P., Dong B., Zhang Y., Mo X. (2011) Functional polymorphism rs189037 in the promoter region of *ATM* gene is associated with angiographically characterized coronary stenosis. *Atherosclerosis.* **219**(2), 694–697.  
<https://doi.org/10.1016/j.atherosclerosis.2011.08.040>
  41. Li Z., Yu J., Zhang T., Li H., Ni Y. (2013) rs189037, a functional variant in *ATM* gene promoter, is associated with idiopathic nonobstructive azoospermia. *Fertil. Steril.* **100**(6), 1536–1541.e1.  
<https://doi.org/10.1016/j.fertnstert.2013.07.1995>
  42. Ding X., Yue J.R., Yang M., Hao Q.K., Xiao H.Y., Chen T., Gao L.Y., Dong B.R. (2015) Association between the rs189037 single nucleotide polymorphism in the *ATM* gene promoter and cognitive impairment. *Genet. Mol. Res.* **14**(2), 4584–4592.  
<https://doi.org/10.4238/2015.May.4.17>
  43. Ding X., He Y., Hao Q., Chen S., Yang M., Leng S.X., Yue J., Dong B. (2018) The association of single nucleotide polymorphism rs189037C>T in *ATM* gene with coronary artery disease in Chinese Han populations: a case control study. *Medicine (Baltimore).* **97**(4), e9747.  
<https://doi.org/10.1097/MD.0000000000009747>

## Regulatory Potential of SNP Markers in the Genes of DNA Repair Systems

N. P. Babushkina<sup>1</sup>, \* and A. N. Kucher<sup>1</sup>

<sup>1</sup>Research Institute of Medical Genetics, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, 634050 Russia

\*e-mail: nad.babushkina@medgenetics.ru

In non-coding regions of the genome, the widest range of SNP markers associated with human diseases and petrogenetically significant features were identified. This raised the critical question of identifying the mechanisms that explain these associations. Previously, we identified a number of associations of polymorphic variants of genes encoding DNA repair proteins with multifactorial diseases. To clarify the possible mechanisms underlying established associations, we carried out a detailed annotation of the regulatory potential of the studied markers using a number of on-line resources (GTXPortal, VannoPortal, Ensemble, RegulomeDB, Polypact, UCSC, GnomAD, ENCODE, GeneHancer, EpiMap Epigenomics 2021, HaploReg, GWAS4D, JASPAR, ORegAnno, DisGeNet, OMIM). The article characterizes the regulatory potential of polymorphic variants rs560191 (in the *TP53BP1* gene), rs1805800 and rs709816 (in the *NBN* gene), rs473297 (*MRE11*), rs189037 and rs1801516 (*ATM*), rs1799977 (*MLH1*), rs1805321 (*PMS2*), rs20579 (*LIG1*). Both the general characteristics of the studied markers and information on their influence on the expression of “own” and co-regulated genes, on changes in binding affinity of transcription factors are given. Known data on both adaptogenic and pathogenicity potential of these SNPs and on histone modifications co-localized with them are presented. The potential involvement in regulatory function of not only genes that contain SNPs studied but also nearby genes may explain the association of the markers with diseases and their clinical phenotypes.

**Keywords:** SNP, associations, gene regulation, splicing, transcription factors, histone code, pathogenicity score, sequence conservation

ГЕНОМИКА.  
ТРАНСКРИПТОМИКА

УДК 575.1[616.89+616.899]

АССОЦИАЦИЯ УРОВНЯ СЫВОРОТОЧНОГО ПРОЛАКТИНА  
С ПОЛИМОРФНЫМИ ВАРИАНТАМИ ГЕНОВ *GRIN2A*, *GRM3*, *GRM7*  
У БОЛЬНЫХ ШИЗОФРЕНИЕЙ, ПРИНИМАЮЩИХ  
КОНВЕНЦИОНАЛЬНЫЕ И АТИПИЧНЫЕ АНТИПСИХОТИКИ

© 2023 г. В. В. Тигунцев<sup>а</sup>, \*, В. И. Герасимова<sup>а</sup>, Е. Г. Корнетова<sup>а</sup>, О. Ю. Федоренко<sup>а</sup>,  
А. Н. Корнетов<sup>б</sup>, А. А. Гончарова<sup>а</sup>, Е. Г. Полтавская<sup>а</sup>, А. С. Бойко<sup>а</sup>

<sup>а</sup>Научно-исследовательский институт психического здоровья Томского национального исследовательского  
медицинского центра, Томск, 634014 Россия

<sup>б</sup>Сибирский государственный медицинский университет, Томск, 634050 Россия

\*e-mail: cristall2009@live.ru

Поступила в редакцию 24.05.2022 г.

После доработки 11.07.2022 г.

Принята к публикации 22.07.2022 г.

Дофаминовая, серотониновая и глутаматная системы тесно взаимодействуют в патогенезе и фармакотерапии шизофрении. Мы сформулировали гипотезу, согласно которой полиморфные варианты генов *GRIN2A*, *GRM3* и *GRM7* могут быть ассоциированы с развитием гиперпролактинемии у больных шизофренией, принимающих конвенциональные и атипичные антипсихотики в качестве базовой терапии. Обследовано 432 пациента славянских национальностей с установленным диагнозом шизофрения. ДНК выделяли из лейкоцитов периферической крови стандартным фенол-хлороформным методом. Для пилотного генотипирования выбрано 12 SNP в гене *GRIN2A*, четыре SNP в гене *GRM3* и шесть в гене *GRM7*. Аллельные варианты этих полиморфизмов определяли методом ПЦР в реальном времени. Уровень пролактина определяли иммуноферментным методом. В группах пациентов с нормальным и повышенным уровнем пролактина, принимающих конвенциональные антипсихотики, выявлены статистически значимые различия в распределении частот генотипов и аллелей полиморфных вариантов *GRIN2A* rs9989388 и *GRIN2A* rs7192557, а также различия в уровнях сывороточного пролактина в зависимости от генотипов полиморфного варианта *GRM7* rs3749380. Обнаружены статистически значимые различия в частотах генотипов и аллелей полиморфного варианта *GRM3* rs6465084 в группах пациентов с гиперпролактинемией и без нее, чьим базовым препаратом был атипичный антипсихотик. Впервые показана ассоциация полиморфных вариантов генов *GRIN2A*, *GRM3* и *GRM7* с развитием гиперпролактинемии у больных шизофренией, принимающих конвенциональные и атипичные антипсихотики. Эти ассоциации не только подтверждают роль глутаматергической системы в развитии шизофрении, но и показывают, что следует учитывать генетическую составляющую в терапии больных с данным расстройством.

**Ключевые слова:** пролактин, гиперпролактинемия, глутамат, однонуклеотидные полиморфизмы, ПЦР в реальном времени, человек, антипсихотики, шизофрения

**DOI:** 10.31857/S0026898423010159, **EDN:** AXSWIL

## ВВЕДЕНИЕ

Существуют три основные гипотезы, объясняющие патогенез шизофрении с точки зрения нарушений в нейромедиаторных системах центральной нервной системы (ЦНС).

Дофаминергическая гипотеза предполагает повышенную активность дофамина в мезолимбической системе. Эту гипотезу подтверждает механизм действия конвенциональных антипсихоти-

ков, направленный на блокаду D2-рецепторов [1]. Серотонинергическая гипотеза основана на том, что гиперфункция рецепторов 5-HT<sub>2A</sub> на глутаматных нейронах приводит к повышенному выбросу глутамата в вентральной области покрышки, а это, в свою очередь, активирует мезолимбический тракт и повышает выброс дофамина в вентральном стриатуме [2].

Глутаматергическая гипотеза представляет шизофрению как следствие гипофункции NMDA-рецепторов на промежуточных ГАМКергических нейронах в коре головного мозга [3]. Такая гипо-

Сокращения: ГПРЛ – гиперпролактинемия; ГАМК – гамма-аминомасляная кислота; PANSS – Positive And Negative Syndrome Scale.

функция может привести к гиперактивации глутаматной нейротрансмиссии в вентральной области покрышки, а также к избытку дофамина в вентральном стриатуме через мезолимбический тракт. Согласно этой же гипотезе, позитивные симптомы шизофрении вызывает избыток дофамина в мезолимбическом тракте, обусловленный нарушениями в функционировании нейронной цепи глутамат–ГАМК–глутамат–дофамин. Негативные и когнитивные симптомы можно объяснить недостаточным высвобождением дофамина в мезокортикальной области из-за нарушения работы нейронной цепи глутамат–ГАМК–глутамат–ГАМК–дофамин [4].

Участие глутамата в патогенезе шизофрении также подтверждается результатами исследований, показавшими способность блокатора NMDA-рецепторов фенциклидина вызывать преходящие шизофреноподобные психозы [5, 6]. Патогенетический механизм этого явления основан на блокировании притока в клетку кальция через управляемые NMDA-рецепторами возбуждающие каналы [7]. Клинические наблюдения подтверждают, что для фенциклидиновых психозов характерны все три группы симптомов шизофрении: позитивные, негативные и когнитивные [8].

Недавние исследования показали, что серотонинергическая и глутаматергическая системы взаимодействуют в патогенезе и фармакотерапии шизофрении. В частности, рецепторы 5-HT<sub>2A</sub> и GRM2 могут регулировать друг друга в функционально антагонистической аллостерической манере [9]. Эта взаимосвязь между рецепторами проявляется на различных уровнях: через регуляцию нижестоящих эффекторов, посредством эпигенетических и генетических факторов или как результат сборки рецепторов 5-HT<sub>2A</sub> и GRM2 в гетеромерный комплекс, который участвует в механизме действия антипсихотиков [10].

Известно, что фармакотерапия антипсихотическими препаратами способствует снижению симптоматики шизофрении и переходу заболевания в состояние стабильной ремиссии [11, 12]. Данные препараты в зависимости от селективности действия подразделяются на конвенциональные (типичные, традиционные, первого поколения) и атипичные (второго поколения).

Эффект конвенциональных антипсихотиков достигается строго за счет блокады D<sub>2</sub>-рецепторов дофамина [13]. Атипичные антипсихотические препараты воздействуют на большее число фармакологических мишеней, многие из них взаимодействуют с 5-HT<sub>2</sub>-рецепторами (рисперидон, оланзапин, клозапин, зипрасидон, сертиндол). Благодаря расширенному спектру фармакологического действия, атипичные антипсихотические препараты показали высокую эффективность в отношении негативных симптомов шизофрении [14].

Один из наиболее распространенных побочных эффектов антипсихотических препаратов – гиперпролактинемия (ГПРЛ), существенно снижает качество жизни пациентов [15]. В основе развития ГПРЛ лежит блокада D<sub>2</sub>-рецепторов дофамина и угнетение дофаминергической передачи в тубероинфундибулярной зоне головного мозга [16].

Традиционные антипсихотики довольно часто вызывают ГПРЛ (примерно у 40–90% пациентов). Большая часть атипичных препаратов также значительно повышает уровень пролактина в сыворотке крови [17], однако некоторые имеют низкий риск развития ГПРЛ (клозапин, оланзапин, зипрасидон), что связано с их большей аффинностью к 5-HT<sub>2A</sub>-рецепторам, чем к D<sub>2</sub>-рецепторам [18]. Этот эффект также можно объяснить антихолинергическими свойствами оланзапина и клозапина [19].

Неоспорима роль генетической составляющей в функционировании нейромедиаторных систем ЦНС, в частности глутаматергической [20, 21]. Ген *GRIN2A* кодирует субъединицу NMDA-рецептора, представителя суперсемейства ионотропных глутаматных рецепторов, которые принимают участие в кальций-опосредованном компоненте возбуждающей синаптической передачи в ЦНС. Гены *GRM3* и *GRM7* кодируют метаботропные рецепторы глутамата, связанные с G-белком. Эти рецепторы принимают участие в ингибировании каскада cAMP, но различаются по селективности агонистов. Полиморфизмы этих генов могут влиять на концентрацию глутамата, на активность глутаматергической нейротрансмиссии и, как следствие, на содержание дофамина в мезолимбическом тракте. В свою очередь, повышение уровня дофамина в этой области ведет к развитию ГПРЛ.

Исходя из того, что дофаминовая, серотонинергическая и глутаматная системы тесно взаимодействуют в патогенезе и фармакотерапии шизофрении, мы сформулировали гипотезу, согласно которой полиморфные варианты генов *GRIN2A*, *GRM3* и *GRM7* могут быть ассоциированы с развитием ГПРЛ у больных, принимающих конвенциональные и атипичные антипсихотики в качестве базовой терапии.

Мы сравнили частоты генотипов и аллелей полиморфных вариантов генов *GRIN2A*, *GRM3* и *GRM7* в группах больных шизофренией с ГПРЛ и без нее, проживающих в Томской области, принадлежащих к славянской этноязыковой общности и принимающих в качестве базовой терапии конвенциональные и атипичные антипсихотики.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Нами обследовано 432 пациента с установленным диагнозом шизофрения, проходивших ста-

ционарное лечение в клиниках НИИ психического здоровья Томского НИМЦ и в Томской клинической психиатрической больнице. Основными критериями для включения пациентов были верифицированный диагноз шизофрении по МКБ-10 (Международная классификация болезней 10 пересмотра), возраст 18–65 лет, подписанное информированное согласие пациента, принадлежность к славянской этноязыковой общности, постоянное проживание на территории Западной Сибири. В исследование не включали пациентов с острыми и хроническими инфекционными, воспалительными, аутоиммунными, соматическими заболеваниями в острой фазе, синдромом зависимости от опиоидов, каннабиноидов, седативных или снотворных веществ, кокаина, психостимуляторов, включая кофеин, галлюциногенов, летучих растворителей.

Среди обследованных было 211 мужчин (48.5%) и 221 женщина (51.5%) в возрасте 18–65 лет, медиана 40 [32; 51] лет.

Длительность заболевания у 265 человек (61.3%) превышала 10 лет, у 77 пациентов (17.8%) – от 5 до 10 лет, у 77 (17.8%) – от 1 г. до 5 лет, у 13 человек (3.0%) не превышала 1 г.

Суточный хлорпромазиновый эквивалент принимаемых антипсихотиков составил 450 [232.5; 800] мг, медиана длительности терапии – 11.7 [3, 18] г. При первичном обследовании пациентов использовали шкалу позитивных и негативных синдромов PANSS (Positive And Negative Syndrome Scale [22]), которая позволяет проводить стандартизированную оценку различных векторов психопатологической симптоматики шизофрении, определять клинический профиль больного и проследить динамику состояния в процессе терапии. Нами получены следующие результаты, представленные в виде медианы, верхнего и нижнего квартилей (табл. 1).

ДНК выделяли из лейкоцитов периферической крови стандартным фенол-хлороформным методом [23]. Для генотипирования выбрано 12 SNP в гене *GRIN2A* (rs11644461, rs11646587, rs11866328, rs7190619, rs7192557, rs7196095, rs7206256, rs9788936, rs9989388, rs1345423, rs4782039, rs8057394), четыре SNP в гене *GRM3* (rs1468412, rs6465084, rs2237562, rs2299225) и шесть SNP в гене *GRM7* (rs779746, rs1450099, rs17031835, rs3749380, rs1396409, rs12491620). Полиморфные варианты отбирали на основании публикаций в сфере генетики и фармакогенетики шизофрении, в которых показано как участие генов *GRIN2A* и *GRM7* в ответе на терапию антипсихотическими препаратами [24, 25], так и констатирована противоречивость данных о вовлеченности гена *GRM3* в патогенез шизофрении [26].

Аллельные варианты определяли методом ПЦР в реальном времени с использованием набо-

**Таблица 1.** Показатели пациентов исследуемой выборки по шкале PANSS

Шкала PANSS	Me [Q1; Q3]
Позитивные симптомы	23 [19, 27]
Негативные симптомы	25 [21, 28]
Общепсихопатологические симптомы	51 [44; 57]
Общий показатель	99 [91; 108]

ров SNP Genotyping Assay на приборе StepOne-Plus (США).

Уровень пролактина определяли иммуноферментным методом с использованием набора реагентов PRL Test System (“MonobindInc”, США) для количественного определения содержания гормона в сыворотке крови. У женщин репродуктивного возраста забор крови производили в первую фазу менструального цикла. Антипсихотик-индуцированную ГПРЛ диагностировали при уровне пролактина выше 25 нг/мл у женщин и 20 нг/мл у мужчин в соответствии с международными критериями [27].

Статистическую обработку данных проводили с использованием пакета программ SPSS 17.0. Применяли расчет средних величин,  $\chi^2$  Пирсона, в том числе, с учетом поправки Йетса, точный критерий Фишера (если не менее чем в одной из исследуемых групп было меньше пяти человек), Н-критерий Крускала–Уоллеса, U-критерий Манна–Уитни и множественный регрессионный анализ. Поскольку часть показателей не подчинялась закону нормального распределения, использовали непараметрические критерии, их описательная статистика представлена медианой и межквартильным интервалом. Критический уровень значимости принят равным 0.05.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Антипсихотик-индуцированная ГПРЛ выявлена у 262 (60.6%) из 432 пациентов нашей выборки и не выявлена у 170 (39.4%). Распределение пациентов по полу в этих группах представлено в табл. 2. Данные о длительности течения заболевания в группах приведены в табл. 3.

Поскольку ГПРЛ развивается при постоянном приеме антипсихотических препаратов в течение нескольких недель, выявленные различия могут быть следствием особенностей выборки, в том числе условного деления на подгруппы в зависимости от длительности антипсихотической терапии. По этой причине эти различия не учитывали при дальнейшем анализе частот генотипов и аллелей изучаемых полиморфных вариантов.

Далее выборку разбили на две группы в зависимости от класса базового антипсихотика с последующим делением по наличию/отсутствию ГПРЛ.

**Таблица 2.** Частота развития гиперпролактинемии у больных шизофренией

Пациенты	ГПРЛ–, <i>n</i> (%)	ГПРЛ+, <i>n</i> (%)	$\chi^2$	<i>p</i>
Мужчины	97 (57.1)	114 (43.5)	7.04	0.008
Женщины	73 (42.9)	148 (56.6)*		
Всего	170 (100)	262 (100)		

**Таблица 3.** Гиперпролактинемия у пациентов с различной длительностью заболевания

Длительность заболевания	ГПРЛ–, <i>n</i> (%)	ГПРЛ+, <i>n</i> (%)	<i>F</i>	<i>p</i>
Менее 1 года	3 (1.8)	11 (4.2)*	11.48	0.009
1–5 лет	17 (10)	60 (22.9)		
5–10 лет	27 (15.9)	50 (19.1)		
Более 10 лет	123 (72.4)	141 (53.8)		
Всего	170 (100.0)	262 (100.0)		

**Таблица 4.** Частота ГПРЛ у пациентов разного пола, получающих конвенциональные антипсихотики

Пол	ГПРЛ–	ГПРЛ+	$\chi^2$	<i>p</i>
Женщины	38 (29%)	93 (71%)*	8.17	0.004
Мужчины	64 (46.7%)	73 (53.3%)		

Не выявлено статистически значимых различий в частотах ГПРЛ в группах пациентов, принимающих в качестве базовой терапии конвенциональные и атипичные антипсихотики ( $\chi^2 = 0.27$ ;  $p = 0.64$ ). ГПРЛ диагностирована у 61.9% пациентов, получающих конвенциональные антипсихотики, и у 59.1% пациентов, получающих атипичные антипсихотики.

Оценили также развитие ГПРЛ в группах определенной половой принадлежности. Оказалось, что среди женщин, получающих конвенциональные антипсихотики, ГПРЛ встречалась статистически значимо чаще, чем у мужчин ( $\chi^2 = 8.17$ ;  $p = 0.004$ ; табл. 4). В группе больных, чьим базовым препаратом был атипичный антипсихотик, подобных различий не обнаружено.

На следующем этапе проведено генотипирование пациентов по выбранным однонуклеотидным полиморфизмам. Анализ распределения частот генотипов и аллелей позволил выявить статистически значимые различия между группами пациентов с нормальным и повышенным уровнем пролактина по полиморфному варианту *GRIN2A* rs9989388 среди принимающих в качестве базовой терапии конвенциональные антипсихотики ( $\chi^2 = 5.99$ ;  $p = 0.03$ ; табл. 5): генотип TC преобладал в группе пациентов с ГПРЛ ( $OR = 1.93$ ; 95% CI 1.14–3.28). Также обнаружены статистически значимые различия в частотах генотипов и аллелей полиморфного варианта rs7192557 того

же гена. Генотип GG ( $\chi^2 = 5.89$ ;  $p = 0.05$ ;  $OR = 0.56$ ; 95% CI 0.34–0.93) и аллель G ( $\chi^2 = 5.30$ ;  $p = 0.02$ ;  $OR = 0.60$ ; 95% CI 0.39–0.91) чаще встречались у пациентов с нормальным уровнем пролактина в сыворотке крови, в то время как аллель A был более характерен для подгруппы с ГПРЛ ( $OR = 1.67$ ; 95% CI 1.10–2.54).

Среди пациентов, чьим базовым препаратом был атипичный антипсихотик, обнаружены статистически значимые различия в частотах генотипов и аллелей полиморфного варианта *GRM3* rs6465084 (табл. 6). Генотип AA ( $F = 8.02$ ;  $p = 0.02$ ;  $OR = 2.02$ ; 95% CI 1.07–3.81) и аллель A ( $\chi^2 = 6.87$ ;  $p = 0.01$ ;  $OR = 2.02$ ; 95% CI 1.07–3.81) предрасполагали к развитию ГПРЛ. Генотип GG ( $OR = 0.21$ ; 95% CI 0.06–0.82) и аллель G ( $OR = 0.50$ ; 95% CI 0.30–0.82), напротив, оказывали протективный эффект в отношении повышения уровня сывороточного пролактина.

Далее сравнили уровни сывороточного пролактина у пациентов с разными генотипами исследуемых полиморфизмов. Удалось обнаружить статистически значимые различия ( $H = 7.36$ ;  $p = 0.025$ ) в частотах генотипов полиморфного варианта *GRM7* rs3749380 в группе пациентов, принимающих в качестве базовой терапии конвенциональные антипсихотики: у носителей генотипа TT уровень пролактина был примерно в 1.5 раза выше, чем у носителей генотипов CC и CT (табл. 7).

**Таблица 5.** Сравнение частот генотипов и аллелей полиморфных вариантов гена *GRIN2A* в группах пациентов с ГПРЛ и без побочного эффекта, принимающих в качестве базовой терапии конвенциональные антипсихотики

Ген <i>GRIN2A</i>		ГПРЛ ( <i>n</i> , %)	Без побочного эффекта ( <i>n</i> , %)	OR		$\chi^2/F$	<i>p</i>
				знач.	95% CI		
Полиморфизм rs9989388							
Генотип	CC	82 (50.3)	60 (61.2)	0.64	0.39–1.07	5.99	0.03
	TC	75 (46.0)	30 (30.6)	1.93	1.14–3.28		
	TT	6 (3.4)	8 (8.2)	0.43	0.15–1.28		
Аллель	C	239 (73.5)	150 (76.5)	0.85	0.56–1.29	0.43	0.51
	T	86 (26.5)	46 (23.5)	1.17	0.78–1.77		
Полиморфизм rs7192557							
Генотип	GG	80 (48.5)	62 (62.7)	0.56	0.34–0.93	5.89	0.05
	GA	72 (43.6)	34 (34.3)	1.48	0.88–2.48		
	AA	13 (7.9)	3 (3.0)	2.74	0.76–9.86		
Аллель	G	232 (70.3)	158 (79.8)	0.60	0.39–0.91	5.30	0.02
	A	98 (29.7)	40 (20.2)	1.67	1.10–2.54		

**Таблица 6.** Сравнение частот генотипов и аллелей полиморфного варианта *GRM3* rs6465084 в группах пациентов с ГПРЛ и без побочного эффекта, принимающих в качестве базовой терапии атипичные антипсихотики

Полиморфизм <i>GRM3</i> rs6465084		ГПРЛ ( <i>n</i> , %)	Безпобочного эффекта ( <i>n</i> , %)	OR		$\chi^2/F$	<i>p</i>
				знач.	95% CI		
Генотип	AA	57 (60.6)	29 (43.3)	2.02	1.07–3.81	8.02	0.02
	GA	34 (36.2)	29 (43.3)	0.74	0.39–1.41		
	GG	3 (3.2)	9 (13.4)	0.21	0.06–0.82		
Аллель	A	148 (74.5)	87 (64.9)	1.99	1.22–3.29	6.87	0.01
	G	40 (25.5)	47 (35.1)	0.50	0.30–0.82		

**Таблица 7.** Распределение уровня пролактина в зависимости от генотипов полиморфного варианта *GRM7* rs3749380 в группе пациентов с шизофренией, принимающих в качестве базовой терапии конвенциональные антипсихотики

Генотип <i>GRM</i> rs3749380	<i>N</i>	Уровень пролактина, Me (Q1, Q3)	<i>H</i>	<i>p</i>
CC	96	30.96 [16.36; 61.56]	7.36	0.025
CT	122	27.40 [13.83; 61.95]		
TT	46	50.11 [22.99; 92.15]		

Дополнительно в общей выборке проведен множественный регрессионный анализ, направленный на выявление потенциального вклада фактора пола в выявленные ассоциации. На основании построения регрессионной модели (табл. 8) показано, что в группе пациентов, принимающих конвенциональные антипсихотики, риск развития ГПРЛ ассоциирован с полом ( $p = 0.004$ ;  $R^2 = 0.0297$ ).

Таким образом, обнаруженные ассоциации полиморфных вариантов *GRIN2A* rs998938 и *GRIN2A* rs7192557 с развитием ГПРЛ, а также полиморфного варианта *GRM7* rs3749380 с уровнем пролактина в сыворотке крови могут быть связаны со статистически значимым различием в частоте ГПРЛ у мужчин и женщин в исследуемой выборке.

**Таблица 8.** Множественный регрессионный анализ влияния пола на риск развития ГПРЛ

Группа	B	p
Принимающие конвенциональные антипсихотики	-0.167	0.004*
Принимающие атипичные антипсихотики	-0.077	0.359

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Взаимосвязь нейромедиаторных систем головного мозга в патогенезе и фармакотерапии шизофрении является одной из малоизученных проблем современной биологической психиатрии. Полиморфные варианты генов, выбранные нами, уже становились предметом изучения в качестве биомаркеров риска шизофрении, однако эти работы касались исключительно глутаматергической системы без вовлечения дофаминовой системы как основного звена патогенеза антипсихотик-индуцированной ГПРЛ.

Нами впервые обнаружены ассоциации полиморфных вариантов генов глутаматергической системы с развитием антипсихотик-индуцированной ГПРЛ у больных шизофренией, принимающих в качестве базовой терапии конвенциональные и атипичные антипсихотики.

Опубликованы данные, доказывающие вовлеченность метаболитного глутаматного рецептора (GRM3) в патогенез шизофрении. Так, лечение пациентов аналогом аминокислоты LY2140023, чье действие направлено на активацию глутаматных рецепторов GRM3/GRM2, приводило к сглаживанию позитивных и негативных симптомов, сравнимому с терапией оланзапином [28]. Выявлена также повышенная частота встречаемости генотипа AA и аллеля A функционального полиморфизма *GRM3* rs6465084 у больных шизофренией по сравнению с контрольной группой [29]. С этим косвенно согласуются полученные нами результаты, согласно которым полиморфный вариант *GRM3* rs6465084 ассоциирован с развитием ГПРЛ у пациентов, принимающих атипичные антипсихотики. Есть и контраргументы участия этого гена в патогенезе шизофрении. Например, коллектив авторов во главе с Norton N. [30] не смог обнаружить различий в частотах генотипов, аллелей и гаплотипов всех исследованных ранее полиморфизмов гена *GRM3* между больными шизофренией и здоровыми лицами в популяции Великобритании. Подобные отрицательные результаты получены и на японской популяции [31]. Дальнейшие комплексные исследования могли бы пролить свет на этот вопрос.

Вклад полиморфных вариантов гена *GRM7* в патогенез шизофрении нашел подтверждение в исследовании на китайской популяции, где Liang W. и соавт. [25] удалось выявить ассоциацию нескольких полиморфизмов с ответом на терапию кветиапином, рисперидоном, галоперидо-

лом и перициaziном. Также частоты аллелей одного полиморфизма указанного гена статистически значимо различались в группах больных шизофренией и в контроле. Эта разница осталась значимой после введения поправки Бонферрони, однако отношение шансов, равное 0.95, заставляет относиться к таким результатам с осторожностью. Франко-канадские исследователи показали ассоциацию полиморфизма *GRM7* rs1396409 с нейрокогнитивным дефицитом на ранних стадиях психоза [32].

Нам удалось обнаружить статистически значимые различия в уровнях сывороточного пролактина в зависимости от генотипов полиморфизма *GRM7* rs3749380 в группе пациентов, получающих конвенциональные антипсихотики, однако не выявлено ассоциации этого локуса с риском развития ГПРЛ. Можно предположить, что отсутствие ассоциаций обусловлено особенностями выборки и тем, что под действием антипсихотических препаратов уровень сывороточного пролактина у некоторых пациентов в зависимости от их фармакогенетических особенностей может повышаться сильнее, чем у других. При этом ГПРЛ с высокой вероятностью диагностируется и у тех, и у других. Различия в повышенных уровнях пролактина становятся видны только при количественном их сравнении, что и показано на примере полиморфного варианта *GRM7* rs3749380.

Коллектив нашего института и ранее искал данные, подтверждающие глутаматергическую теорию шизофрении. Например, ранее на выборке пациентов Сибирского региона нами показана [33] ассоциация полиморфизма *GRIN2A* rs7206256 с ранним возрастом манифестации шизофрении. С ведущей симптоматикой и типом течения заболевания подобных ассоциаций не обнаружено. Участие указанного гена в патогенезе шизофрении подтверждают и результаты нашего исследования.

Множественный регрессионный анализ, проведенный на общей выборке, показал связь риска развития ГПРЛ с полом пациентов, принимающих конвенциональные антипсихотики. Как наши данные [34], так и результаты других ученых [35], подтверждают, что женский пол является фактором риска развития антипсихотик-индуцированной ГПРЛ. Антипсихотики первого поколения вызывают ГПРЛ примерно у 40–90% пациентов, поскольку их действие направлено строго на блокаду D2-дофаминовых рецепторов [36]. Таким

образом, в основе выявленной ассоциации может лежать профиль принимаемых пациентами конвенциональных антипсихотиков, которые с высокой частотой вызывают развитие ГПРЛ. В свою очередь, среди атипичных антипсихотиков, которые назначались пациентам исследуемой выборки, преобладали кветиапин и оланзапин, имеющие низкий риск повышения сывороточного пролактина.

Несмотря на довольно активное изучение патогенеза и фармакотерапии шизофрении, нам не удалось найти в международных базах данных публикаций, направленных на изучение вклада генов глутаматной системы в развитие антипсихотик-индуцированной ГПРЛ. Наше исследование является пилотным и имеет ряд ограничений, поскольку мы не использовали поправку на множественные сравнения. Кроме того, рассматривается возможность анализа частот гаплотипов в исследованных группах.

Результаты данного исследования еще раз подтверждают тесную связь дофаминергической, серотонинергической и глутаматергической систем в развитии шизофрении, однако требуют более детального изучения, в том числе, на других популяциях.

Работа получила поддержку Российского научного фонда (грант № 21-15-00212 2021–2023 гг. для О.Ю. Федоренко).

Протокол исследования одобрен этическим комитетом НИИ психического здоровья Томского НИМЦ (от 14 мая 2021 г., Дело № 142/2.2021). Все процедуры, выполненные в настоящем исследовании, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики. От каждого из включенных в исследование участников получено информированное добровольное согласие.

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

#### СПИСОК ЛИТЕРАТУРЫ

1. McCutcheon R.A., Krystal J.H., Howes O.D. (2020) Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. *World Psychiatry*. **19**(1), 15–33. <https://doi.org/10.1002/wps.20693>
2. Stahl S. (2018) Beyond the dopamine hypothesis of schizophrenia to three neural networks of psychosis: dopamine, serotonin, and glutamate. *CNS Spectrums*. **23**(3), 187–191. <https://doi.org/10.1017/S1092852918001013>
3. Гареева А.Э. (2019) Современный взгляд на нейробиологические гипотезы шизофрении. *Журн. высшей нервной деятельности им. И.П. Павлова*. **69**(4), 437–455. <https://doi.org/10.1134/S0044467719040038>
4. Schwartz T.L., Sachdeva S., Stahl S.M. (2012) Glutamate neurocircuitry: theoretical underpinnings in schizophrenia. *Front. Pharmacol.* **3**, 195.
5. Cadinu D., Grayson B., Podda G., Harte M.K., Doostdar N., Neill J.C. (2018) NMDA receptor antagonist rodent models for cognition in schizophrenia and identification of novel drug treatments, an update. *Neuropharmacology*. **142**, 41–62. <https://doi.org/10.1016/j.neuropharm.2017.11.045>
6. Adell A. (2020) Brain NMDA receptors in schizophrenia and depression. *Biomolecules*. **10**(6), 947. <https://doi.org/10.3390/biom10060947>
7. Javitt D.C., Jotkowitz A., Sircar R., Zukin S.R. (1987) Non-competitive regulation of phencyclidine/sigma-receptors by the *N*-methyl-*D*-aspartate receptor antagonist *D*-(-)-2-amino-5-phosphonovaleric acid. *Neurosci. Lett.* **78**(2), 193–198. [https://doi.org/10.1016/0304-3940\(87\)90632-x](https://doi.org/10.1016/0304-3940(87)90632-x)
8. Fujigaki H., Mouri A., Yamamoto Y., Nabeshima T., Saito K. (2019) Linking phencyclidine intoxication to the tryptophan-kynurenine pathway: therapeutic implications for schizophrenia. *Neurochem. Int.* **125**, 1–6. <https://doi.org/10.1016/j.neuint.2019.02.001>
9. Shah U.H., González-Maeso J. (2019) Serotonin and glutamate interactions in preclinical schizophrenia models. *ACS Chem. Neurosci.* **10**(7), 3068–3077. <https://doi.org/10.1021/acschemneuro.9b00044>
10. Ibi D., de la Fuente Revenga M., Kezunovic N., Mugu-ruza C., Saunders J.M., Gaitonde S.A., Moreno J.L., Ijaz M.K., Santosh V., Kozlenkov A., Holloway T., Seto J., García-Bea A., Kurita M., Mosley G.E., Jiang Y., Christoffel D.J., Callado L.F., Russo S.J., Dracheva S., López-Giménez J.F., Ge Y., Escalante C.R., Meana J.J., Akbarian S., Huntley G.W., González-Maeso J. (2017) Antipsychotic-induced Hdac2 transcription via NF-κB leads to synaptic and cognitive side effects. *Nat. Neurosci.* **20**(9), 1247–1259. <https://doi.org/10.1038/nn.4616>
11. Мосолов С.Н. (2018) Психозы дофаминовой гиперчувствительности на современном этапе антипсихотической фармакотерапии шизофрении: что нужно знать практикующему врачу. *Современная терапия психических расстройств*. **4**, 41–49.
12. Иванов М.В., Незнанов Н.Г., Костерин Д.Н. (2014) Антипсихотики. В: *Рациональная фармакотерапия в психиатрической практике. Руководство для практикующих врачей*. Сер. “Рациональная фармакотерапия”; под ред. Александровского Ю.А., Незнанова Н.Г. М.: Литтерра, 142–175.
13. Mittal S., Prasad S., Ghosh A. (2018) Antipsychotic-induced hyperprolactinaemia: case studies and review. *Postgrad. Med. J.* **94**(1110), 226–229. <https://doi.org/10.1136/postgradmedj-2017-135221>
14. Jeong H., Lee M. (2013) Long-acting injectable antipsychotics in first-episode schizophrenia. *Clin. Psychopharmacol. Neurosci.* **11**(1), 1–6.
15. Горобец Л.Н., Буланов В.С., Литвинов А.В. (2016) Особенности формирования нейроэндокринных дисфункций у больных шизофренией в амбулатор-

- ной практике (натуралистическое исследование). *Фарматека*. **S4**, 41–45.
16. Османова Д.З., Бойко А.С., Федоренко О.Ю., Пожидаев И.В., Фрейдин М.Б., Стегний В.Н., Корнетова Е.Г., Иванова С.А. (2018) Роль генов дофаминергической системы в развитии антипсихотик-индуцированной гиперпролактинемии у больных шизофренией. *Психическое здоровье*. **16**(5), 25–27.
  17. Leucht S., Cipriani A., Spineli L., Mavridis D., Orey D., Richter F., Samara M., Barbui C., Engel R.R., Geddes J.R., Kissling W., Stapf M.P., Lässig B., Salanti G., Davis J.M. (2013) Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet*. **382**(9896), 951–962. [https://doi.org/10.1016/S0140-6736\(13\)60733-3](https://doi.org/10.1016/S0140-6736(13)60733-3)
  18. Meltzer H.Y., Elkis H., Vanover K., Weiner D.M., van Kammen D.P., Peters P., Hacksell U. (2012) Pimavanserin, a selective serotonin (5-HT)<sub>2A</sub>-inverse agonist, enhances the efficacy and safety of risperidone, 2 mg/day, but does not enhance efficacy of haloperidol, 2 mg/day: comparison with reference dose risperidone, 6 mg/day. *Schizophr. Res.* **141**(2–3), 144–152. <https://doi.org/10.1016/j.schres.2012.07.029>
  19. Baldessarini R.J., Tarazi F.I. 2001. Drugs and the treatment of psychiatric disorders: antipsychotic and anti-manic agents. In *Goodman and Gilman's the Pharmacological Basis of Therapeutics*. 10th ed. Hardman J.G., Limbird L.E., Gilman A.G., Eds. New York: McGraw-Hill Press, 485–528.
  20. Uno Y., Coyle J.T. (2019) Glutamate hypothesis in schizophrenia. *Psychiatry Clin. Neurosci.* **73**(5), 204–215. <https://doi.org/10.1111/pcn.12823>
  21. Egerton A., Grace A.A., Stone J., Bossong M.G., Sand M., McGuire P. (2020) Glutamate in schizophrenia: neurodevelopmental perspectives and drug development. *Schizophr. Res.* **223**, 59–70. <https://doi.org/10.1016/j.schres.2020.09.013>
  22. Kay S.R., Fiszbein A., Opler L.A. (1987) The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* **13**(2), 261–276.
  23. Sambrook J., Fritsch E.R., Maniatis T. (1989) *Molecular cloning: a laboratory manual* 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Lab. Press.
  24. Bakker P.R., Al Hadithy A.F., Amin N., van Duijn C.M., van Os J., van Harten P.N. (2012) Antipsychotic-induced movement disorders in long-stay psychiatric patients and 45 tag SNPs in 7 candidate genes: a prospective study. *PLoS One*. **7**(12), e50970. <https://doi.org/10.1371/journal.pone.0050970>
  25. Liang W., Yu H., Su Y., Lu T., Yan H., Yue W., Zhang D. (2020) Variants of *GRM7* as risk factor and response to antipsychotic therapy in schizophrenia. *Transl. Psychiatry*. **10**(1), 83. <https://doi.org/10.1038/s41398-020-0763-4>
  26. Yang X., Wang G., Wang Y., Yue X. (2015) Association of metabotropic glutamate receptor 3 gene polymorphisms with schizophrenia risk: evidence from a meta-analysis. *Neuropsychiatr. Dis. Treat.* **11**, 823–833. <https://doi.org/10.2147/NDT.S77966>
  27. Peuskens J., Pani L., Detraux J., De Hert M. (2014) The effects of novel and newly approved antipsychotics on serum prolactin levels: a comprehensive review. *CNS Drugs*. **28**(5), 421–453.
  28. Patil S.T., Zhang L., Martenyi F., Lowe S.L., Jackson K.A., Andreev B.V., Avedisova A.S., Bardenstein L.M., Gurovich I.Y., Morozova M.A., Mosolov S.N., Neznanov N.G., Reznik A.M., Smulevich A.B., Tochilov V.A., Johnson B.G., Monn J.A., Schoepp D.D. (2007) Activation of mGlu2/3 receptors as a new approach to treat schizophrenia: a randomized phase 2 clinical trial. *Nat. Med.* **13**, 1102–1107.
  29. Mössner R., Schuhmacher A., Schulze-Rauschenbach S., Kühn K.U., Rujescu D., Rietschel M., Zobel A., Franke P., Wölwer W., Gaebel W., Häfner H., Wagner M., Maier W. (2008) Further evidence for a functional role of the glutamate receptor gene *GRM3* in schizophrenia. *Eur. Neuropsychopharmacol.* **18**(10), 768–772. <https://doi.org/10.1016/j.euroneuro.2008.05.007>
  30. Norton N., Williams H.J., Dwyer S., Ivanov D., Preece A.C., Gerrish A., Williams N.M., Yerassimou P., Zammit S., O'Donovan M.C., Owen M.J. (2005) No evidence for association between polymorphisms in *GRM3* and schizophrenia. *BMC Psychiatry*. **5**, 23. <https://doi.org/10.1186/1471-244X-5-23>
  31. Tochigi M., Suga M., Ohashi J., Otowa T., Yamasue H., Kasai K., Kato T., Okazaki Y., Kato N., Sasaki T. (2006) No association between the metabotropic glutamate receptor type 3 gene (*GRM3*) and schizophrenia in a Japanese population. *Schizophr. Res.* **88**(1–3), 260–264. <https://doi.org/10.1016/j.schres.2006.07.008>
  32. Chaumette B., Sengupta S.M., Lepage M., Malla A., Iyer S.N., Kebir O.; ICAAR study group, Dion P.A., Rouleau G.A., Krebs M.O., Shah J.L., Joober R. (2020) A polymorphism in the glutamate metabotropic receptor 7 is associated with cognitive deficits in the early phases of psychosis. *Schizophr. Res.* **S0920-9964**(20), 30371–30376. <https://doi.org/10.1016/j.schres.2020.06.019>
  33. Poltavskaya E.G., Fedorenko O.Y., Kornetova E.G., Loonen A.J.M., Kornetov A.N., Bokhan N.A., Ivanova S.A. (2021) Study of early onset schizophrenia: associations of *GRIN2A* and *GRIN2B* polymorphisms. *Life (Basel)*. **11**(10), 997. <https://doi.org/10.3390/life11100997>
  34. Корнетова Е.Г., Иванова С.А., Тигунцев В.В., Бойко А.С., Лобачёва О.А., Корнетов А.Н., Семке А.В. (2019) Антипсихотик-индуцированная гиперпролактинемия у больных шизофренией: социальные, клинические, иммунологические и терапевтические особенности. Томск: Интегральный Переплет.
  35. Янковская А.Г. (2017) Уровень пролактина у женщин с шизофренией на ранних этапах заболевания в условиях психофармакотерапии. *Журн. гродненского гос. мед. университета*. **15**(4), 437–441.
  36. Mittal S., Prasad S., Ghosh A. (2017) Antipsychotic-induced hyperprolactinaemia: case studies and review. *Postgrad. Med. J.* **0**, 1–4. <https://doi.org/10.1136/postgradmedj-2017-135221>

## Association of the Level of Serum Prolactin and Polymorphic Variants of the *GRIN2A*, *GPM3*, *GPM7* GENES in Patients with Schizophrenia Taking Conventional and Atypical Antipsychotics

V. V. Tiguntsev<sup>1, \*</sup>, V. I. Gerasimova<sup>1</sup>, E. G. Kornetova<sup>1</sup>, O. Yu. Fedorenko<sup>1</sup>, A. N. Kornetov<sup>2</sup>, A. A. Goncharova<sup>1</sup>, E. G. Poltavskaya<sup>1</sup>, and A. S. Boyko<sup>1</sup>

<sup>1</sup>Mental Health Research of the Tomsk National Research Medical Center, Tomsk, 634014 Russia

<sup>2</sup>Siberian State Medical University, Tomsk, 634050 Russia

\*e-mail: cristall2009@live.ru

Dopamine, serotonin and glutamate systems are jointly involved in the pathogenesis and pharmacotherapy of schizophrenia. We formulated a hypothesis that polymorphic variants of the *GRIN2A*, *GRM3*, and *GRM7* genes may be associated with the development of hyperprolactinemia in patients with schizophrenia taking conventional and atypical antipsychotics as basic treatment. 432 Caucasian patients diagnosed with schizophrenia were examined. DNA was isolated from peripheral blood leukocytes by the standard phenol-chloroform method. For pilot genotyping, 12 SNPs in the *GRIN2A* gene, 4 SNPs in the *GRM3* gene, and 6 SNPs in the *GRM7* gene were selected. Allelic variants of the studied polymorphisms were determined by real-time PCR. The level of prolactin was determined by enzyme immunoassay. Among persons taking conventional antipsychotics, there were statistically significant differences in the distribution of genotype and allele frequencies in groups of patients with normal and elevated prolactin levels for the *GRIN2A* rs9989388 and *GRIN2A* rs7192557 polymorphic variants, as well as differences in serum prolactin levels depending on the genotypes of the *GRM7* rs3749380 polymorphic variant. Among persons taking atypical antipsychotics, statistically significant differences were found in the frequencies of genotypes and alleles of the *GRM3* rs6465084 polymorphic variant. For the first time, an association of polymorphic variants of the *GRIN2A*, *GRM3*, and *GRM7* genes with the development of hyperprolactinemia in patients with schizophrenia taking conventional and atypical antipsychotics has been established. They not only confirm the close connection of the dopaminergic, serotonergic, and glutamatergic systems in the development of schizophrenia, but also demonstrate the potential of taking into account the genetic component for its therapy.

**Keywords:** prolactin, glutamate, single nucleotide polymorphisms, real-time PCR, ELISA, human, antipsychotics, molecular genetics, schizophrenia

ГЕНОМИКА.  
ТРАНСКРИПТОМИКА

УДК 575.113

СТРУКТУРА И ЭВОЛЮЦИЯ ГЕНА *AqE* У НАСЕКОМЫХ

© 2023 г. Л. В. Пузакова<sup>а</sup>, \*, М. В. Пузаков<sup>а</sup>

<sup>а</sup>Институт биологии южных морей им. А.О. Ковалевского Российской академии наук,  
Севастополь, 299011 Россия

\*e-mail: [kvluda@yandex.ru](mailto:kvluda@yandex.ru)

Поступила в редакцию 21.04.2022 г.

После доработки 07.07.2022 г.

Принята к публикации 24.07.2022 г.

Ген *AqE*, кодирующий сульфолатдегидрогеназа-подобный фермент семейства LDH2/MDH2-оксидоредуктаз, найден в геномах бактерий и грибов, животных и растений, образ жизни которых связан с водной средой, а также членистоногих, в частности, насекомых, которые ведут преимущественно сухопутный образ жизни. Чтобы проследить эволюционную судьбу гена *AqE*, мы изучили распространённость этого гена и его структуру в классе насекомых. Обнаружено, что ген *AqE* представлен не во всех отрядах/подотрядах насекомых, наблюдается потеря гена. В некоторых отрядах он дублирован или мультиплицирован. Установлена вариабельность гена как по длине, так и по экзон-интронной структуре – от безинтронной до многоинтронной. Обнаружено, что мультипликация гена *AqE* насекомых имеет древнюю природу, однако есть и “молодые” дубликации. Не исключено, что в связи с появлением паралогов ген приобрёл новую функцию.

**Ключевые слова:** эволюция генов, дубликация, мультипликация, потеря генов, насекомые, сульфолатдегидрогеназа

**DOI:** 10.31857/S0026898423010135, **EDN:** AXMDWN

ВВЕДЕНИЕ

Семейство LDH2/MDH2-оксидоредуктаз (LDH/MDH-оксидоредуктазы типа 2) включает ферменты, которые участвуют в разнообразных метаболических процессах и на основании филогенетического анализа подразделяются на семь групп [1–4]. Охарактеризованы функции ферментов клады ComC (*L*-сульфолатдегидрогеназа, SLDH), клады LDH (лактатдегидрогеназа типа 2), клады AllD (уреидогликолатдегидрогеназа), клады YiaK (2,3-дикето-*L*-гулонатредуктаза) и клады DpkA ( $\Delta$ 1-пиперидин-2-карбоксилат/ $\Delta$ 1-пирролин-2-карбоксилатредуктаза), в то время как функция белков трех других клад: термофильных архей, YbiC и YbiS, остается неизвестной. Поскольку в филогенетический анализ было включено только 20 аминокислотных последовательностей [4], то, возможно, в дальнейшем такая классификация будет пересмотрена.

Функция LDH2/MDH2-оксидоредуктаз на данный момент охарактеризована только у архей и бактерий [3, 4]. Недавно у эукариот выявили гены, которые кодируют белки, гомологичные SLDH архей [5]. У архей и бактерий этот фермент кодируется геном *comC* [6–8]. *ComC*-подобные гены найдены в геномах грибов, растений и животных. Однако в царстве растений гомологи *comC* обнаружены только у водорослей, тогда как у

мхов и семенных растений они отсутствуют. Гомологи гена выявлены у преобладающего числа эволюционных групп животных, за исключением четвероногих (земноводных, рептилий, птиц и млекопитающих) [5]. *ComC*-подобные гены эукариот условно названы *AqE* (aquatic enzyme, водный фермент) в связи с его встречаемостью преимущественно в тех таксонах, образ жизни или стадии онтогенеза которых ассоциированы с водной (жидкой) средой. Подробный анализ встречаемости и структуры гена *AqE* у костистых рыб показал, что в большинстве исследованных отрядов сохраняется идентичная экзон-интронная организация (11 экзонов) и характерные консервативные мотивы этого гена [9]. Однако у карпообразных (Cypriniformes) этот ген утерян полностью, а у представителей лососеобразных (Salmoniformes) претерпел значительные делеции и, по всей видимости, нефункционален [9]. Ранее предполагалось, что SLDH-подобный белок, кодируемый геном *AqE*, участвует в анаэробном энергообмене и его утрата отдельными таксонами как животных, так и растений связана с сопутствующим выходу на сушу перестройкой метаболизма, обусловленной насыщенностью среды кислородом и отсутствием естественной гипоксии, характерной для гидробионтов [5]. Утрата гена *AqE* у карпообразных и лососеобразных мог-

ла произойти из-за индивидуальной эволюции этих таксонов и соответствующей реструктуризации метаболических путей. Известно, что у предков *Cypriniformes* и *Salmoniformes* происходили независимые полногеномные дубликации (так называемая четвертая полногеномная дубликация позвоночных) [10, 11]. Избыток оксидоредуктаз, возникший в результате дубликации генома, мог стать причиной потери значимости гена *AqE* для метаболизма, так как появился большой “ресурс” для негомологичной замены или появления новых ферментов и новых метаболических путей [9]. Анализ дифференциальной экспрессии *AqE* желтого горбыля (*Larimichthys crocea*) показал, что уровень активности этого гена высок в печени, жабрах и коже, тогда как в других тканях он низкий. В связи с этим предположили, что функция SLDH-подобного белка рыб может быть связана с выведением продуктов метаболизма [12].

Среди таксонов, представители которых сохранили ген *AqE*, большой интерес представляют насекомые, поскольку они являются сухопутными организмами. Насекомые – очень разнообразный класс организмов, обитающих в различных условиях, в том числе в средах, где могут сталкиваться с дефицитом кислорода. Так, например, представители отрядов Trichoptera, Megaloptera, Odonata проходят личиночную стадию развития в водной среде и дышат растворенным в воде кислородом. Также представители некоторых отрядов насекомых развиваются в почве – Auchenorrhyncha, Isoptera, Diplura, Protura, в древесине – Zoraptera, Hymenoptera (например *Orussus abietinus*) или Polyphaga (например *Dendroctonus ponderosae*, *Anoplophora glabripennis*), в разлагающихся органических субстратах – некоторые представители отряда Diptera. Паразитические виды, такие как Hymenoptera, Strepsiptera, живут в тканях животных, используя для дыхания кислород, растворенный в гемолимфе. Есть среди насекомых и экстремофилы, обитающие в холодных условиях, около ледников – Grylloblattodea.

Нами изучено разнообразие, структура и эволюция гена *AqE* в классе насекомых с таксономическим разрешением до отряда. Присутствие гена *AqE* в геноме насекомых, в частности комара *Aedes albopictus* (отряд Diptera), было отмечено ранее [5], однако детально этот ген изучен впервые.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Поиск генов *AqE* у насекомых.** Поиск гомологов гена *AqE* у насекомых проводили с использованием алгоритма локального выравнивания tBLASTn [13]. В качестве референсной последовательности выбрали аминокислотную последовательность, кодируемую *AqE* комара *A. albopictus*. Эта последовательность была идентифицирована при первичном исследовании генов *AqE* [5]. Проанализированы

полногеномные ДНК представителей 17 отрядов. Еще 10 отрядов исследованы с помощью поиска гомологий в базе данных транскриптомов. Используются геномные и транскриптомные последовательности, размещенные на серверах NCBI. Уникальные идентификаторы полногеномных и транскриптомных сборок представителей класса насекомых, а также таксонов Diplura и Protura приведены в табл. 1.

Границы экзонов у обнаруженных генов *AqE* уточняли визуально по наибольшему сходству последовательностей между референсной и исследуемой последовательностями и наличию 5'- и 3'-границы сайта сплайсинга. Кодирующие последовательности *AqE*, полученные в результате анализа, использовали для поиска транскрибируемых последовательностей РНК. Копиям *AqE* в случае выявления более одной в геноме, присваивали порядковый номер в соответствии с очередностью обнаружения.

**Филогенетический анализ.** Множественное выравнивание аминокислотных последовательностей выполняли с помощью MUSCLE [14], а полученные данные использовали для построения филогенетического дерева с использованием программного обеспечения MEGA X [15] и метода максимального правдоподобия (ML). Использовали следующие параметры: бутстреп, 100 повторов; модель замещения аминокислот Ле-Гаскуэль-2008; и гамма-распределение. Анализ включал 91 аминокислотную последовательность *AqE* под-типа Hexapoda.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

### *Представленность гена *AqE* в геномах насекомых*

В ходе исследования гена *AqE* мы проанализировали 51 вид насекомых, представляющих 27 отрядов. Полногеномные последовательности ДНК на момент исследования были получены только у представителей 17 анализируемых отрядов. Другие 10 отрядов мы смогли проанализировать, используя базу данных транскриптомов. В качестве референсного белка (шаблона) мы использовали аминокислотную последовательность, кодируемую *AqE* комара *A. albopictus*. Этот белок был выбран в качестве референсного, поскольку экзон-интронная организация его гена имеет высокое сходство с классическим *AqE* [5, 9]. В результате установлена значительная вариабельность присутствия, структуры и количества копий *AqE* у представителей насекомых (табл. 1). Нами определены три основные тенденции: (1) сохранение единственной копии *AqE*; (2) полная потеря *AqE*; и (3) мультипликация *AqE*, варьирующая у разных видов от 2 до 9 копий.

Группы, которые потеряли весь ген, представлены в меньшинстве (табл. 1). Это отряд Strepsiptera

Таблица 1. Характеристики вариантов гена *AqE* у представителей подтипа *Hexaroda*

Таксон	Отряд	Представитель/ идентификатор геномной или транскриптомной сборки	Ген	Количество экзонов	Блок, а.о.	Перекрытие с РП, %	Идентичность к РП, %	Дополнительная информация	Транскрипт*	
Monocniphya	Archaeognatha	<i>Machilis hrabei</i> / GAUM000000000	<i>AqE1</i>	ND	363	97	59.89		Есть	
			<i>AqE1</i>	ND	356	96	56.57		Есть	
			<i>Xenocantopsis brachyvenus</i> / GCA_900249655	1	364	96	61.02		НО	
	Orthoptera		<i>Laupala kohalensis</i> / GCA_002313205	<i>AqE1</i>	5	313	94	50.74	ДТ1	НО
				<i>AqE2</i>	6	309	96	42.65	Делеция С-конца, 10 а.о. ДТ1 Стоп-колон в ОРС	НТ Orthoptera – есть
				<i>AqE3</i>	1	48	12	55.56	Сохранился короткий фрагмент	
				<i>AqE1</i>	1	351	94	33.33	Есть полиА-хвост и прямые повторы	НО
				<i>AqE2</i>	1	350	96	37.7	Старт-колон не обнаружен	НО
				<i>AqE3</i>	1	363	97	39		НО
Polyneoptera		<i>Blattella germanica</i> / GCA_003018175	<i>AqE4</i>	1	328	89	33.6	Старт-колон не обнаружен. Делеция N-конца, 30 а.о.	НО	
			<i>AqE5</i>	4	336	97	50.7	ДТ1	Есть	
			<i>AqE6</i>	4	358	98	62.64	Старт-колон не обнаружен	Есть	
			<i>AqE7</i>	6	343	91	57.83		Есть	
			<i>AqE8</i>	3	150	41	42.86	Протяженная делеция N-конца, 158 а.о. Делеция С-конца, 55 а.о. Старт- и стоп-колоны не обнаружены	Есть	
			<i>AqE9</i>	5	276	82	47	ДТ1 Делеция С-конца, 60 а.о	Есть	
			<i>AqE1</i> isozyme 1	10	381	96	65.34	Удлинение N-конца на 18 а.о.	НТ Isoptera – НО	
			<i>AqE1</i> isozyme 2	9	354	96	65.34		НТ Isoptera – НО	
			<i>Zootermopsis nevadensis</i> / GCA_000696155	9	354	96	64		НТ Isoptera – НО	

Таблица 1. Продолжение

Таксон	Отряд	Представитель/ идентификатор геномной или транскриптомной сборки	Ген	Количество экзонов	Белок, а.о.	Перекрытие с РП, %	Идентичность к РП, %	Дополнительная информация	Транскрипт*	
Mantodea		<i>Metallyticus splendidus</i> /GATV000000000	<i>AqE1</i>	ND	360	96	59.71		Есть	
		<i>Stagmatoptera biocellata</i> /GDBT000000000	<i>AqE1</i>	ND	365	96	61.36		Есть	
Plasmatodea			Нет данных по таксону						НТ	
Manoptorhasmato- dea			Нет данных по таксону						НО	
Grylloblattode			Нет данных по таксону						НО	
Dermaptera		<i>Diplatys sp.</i> /GDSCS010000000	<i>AqE1</i>	ND	355	96	58.36		Есть	
			<i>AqE1</i>	ND	355	98	51.12		Есть	
		Heteroptera	<i>Cimex lectularius</i> / GCA_000648675	<i>AqE1</i> isozyme 1	10	368	96	56.82		Есть
				<i>AqE1</i> isozyme 2	9	356	96	56.82		Есть
Sternorrhyncha			НО					НО		
Hemiptera		Coleorrhyncha	Нет данных по таксону						НО	
			<i>AqE1</i>	ND	356	96	66		Есть	
		<i>Auchenorrhyncha</i>	<i>AqE1</i>	8	337	92	64.29		НТ <i>Auchenorrhyncha</i> – есть	
			<i>AqE2</i>	6	301	95	47.67	ДП1	НТ <i>Auchenorrhyncha</i> – есть	
		<i>Badonnelia titel</i> / GDFG010000000	<i>AqE1</i>	ND	361	97	67.8		Есть	
Psocoptera	<i>Elipocus karitensis</i> / GDFL010000000	<i>AqE1</i>	ND	358	98	61.58		Есть		
		<i>AqE2</i>	ND	357	97	61.13		Есть		
Phthiraptera		<i>Pediculus humanus</i> /GCA_000006295	<i>AqE1</i>	6	345	92	64.09	Старт-колон не обнаружен	НТ <i>Phthiraptera</i> – есть	
Thysanoptera		<i>Frankliniella occidentalis</i> / GCA_000697945	<i>AqE1</i>	8	344	96	59.26	Старт-колон не обнаружен	Есть	

Paraneoptera/Condylognatha

Таблица 1. Продолжение

Таксон	Отряд	Представитель/ идентификатор геномной или транскриптомной сборки	Ген	Количество экзонов	Блок, а.о.	Перекрытие с РП, %	Идентичность к РП, %	Дополнительная информация	Транскрипт*	
Holometabola	Trichoptera		<i>AqE1</i>	7	360	96	56.23	Делеция в ОРС, 16 а.о	НТ Trichoptera – есть	
			<i>AqE1</i>	8	360	96	56		Есть	
	Lepidoptera		<i>AqE2</i>	8	356	97	56.09		Есть	
			<i>AqE1</i>	3	437	98	69.47	Удлинение N-конца на 80 а.о	НТ Diptera – есть	
	Diptera		<i>Drosophila busckii</i> /GCA_001277935	<i>AqE2</i>	3	437	98	69.47	Удлинение N-конца на 80 а.о	НТ Diptera – есть
				<i>AqE3</i>	3	437	98	69.47	Удлинение N-конца на 80 а.о	НТ Diptera – есть
				<i>AqE4</i>	1	396	99	38.89	Удлинение N-конца на 30 а.о.	НТ Diptera – есть
				<i>AqE5</i>	1	474	95	35.80	Удлинение N-конца на 112 а.о. Есть полиА-хвост и прямые повторы	НТ Diptera – есть
				<i>AqE1</i>	3	440	98	69.92	Удлинение N-конца на 81 а.о.	Есть
				<i>AqE2</i>	1	384	97	40.23	Удлинение N-конца на 25 а.о.	Есть
				<i>AqE3</i>	1	464	98	38.44	Удлинение N-конца на 103 а.о. Есть короткий полиА-хвост и прямые повторы	Есть
				<i>AqE1</i>	5	486	100	98.62	Удлинение N-конца на 123 а.о.	Есть
				<i>AqE1</i>	5	487	100	99.72	Удлинение N-конца на 124 а.о.	Есть
				<i>AqE2</i>	4	363	100	100		Есть
	Siphonaptera		<i>Ctenocephalides felis</i> /GCA_003426905	<i>AqE3</i>	5	443	100	99.72	Удлинение N-конца на 80 а.о.	Есть
				<i>AqE1</i>	5	501	98	83.57	Удлинение N-конца на 139 а.о.	НО
				<i>AqE2</i>	4	364	98	83.57		НО
				<i>AqE1</i>	5	403	97	55	Удлинение N-конца на 37 а.о.	Есть
	Mecoptera		<i>Nannochorista philipoti</i> /GADB00000000	<i>AqE2</i>	4	297	68	59	Делеция С-конца 104 а.о. Удлинение N-конца на 37 а.о.	Есть
				<i>AqE1</i>	5	391	97	51	Удлинение N-конца на 39 а.о. Делеция в ОРС – 16 а.о	Есть
<i>AqE1</i>				ND	368	97	72.47		Есть	
Megaloptera		<i>Boreus hyemalis</i> /GAYK00000000	<i>AqE1</i>	ND	370	98	68.27		Есть	
			<i>AqE1</i>	ND	366	96	62.50		Есть	

Таблица 1. Продолжение

Таксон	Отряд	Представитель/ идентификатор геномной или транскриптомной сборки	Ген	Количество экзонов	Белок, а.о.	Перекрытие с РП, %	Идентичность к РП, %	Дополнительная информация	Транскрипт*
Neuroptera		<i>Osmylius fulvicornis</i> /GAYC000000000	<i>AqE1</i>	ND	365	96	61.25		Есть
			<i>AqE2</i>	ND	387	96	53.85	Удлинение N-конца на 27 а.о. Старт- и стоп-кодоны не обнаружены	Есть
			<i>AqE1</i>	ND	355	95	40.57		Есть
			<i>AqE1</i>	ND	365	98	58.99		Есть
			<i>AqE1</i>	ND	364	98	57.58		Есть
Coleoptera		Polyphaga	HO						HO
			HO						HO
			<i>AqE1</i> isozyme 1	4	358	98	53.76	ДТ2	НТ Coleoptera – есть
			<i>AqE1</i> isozyme 2	4	322	85	54.19	ДТ2 Делеция С-конца, 53 а.о	НТ Coleoptera – есть
			<i>AqE2</i>	4	352	96	55.97	ДТ2	НТ Coleoptera – есть
			<i>AqE4</i>	7	391	97	60.97	Удлинение N-конца на 31 а.о.	НТ Coleoptera – есть
			<i>AqE1</i>	8	414	96	56.7	Удлинение N-конца на 42 а.о.	Есть
			<i>AqE1</i>	8	412	96	60.51	Удлинение N-конца на 47 а.о.	Есть
			<i>AqE2</i>	3	147	43	44.55	Делеция С-конца, 244 а.о. Удлинение N-конца на 26 а.о.	Есть
			<i>AqE3</i>	5	264	63	41.99	Делеция N-конца, 125 а.о. Удлинение С-конца на 19 а.о.	Есть
<i>AqE4</i>	8	363	96	54.13		Есть			
<i>AqE5</i> isozyme 1	8	363	96	56.53		Есть			
<i>AqE5</i> isozyme 2	7	312	88	57.14	Делеция N-конца, 52 а.о	Есть			

Таблица 1. Окончание

Таксон	Отряд	Представитель/ идентификатор геномной или транскриптомной сборки	Ген	Количество экзонов	Блок, а.о.	Перекрытие с РП, %	Идентичность к РП, %	Дополнительная информация	Транскрипт*	
Strepsiptera		<i>Lerptinoiarsa decemlineata</i> /GCA_000500325	<i>AqE1</i>	3	166	45	44.58	Делеция N-конца, 193 а.о	Есть	
			<i>AqE2</i>	7	409	59	52.94	Негомологичный фрагмент на N-конце, 97 а.о. Негомологичный фрагмент на C-конце – 125 а.о	Есть	
			<i>AqE3</i>	7	302	71	65.61	Негомологичный фрагмент на N-конце, 56 а.о. Делеция в ОРС, 64 а.о.	Есть	
		<i>Mengilla moldrzyki</i>								
			<i>Nasonia vitripennis</i> /GCA_000002325	<i>AqE1</i>	2	412	97	54.24	Удлинение N-конца на 49 а.о.	Есть
			<i>Apis mellifera</i> /GCA_000002195	<i>AqE1</i>	2	434	97	48.59	Удлинение N-конца на 72 а.о.	Есть
			<i>Cephus cinctus</i> /GCA_000341935	<i>AqE1</i>	2	395	97	53.26	Удлинение N-конца на 32 а.о.	Есть
			<i>Onuscus abietinus</i> /GCA_000612105	<i>AqE1</i>	2	403	97	51.13	Удлинение N-конца на 32 а.о.	Есть
			<i>Neodiprion lecontei</i> /GCA_001263575	<i>AqE1</i>	2	392	98	53.98	Негомологичный фрагмент на N-конце, 26 а.о.	Есть
			Data on taxon unavailable							
Polyneoptera			<i>AqE1</i>	ND	359	96	64.10		Есть	
			<i>AqE1</i>	ND	359	97	64.02	Старт-кодон не обнаружен	Есть	
			<i>AqE1</i>	ND	357	96	62.39	Короткий скэффолд – C-конец не обнаружен	Есть	
Zoraptera			<i>AqE1</i>	ND	382	96	57.67	Удлинение C-конца на 20 а.о.	Есть	
			<i>AqE2</i>	ND	370	98	42.62		Есть	
			<i>AqE3</i>	ND	380	97	35.03	Удлинение N-конца на 19 а.о., C-конца на 12 а.о.	Есть	
Odonata			<i>AqE1</i>	7	334	97	54.81	ДТ1, Делеция C-конца, 17 а.о.	Есть	
			<i>AqE2</i>	6	306	94	46.65	Старт-кодон не обнаружен. ДТ1, Делеция C-конца, 19 а.о.	Есть	
Ephemeroptera			<i>AqE1</i>	8	347	94	60.82	Делеция C-конца, 19 а.о.	Есть	
			<i>AqE1</i>	ND	358	96	53.28		Есть	
Diptera			<i>AqE2</i>	ND	357	97	53.26		Есть	
			<i>AqE1</i>	ND	367	97	51.97		Есть	
Protura			<i>AqE1</i>	ND	367	97	51.97		Есть	
			<i>AqE1</i>	ND	367	97	51.97		Есть	

\* По результатам анализа базы данных транскриптомов.

Примечания. РП – референсная последовательность; НД – нет данных (ген найден через анализ транскриптома); НО – не обнаружено; НТ – нет транскриптома; ДТ1 – делеция типа 1: “KRSNHUIGIAGWYTLRAMNAGCVGMSMTNTSPASPTRSKEA”; ДТ2 – делеция типа 2: “RTGHPRTDANLAF”; ОРС – открытая рамка считывания.

(Holometabola), подотряды Adepnaga, Archostemata (Coleoptera, Holometabola) и подотряд Sternorrhyncha (Hemiptera, Condylgnatha). В подотряде Sternorrhyncha в 11 просмотренных сборках не выявлено гомологий с *AqE* – среди них Aphidomorpha, Aleyrodoidea, Psylloidea. У *Dactylopius coccus*, относящегося к Dactylopiidae, обнаружены два гена *AqE*, но при дополнительной проверке выяснилось, что оба гена – это фрагменты бактериальной ДНК, вероятно попавшей в пробу. Этот факт подтверждается высоким сходством кодируемых белков *D. coccus* с белками бактерий: *AqE1 D. coccus* с *AqE Bosea* sp. (идентичность 72.93% (CP014301.1)) и *AqE2 D. coccus* с *AqE Ochrobactrum anthropi* (идентичность 98.50% (LT671862.1)). При этом наблюдается низкая идентичность к референсному *AqE* комара *A. albopictus*: *AqE1* (29.29%) и *AqE2* (29.64%).

У представителей отрядов Mantophasmatodea, Grylloblattodea, Embioptera и подотряда Coleorrhyncha изучены только последовательности транскриптомов. Не выявлено транскриптов, сходных с транскриптом, кодирующим референсный белок *AqE*. Однако в этом случае нельзя сделать однозначный вывод о потере гена.

Остальные группы распределились следующим образом – представители 10 отрядов и одного подотряда содержат единственную копию гена, в 15 отрядах и одном подотряде ген мультиплицирован (табл. 1). Разные виды тех отрядов, где *AqE* мультиплицирован, имеют разное число копий гена. Так, например, у представителя отряда Orthoptera (Polyneoptera) *Xenocatantops brachycerus* одна копия гена *AqE*, а у *Laupala kohalensis* – три. В отряде Diptera *A. aegypti* имеет одну копию гена, а *A. albopictus* – три. В отряде Coleoptera (Polyphaga) у *Dendroctonus ponderosae* одна копия *AqE*, у *Tribolium castaneum* и *Leptinotarsa decemlineata* – три, у *Anoplophora glabripennis* – пять. Некоторые виды содержат значительное число копий гена *AqE*, так в геноме *Blattella germanica* число копий достигает 9.

#### Анализ транскриптомных баз данных насекомых

Для качественной оценки потенциальной функциональности генов *AqE* насекомых проведен анализ транскриптомных баз данных NCBI. Транскрипты гена *AqE* обнаружены в большинстве случаев. У представителей некоторых отрядов транскрипты гена *AqE* не были найдены. Поэтому предприняли попытки обнаружить транскрипты гена *AqE* у других представителей этих отрядов и в ряде случаев получили положительный результат. В некоторых случаях транскрипты не обнаружены ни у анализируемых видов, ни в таксоне в целом. Данные для некоторых организмов в базе TSA отсутствовали (табл. 1).

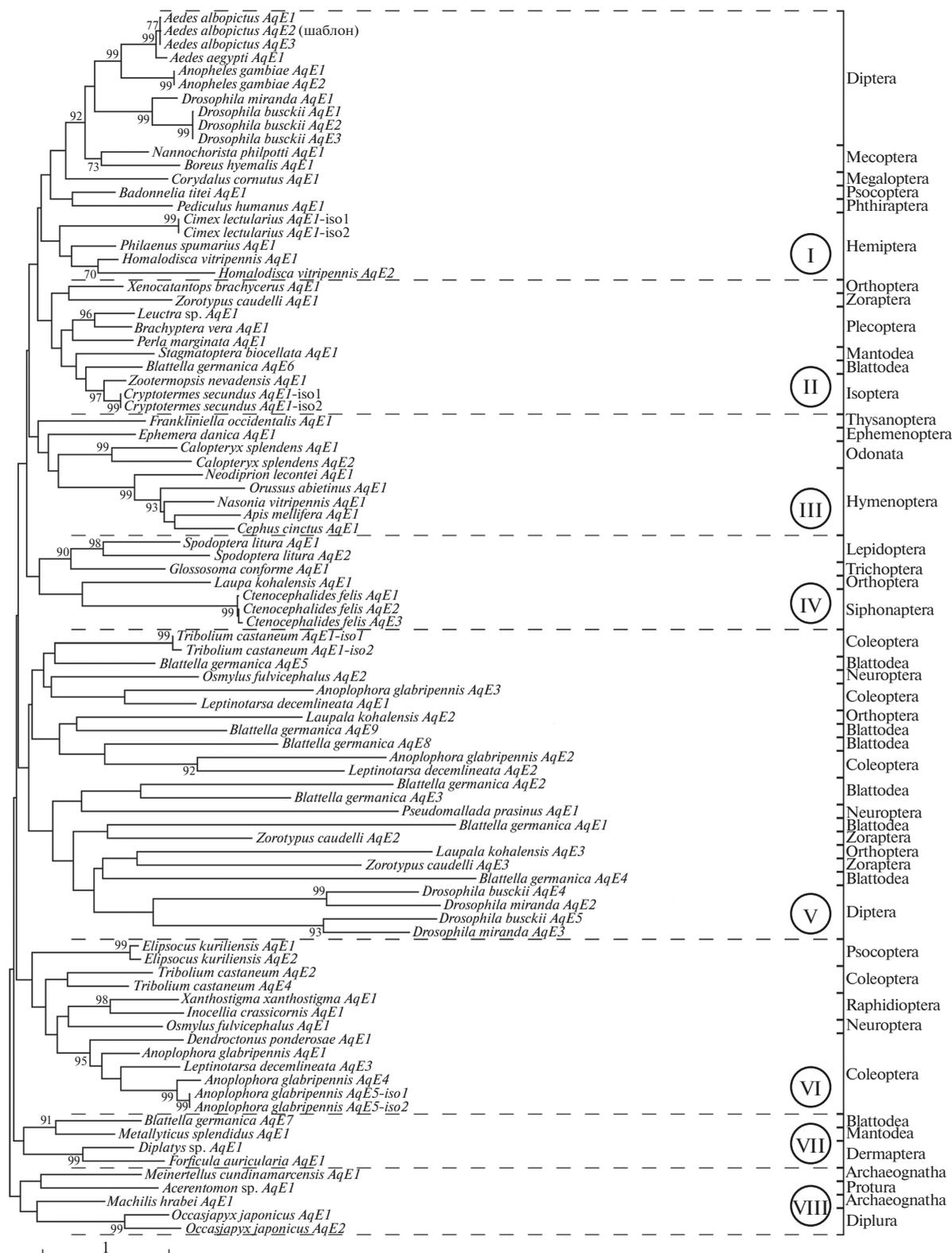
Рассмотрение структуры гена *AqE* показывает, что в большинстве случаев этот ген потенциально функционален. В одном случае (*L. kohalensis AqE2*) открытая рамка считывания (ОРС) содержит стоп-кодон, следовательно, ген нефункционален. В некоторых случаях *AqE* представлен сильно де-летированной копией, либо небольшим фрагментом (*L. kohalensis AqE3* (144 п.н./48 а.о.) или *B. germanica AqE8* (450 п.н./150 а.о.)). Очевидно, что такие фрагменты не могут быть функциональными. Однако не исключена возможность присутствия транскриптов даже при нефункциональности гена. Так, в базах TSA обнаружены транскрипты *An. glabripennis AqE2* протяженностью 441 п.н./47 а.о. или *An. glabripennis AqE3* (738 п.н./246 а.о.). Известно еще несколько подобных случаев (табл. 1). Кроме того, отсутствие старт-кодона, выявленное в некоторых случаях, мы связываем с ошибками при поиске (низкая идентичность). Например, у *B. germanica AqE6* не обнаружен метионин, однако транскрипты представлены в базе TSA (табл. 1).

#### Филогенетические взаимоотношения генов *AqE* насекомых

Чтобы установить взаимосвязь между генами *AqE* насекомых, проведен филогенетический анализ, основанный на методе максимального правдоподобия, в который была включена 91 аминокислотная последовательность, кодируемая этими генами. Полученное дерево не позволило выявить четкие эволюционные отношения между различными вариантами *AqE*, потому что статистическая значимость преобладающей доли визуально отличных клад не превышала 70% (рис. 1). Тем не менее, мы разделили дерево по топологии на восемь условных кластеров и установили некоторые закономерности.

Так, например, в каждом из кластеров I–IV и VI–VIII сгруппированы гены организмов, не встречающиеся ни в одном из остальных шести кластеров. (кроме *B. germanica*). При этом одна группа (кластер V) отличается визуально более длинными эволюционными расстояниями и содержит гены *AqE* организмов, которые представлены также в других группах (кластеры I, II, VI, VII). Единственное исключение – это ген *Pseudomallada prasinus* (отряд Neuroptera). Гены *AqE* этого насекомого не попадают ни в один из других семи кластеров (I–IV и VI–VIII). Однако у одного представителя отряда Neuroptera (*Osmylus fulvicephalus*) гены *AqE* представлены как в кластере V, так и в кластере VI.

Кластеры I–IV и VI–VIII включают *AqE*, которые эволюционируют медленнее и присутствуют практически у всех изученных организмов, а кластер V содержит гены, эволюционирующие с более высокой скоростью, и многие группы, по-ви-



**Рис. 1.** Эволюционные взаимоотношения генов *AqE* насекомых. Филогенетический анализ выполнен в программе MEGA X с помощью метода максимального правдоподобия. Приведены только бутстреп-значения, превышающие 50%. Справа от дендрограммы указаны отряды, к которым принадлежат соответствующие организмы.

димому, утратили этот вариант *AqE*. Гомологи гена *AqE* из кластера V выявлены у 10 видов: *T. castaneum* (*AqE1*), *B. germanica* (*AqE1*, *AqE2*, *AqE3*, *AqE4*, *AqE5*, *AqE8* и *AqE9*), *O. fulvicephalus* (*AqE2*), *An. glabripennis* (*AqE2* и *AqE3*), *Leptinotarsa decemlineata* (*AqE1* и *AqE2*), *L. kohalensis* (*AqE2* и *AqE3*), *P. prasinus* (*AqE1*), *Zorotypus caudelli* (*AqE2* и *AqE3*), *Drosophila busckii* (*AqE4* и *AqE5*) и *D. miranda* (*AqE2* и *AqE3*).

Другое важное явление — молодые дубликации, которые хорошо видны на филогенетическом дереве (рис. 1) и выявлены у четырех видов: *A. albopictus*, *An. gambiae*, *Stenocephalides felis* и *D. busckii*. Так, например, у комара *A. albopictus* выявлены три копии *AqE*. Сходство между аминокислотными последовательностями, кодируемыми этими генами, превышает 99%. Белковый продукт гена *AqE2* отличается одной аминокислотной заменой от белков, кодируемых *AqE1* и *AqE3*. Однако белки *AqE1* и *AqE3* содержат на N-конце дополнительный фрагмент длиной 124 и 80 а.о. соответственно. Еще один комар (*An. gambiae*) имеет две копии гена, идентичные примерно на 100%, при этом белковый продукт гена *AqE1* также содержит дополнительный N-концевой фрагмент (139 а.о.). У блохи *C. felis* обнаружены три гомолога гена *AqE*. Идентичность белков, кодируемых этими генами, колеблется от 93 до 97%, при этом присутствуют также вставки и делеции (табл. 1). У мухи *D. busckii* белковые продукты трех из пяти выявленных копий гена сходны на 99.77–100%.

#### Структурные различия генов *AqE* насекомых

Определена экзон-интронная структура ОРС каждого выявленного гена *AqE*. Результаты оказались неожиданными — гены *AqE* насекомых существенно варьировали по данному признаку. Количество экзонов у разных представителей колеблется от 1 до 10, при этом в геноме одного организма могут одновременно присутствовать гены *AqE* с разным количеством экзонов. Например, один из трех генов *D. miranda* трехэкзонный (входит в кластер I), а два одноэкзонных (кластер V). Или у *T. castaneum* два гена содержат четыре экзона (*AqE1*, *AqE2*), а один семь (*AqE4*), при этом *AqE1* входит в филогенетический кластер V, а *AqE2* и *AqE4* — в кластер VI.

Основная тенденция, наблюдаемая у насекомых, — сокращение числа интронов в гене *AqE*, согласуется с общей тенденцией в эволюции геномов. В большинстве линий эукариот эволюция связана, главным образом, с потерей интронов [16]. Классический *AqE* у рыб имеет 11 экзонов, причем это довольно консервативный признак [9]. Возможно, у рыб, которые обитают в более однородных условиях, чем насекомые, продукт этого гена участвует в некоторых ключевых реакциях

обмена и подвержен более строгому отбору. У насекомых строгость отбора была ослаблена, за короткий срок накопились значительные изменения в структуре этого гена. Это может быть связано с преобразованием функции фермента в тех условиях, в которых обитает и развивается конкретный вид, а разнообразие условий обитания насекомых очень велико: от жидкой среды (в которой личинка дышит кислородом, растворенным в воде, субстрате или межклеточной жидкости) и почвы до атмосферной среды, где нет дефицита кислорода.

Кроме того, встречаются варианты *AqE* с одним экзоном. Этот процесс может быть связан с ретрогенезацией — вставкой в геном ДНК, полученной в процессе обратной транскрипции с мРНК исходного гена и, соответственно, свободной от интронов. Проверка возможной ретрогенезации у видов, имеющих одноэкзонные копии *AqE*, показала следующее: у *B. germanica* только один (*AqE1*) из четырех одноэкзонных генов имеет полиА-хвост и прямые повторы, которые обычно свидетельствуют о произошедшем событии ретрогенезации. Остальные три одноэкзонные копии *B. germanica* таких структур не имеют. У одноэкзонного *AqE4* *D. busckii* нет ни полиА-хвоста, ни повторов, в отличие от *AqE5*, имеющего обе эти структуры. У *AqE2* *D. miranda* нет ни полиА-хвоста, ни повторов, у *AqE3* — повторы есть, но полиА-хвост короткий. Таким образом, *AqE*, имеющие прямые повторы и полиА-хвост, подверглись, очевидно, событию ретрогенезации. Что касается других безинтронных копий, то прямые повторы и полиА-хвост могут отсутствовать у более старых ретрогенов, но одновременное присутствие в геноме копий и с интронами, и без интронов предполагает, что безинтронные гены созданы путем ретрогенезации [17].

#### Модификации в аминокислотных последовательностях *AqE*

Множественное выравнивание белков *AqE* насекомых показало, что некоторые гены в процессе эволюции претерпели существенные модификации — делеции крайних или центральных зон, а также вставки добавочных участков. Протяженность *AqE* значительно варьирует, что связано, в основном, с делециями или вставками. Есть случаи, когда ген фрагментирован, но здесь мы, скорее всего, имеем дело с псевдогенезацией и остатком гена.

Копии *AqE*, имеющие дополнительный фрагмент, кодирующий 25–139 а.о. (табл. 1), представлены у всех рассмотренных нами представителей Diptera. У *D. busckii* все пять копий *AqE*, а у *D. miranda* все три копии *AqE* содержат дополнительный фрагмент. У *A. albopictus* и *An. gambiae* наряду с удлиненными копиями *AqE* есть и копии

без вставок. У представителя *C. felis* (Siphonaptera) все три гена *AqE* кодируют дополнительные фрагменты из 37 а.о. У *O. fulvicephalus* (Neuroptera) в начале гена *AqE2* находится участок, кодирующий 27 добавочных аминокислот. Вставки найдены и в *AqE4 T. castaneum* (Coleoptera), и в *AqE1 Dendroctonus ponderosae* (Coleoptera). У *An. glabripennis* в трех из пяти генов *AqE* (*AqE1*, *AqE2*, *AqE3*) также есть удлинения (табл. 1).

В некоторых отрядах выявлены гены *AqE* с делециями, локализованными в основном в двух участках (табл. 1). Так, первый вариант делеции в центральной части гена представлен в двух из трех генов *AqE L. kohalensis* (Orthoptera), в двух из девяти генов *B. germanica* (Blattodea), в одной из двух копий у *Homalodisca vitripennis* (Hemiptera). Эту же делецию содержат оба варианта *AqE* у *Calopteryx splendens* (Odonata). Все эти элементы распределены по дереву без всякой взаимосвязи друг с другом, следовательно, делеция, по-видимому, возникала независимо.

Второй тип делеции встречается в генах *AqE1* и *AqE2 T. castaneum* (Coleoptera), тогда как *AqE4* не содержит такой делеции. При этом копии расположены на дереве без корреляции с делециями. Таким образом, варианты *AqE* с делециями не имеют общего корня; делеция, по всей видимости, возникала независимо всякий раз в каждом конкретном случае.

Есть также и фланкирующие делеции, но они встречаются в единичных случаях.

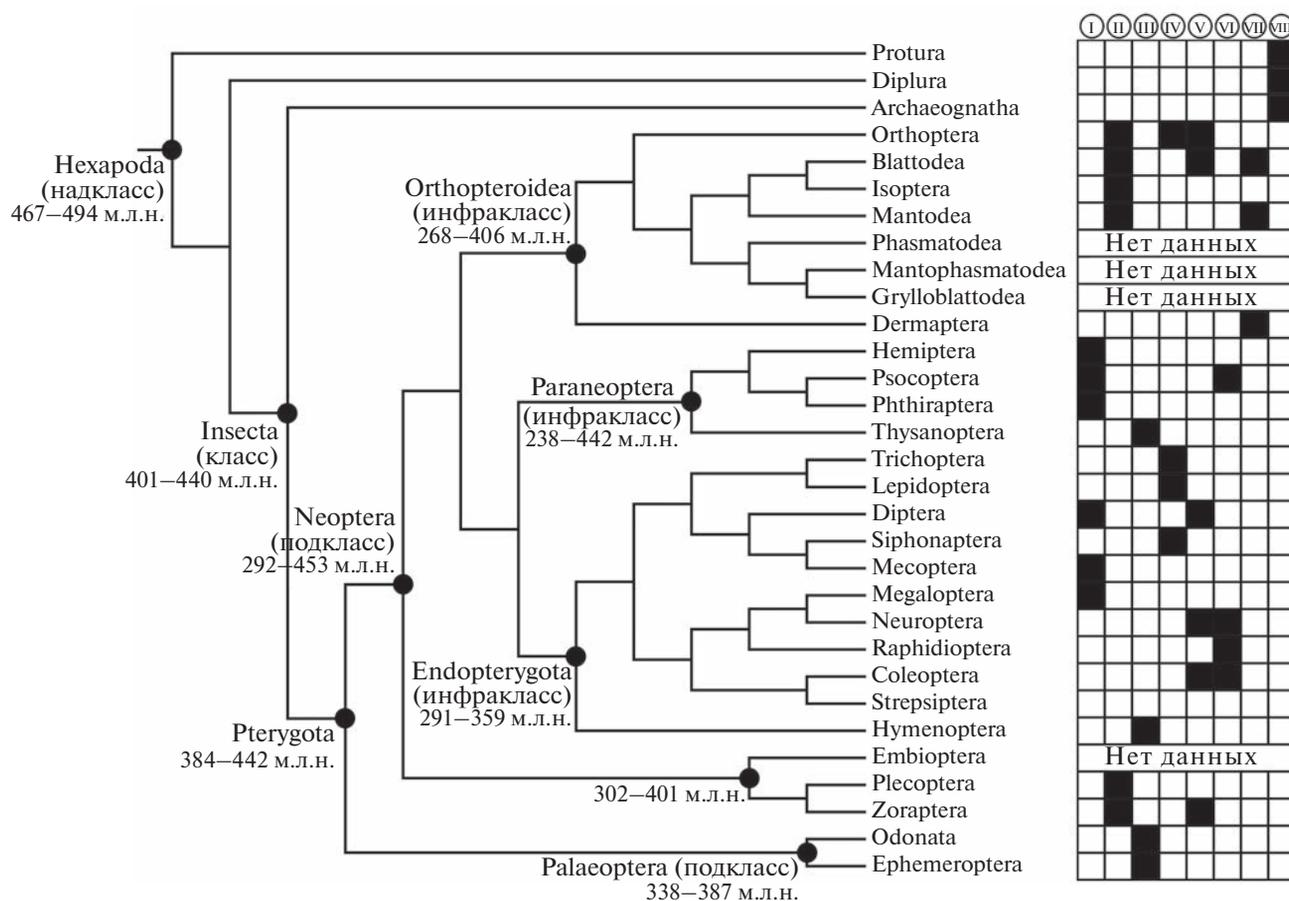
## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Известно, что гены остаются стабильными, если не происходит каких-то существенных изменений во внешних условиях и стабилизирующий отбор “выбрасывает” любые мутации, изменяющие или нарушающие функцию белка. Ген *AqE* присутствует почти во всех таксонах водных хордовых [9]. Его протяженность, интрон-экзонная структура и консервативные домены в целом достаточно постоянны, а присутствие транскриптов свидетельствует о “работоспособности” гена. У представителей класса насекомых ситуация противоположная — наблюдается выраженная неоднородность в наличии/отсутствии гена в геноме, в числе копий гена, его размере и экзон-интронной структуре (табл. 1). Причем эта неоднородность проявляется вплоть до вида, что наталкивает на мысль, что у насекомых, оказавшихся в новых условиях обитания, ген *AqE* попал под действие движущего отбора и активно эволюционировал/эволюционирует вплоть до настоящего момента, так как относительно недавно дивергировавшие группы организмов имеют значительные различия.

Филогенетический анализ показал, что все разнообразие *AqE* у насекомых делится на восемь кластеров (рис. 1). При этом статистическая значимость обособления кластеров ниже 50%. Если рассматривать распределение отрядов по кластерам, то видна выраженная мозаичность, которая не позволяет предположить пути эволюции гена *AqE* у насекомых (рис. 2). Возможно, что активная мультипликация *AqE* началась еще при обособлении эволюционной группы Pterygota около 384–442 млн лет назад. Новые копии эволюционировали независимо, а в дальнейшем возникали как новые дубликации, так и потери уже существующих генов, в связи с чем и сформировалось мозаичное распределение без корреляции филогении *AqE* с филогенией видов. Многочисленные события дубликаций и потери гена, а также быстрой эволюции свидетельствуют о том, что *AqE* у насекомых утратил свою консервативность и значимость в метаболических процессах. Так, например, еще одна подробно изученная группа животных Actinopterygii (костистые рыбы) дивергировала от других позвоночных примерно в тот же период: 364–409 млн лет назад. Однако ген *AqE* у этих животных строго консервативен [9].

Дубликация — один из основных источников изменений, способствующих быстрой эволюции генома [17, 18]. Судьба большинства дублицированных генов — псевдогенизация и/или элиминация из генома [19, 20]. У насекомых встречаются случаи псевдогенизации — у *L. kohalensis* в гене *AqE2* рамка считывания нарушена стоп-кодоном, ген *AqE3* остался в виде фрагмента, кодирующего продукт длиной 48 а.о. У *B. germanica* *AqE8* представляет собой фрагмент, кодирующий пептид длиной 150 а.о. Продукт гена *AqE2 An. glabripennis* состоит из 147 а.о. (при этом выявлены транскрипты). У *Leptinotarsa decemlineata* все три гена *AqE* неполноценны (также выявлены транскрипты). В большинстве же случаев ген остается визуально функциональным, имеет неповрежденную ОРС, а зачастую и транскрипты. Следовательно, в подавляющем числе случаев мы имеем дело с дубликацией, которая успешно закрепилась в геноме насекомых. Ведущий эволюционный биохимик Shelley Copley предполагает, что средний фермент может иметь до 10 разных функций [21]. Соответственно, при дубликации гена любая смежная активность может быть отправной точкой для эволюции нового фермента, функция которого сходна с оригинальной или довольно сильно от нее отличается.

Дублицированные гены могут закрепиться в геноме в результате процессов: а) неофункционализации — приобретения возникшим паралогом новой функции; б) ухода от адаптивного конфликта, когда исходный ген выполняет в организме две функции или более, а после дубликации бифункционального гена происходит разде-



**Рис. 2.** Присутствие разных филогенетических кластеров гена *AqE* среди подтипа Hexapoda. Филогенетические отношения между отрядами и данные о времени дивергенции взяты из базы данных TimeTree (<https://timetree.org/>). м.л.н. – миллионы лет назад.

ление функций между возникшими копиями; в) субфункционализации, при которой из-за накопления мутаций оба паралога становятся необходимыми для выполнения функции, которую ранее обеспечивал предковый ген; г) консервации или мутации типа CNV (вариации числа копий), при которой копии сохраняются в неизменном состоянии, но увеличение их количества дает организму возможность синтеза большего количества специфических РНК или белков [17, 22, 23].

Мы предполагаем, что дубликации закрепились в геноме насекомых как результат неофункционализации или ухода от адаптивного конфликта. Если бы у предка насекомых дубликация привела к субфункционализации, то в настоящий момент среди потомков не должно быть организмов с одной копией гена *AqE*, так как при субфункционализации оба паралога необходимы для выполнения исходной функции гена. Нами же обнаружены организмы, имеющие единственный функциональный ген *AqE* (табл. 1). Вариация числа копий предполагает простое увеличение копияности, мы же имеем дело с очень измененными копиями *AqE*.

На рис. 1 видно, что *AqE* кластера V имеют более длинные эволюционные расстояния, чем другие. Это может быть косвенным свидетельством того, что эта группа *AqE* могла приобрести новую функцию. Известно, что по мере того, как организмы приобретали большее количество специализированных типов клеток, дубликация генов способствовала образованию совершенно разных ферментов, приспособленных к этим клеткам, а также увеличивала регуляторную специфичность в отношении времени их экспрессии в развитии [24–26]. Новые ферменты могут образовываться, поскольку избыточность копий позволяет одной последовательности свободно мутировать и приобретать новые каталитические функции. Таким образом, возможно, произошло формирование большого разнообразия дегидрогеназ [27].

В ходе эволюции конкретных видов насекомых возникало множество дубликаций, которые успешно закреплялись в геноме. Среди них, помимо случаев неофункционализации или ухода от адаптивного конфликта, может быть и субфункционализация, например, у *B. germanica* все 9 дубликаций гена накопили достаточно мутаций

и имеют давнее происхождение, судя по их расхождению на филогенетическом древе. Не исключена и возможность консервации, так у *D. busckii* идентичность копий 1, 2 и 3 (всего их пять) гена составляет 99.77–100%. Однако сказать наверняка вследствие чего закрепились дубликации в каждом конкретном случае мы не можем из-за недостаточности данных.

Десять отрядов насекомых сохранили единственную копию гена, потеряв дублированную. Но есть случаи и полной потери гена — четыре отряда вообще не имеют гена *AqE* (табл. 1). Существуют различные сценарии эволюционной инактивации и/или потери гена: медленное накопление мутационных изменений в гене и превращение его в псевдоген с дальнейшей постепенной деградацией (фрагментация), или же внезапная и полная потеря гена — делеция вследствие неравного кроссинговера при мейозе или перемещения мобильных генетических элементов. Мы не обнаружили никаких остатков гена в исследованных нами геномах в группах, потерявших ген *AqE*, поэтому более достоверным представляется, что делеция, а, возможно, также и псевдогенизация привела к такой сильной деградации, что ген не обнаруживается при поиске.

Потеря генов — широко распространенный в эволюции феномен. Показано, что потеря генов может быть адаптивной или нейтральной. Если ген становится избыточным, то его потеря оказывает нейтральное влияние на жизнедеятельность [28, 29]. С другой стороны, потеря гена может существенно повышать адаптивный потенциал вида. Мутации, приводящие к потере функциональности гена, происходят чаще, чем мутации, которые приводят к приобретению новой функции [30]. Следовательно, потери гена сильнее влияют на адаптивную эволюцию, особенно во время быстрой адаптации к резким изменениям окружающей среды [31].

Мы не можем делать выводы об адаптивном характере потери гена у насекомых из-за отсутствия данных, свидетельствующих в пользу такой возможности. Вероятнее всего, виды, потерявшие ген, приобрели иные пути метаболизма, вследствие чего *AqE* утратил свою значимость и элиминировался из генома, или же аналогичные гены взяли на себя его функцию. Известны примеры негомологичной замены генов. Так, например, ферменты SLDH, которые используют оксалоацетат в качестве субстрата с относительно высокой эффективностью [32], могут действовать в качестве аналога MDH, компенсирующего отсутствие специфической MDH у метаногенной археи [6].

Насекомые — это единственный класс животных, обитающий на суше и сохранивший ген *AqE*, кодирующий SLDH-подобный белок [5]. Не ис-

ключено, что у насекомых *AqE* включен в некие метаболические пути, которые оказались полезными в условиях их обитания. Личинки насекомых, а иногда и имаго, зачастую сталкиваются с недостатком кислорода, избытком сероводорода, метана и иных газов, образующихся при брожении и разложении органических веществ. В таких стрессовых условиях, возможно, необходимы новые метаболические пути, способные обеспечить существование организма в подобных агрессивных средах. Наиболее важная функция оксидоредуктаз связана с их эколого-биохимической ролью в развитии адаптивных реакций, выражающихся, как правило, в регуляции баланса аэробных и анаэробных процессов [33]. Показано, что оксидоредуктазы могут повышать устойчивость растений к холодовому и солевому стрессу [34–36].

Функция гена *AqE* пока не установлена, но, возможно, его продукт является резервным метаболитом, который включается в анаэробный обмен в экстремальных ситуациях.

## ЗАКЛЮЧЕНИЕ

В результате изучения распространенности и разнообразия гена *AqE* установлено, что этот ген представлен не во всех отрядах/подотрядах насекомых, при том что у некоторых видов ген *AqE* имеет два и более гомолога. При этом даже в пределах одного отряда/подотряда наблюдается гетерогенность как в количестве гомологов, так и в потере генов. Предполагается, что мультипликация гена *AqE* насекомых имеет древнюю природу, однако возможны и молодые дубликации. Более того, установлена варибельность как длины гена, так и числа экзонов в нем (от 1 до 10). Не исключено, что в связи с появлением паралогов ген приобрел новую функцию. Такое разнообразие в представленности и в структуре создает большие сложности с идентификацией паралогов и установлением эволюционных взаимосвязей. Возможно, при более детальном (на уровне семейств и родов) исследовании разнообразия *AqE* удастся получить более точные данные. Дополнить картину эволюции позволит также изучение этого гена у всех членистоногих или более древних таксонов.

Предполагается, что необычное поведение гена *AqE*, может быть связано как с потерей его значимости для метаболизма насекомых, так и с приобретением новых функций и/или с необходимостью адаптации к разнообразным условиям, в которых обитают представители этого класса.

Исследование проведено в рамках государственного задания ФГБУН ИМБИ «Функциональные, метаболические и токсикологические аспекты существования гидробионтов и их популя-

ций в биотопах с различным физико-химическим режимом” (номер гос. регистрации 121041400077-1).

Настоящая статья не содержит исследований с использованием животных в качестве объектов.

Авторы заявляют об отсутствии конфликта интересов.

### СПИСОК ЛИТЕРАТУРЫ

- Honka E., Fabry S., Niermann T., Palm P., Hensel R. (1990) Properties and primary structure of the *L*-malate dehydrogenase from the extremely thermophilic archaeobacterium *Methanothermus fervidus*. *Eur. J. Biochem.* **188**, 623–632. <https://doi.org/10.1111/j.1432-1033.1990.tb15443.x>
- Jendrossek D., Kratzin H.D., Steinbuchel A. (1993) The *Alcaligenes eutrophus* *ldh* structural gene encodes a novel type of lactate dehydrogenase. *FEMS Microbiol. Lett.* **112**, 229–235. <https://doi.org/10.1111/j.1574-6968.1993.tb06453.x>
- Muramatsu H., Mihara H., Kakutani R., Yasuda M., Ueda M., Kurihara T., Esaki N. (2005) The putative malate/lactate dehydrogenase from *Pseudomonas putida* is an NADPH-dependent delta 1-piperidine-2-carboxylate/delta 1-pyrroline-2-carboxylate reductase involved in the catabolism of *D*-lysine and *D*-proline. *J. Biol. Chem.* **280**(7), 5329–5335. <https://doi.org/10.1074/jbc.M411918200>
- Muramatsu H., Mihara H., Goto M., Miyahara I., Hirotsu K., Kurihara T., Esaki N. (2005) A new family of NAD(P)H-dependent oxidoreductases distinct from conventional Rossmann-fold proteins. *J. Biosci. Bioeng.* **99**, 541–754. <https://doi.org/10.1263/jbb.99.541>
- Puzakova L.V., Puzakov M.V., Soldatov A.A. (2019) Gene encoding a novel enzyme of LDH2/MDH2 family is lost in plant and animal genomes during transition to land. *J. Mol. Evol.* **87**, 52–59. <https://doi.org/10.1007/s00239-018-9884-2>
- Irimia A., Madern D., Zaccari G., Vellieux F.M. (2004) Methanoarchaeal sulfolactate dehydrogenase: prototype of a new family of NADH-dependent enzymes. *EMBO J.* **23**, 1234–1244. <https://doi.org/10.1038/sj.emboj.7600147>
- Denger K., Cook A.M. (2010) Racemase activity effected by two dehydrogenases in sulfolactate degradation by *Chromohalobacter salexigens*: purification of (S)-sulfolactate dehydrogenase. *Microbiology* (Reading). **156**, 967–974. <https://doi.org/10.1099/mic.0.034736-0>
- Zhang Y., Schofield L.R., Sang C., Dey D., Ronimus R.S. (2017) Expression, purification, and characterization of (R)-sulfolactate dehydrogenase (ComC) from the rumen methanogen *Methanobrevibacter millerae* SM9. *Archaea.* **6**, 5793620. <https://doi.org/10.1155/2017/5793620>
- Puzakova L.V., Puzakov M.V., Gostyukhina O.L. (2021) Newly discovered *AqE* gene is highly conserved in non-tetrapod vertebrates. *J. Mol. Evol.* **89**, 214–224. <https://doi.org/10.1007/s00239-021-09997-x>
- Berthelot C., Brunet F., Chalopin D., Juanchich A., Bernard M., Noël B., Bento P., Da Silva C., Labadie K., Alberti A., Aury J. M., Louis A., Dehais P., Bardou P., Montfort J., Klopp C., Cabau C., Gaspin C., Thorgaard G.H., Boussaha M., Quillet E., Guymard R., Galiana D., Bobe J., Volff J.N., Genêt C., Wincker P., Jaillon O., Roest Crollius H., Guiguen Y. (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* **5**, 3657.
- Petit J., David L., Dirks R., Wiegertjes G.F. (2017) Genomic and transcriptomic approaches to study immunology in cyprinids: what is next? *Devel. Compar. Immunol.* **75**, 48–62.
- Пузакова Л.В., Пузаков М.В. (2022) Тканеспецифичность активности гена *AqE* у желтого горбыля *Larimichthys crocea*. *Генетика.* **58**, 540–549.
- Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Edgar R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**, 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Kumar S., Stecher G., Li M., Knyaz C., Tamura K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Rogozin I.B., Carmel L., Csuros M., Koonin E.V. (2012) Origin and evolution of spliceosomal introns. *Biol. Direct.* **7**, 11. <https://doi.org/10.1186/1745-6150-7-11>
- Cardoso-Moreira M., Long M. (2012) The origin and evolution of new genes. In: *Evolutionary Genomics. Methods Mol. Biol. (Methods and Protocols)*. Ed. Anisimova M. 856. Humana Press, 161–186. [https://doi.org/10.1007/978-1-61779-585-5\\_7](https://doi.org/10.1007/978-1-61779-585-5_7)
- Taylor J.S., Raes J. (2004) Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* **38**, 615–643. <https://doi.org/10.1146/annurev.genet.38.072902.092831>
- Lynch M., Conery J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science.* **290**, 1151–1155. <https://doi.org/10.1126/science.290.5494.1151>
- Журавлева Г.А., Инге-Вечтомов С.Г. (2009) Возникновение новых белков за счет дубликации генов – что общего в эволюции зрительных чувствительных белков и факторов терминации трансляции. *Молекуляр. биология.* **43**, 759–771.
- Copley S.D. (2017) Shining a light on enzyme promiscuity. *Curr. Opin. Struct. Biol.* **47**, 167–175. <https://doi.org/10.1016/j.sbi.2017.11.001>
- Ohno S. (1970) Introduction. In: *Evolution by Gene Duplication*. Berlin, Heidelberg: Springer, [https://doi.org/10.1007/978-3-642-86659-3\\_1](https://doi.org/10.1007/978-3-642-86659-3_1)
- Hahn M.W. (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* **100**, 605–617. <https://doi.org/10.1093/jhered/esp047>

24. Markert C.L. (1971) Developmental Genetics. Heinrich Ursprung. 214 p.
25. Markert C.L., Shaklee J.B., Whitt G.S. (1975) Evolution of a gene. Multiple genes for LDH isozymes provide a model of the evolution of gene structure, function and regulation. *Science*. **189**, 102–114. <https://doi.org/10.1126/science.1138367>
26. Zuckerkandl E. (1978) Multilocus enzymes, gene regulation, and genetic sufficiency. *J. Mol. Evol.* **12**, 57–89. <https://doi.org/10.1007/BF01732545>
27. Eventhoff W., Rossman M. G. (1975) The evolution of the dehydrogenases and kinases. *CRC Crit. Rev. Biochem.* **3**, 111–140.
28. Moreau R., Dabrowski K. (1998) Body pool and synthesis of ascorbic acid in adult sea lamprey (*Petromyzon marinus*): an agnathan fish with gulonolactone oxidase activity. *Proc. Natl. Acad. Sci. USA*. **95**, 10279–10282. <https://doi.org/10.1073/pnas.95.17.10279>
29. Drouin G., Godin J.R., Pagé B. (2011) The genetics of vitamin C loss in vertebrates. *Curr Genomics*. **12**, 371–378. <https://doi.org/10.2174/138920211796429736>
30. Albalat R., Cañestro C. (2016) Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391. <https://doi.org/10.1038/nrg.2016.39>
31. Greenberg A.J., Moran J.R., Coyne J.A., Wu C.I. (2003) Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science*. **302**, 1754–1757. <https://doi.org/10.1126/science.1090432>
32. Graupner M., Xu H., White R.H. (2000) Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing reactions involved in coenzyme biosynthesis in methanarchaea. *J. Bacteriol.* **182**, 3688–3692.
33. Мещерякова О.В. (2004) Динамика активности изоферментов лактатдегидрогеназы, малатдегидрогеназы и  $\alpha$ -глицерофосфатдегидрогеназы в процессе адаптации рыб к различным факторам окружающей среды. Автореферат дисс. канд. биол. наук. Петрозаводск. 24 с.
34. Kandoi D., Mohanty S., Tripathy B.C. (2018) Overexpression of plastidic maize NADP-malate dehydrogenase (ZmNADP-MDH) in *Arabidopsis thaliana* confers tolerance to salt stress. *Protoplasma*. **255**, 547–563. <https://doi.org/10.1007/s00709-017-1168-y>
35. Wang Q.J., Sun H., Dong Q.L., Sun T.Y., Jin Z.X., Hao Y.J., Yao Y.X. (2016) The enhancement of tolerance to salt and cold stresses by modifying the redox state and salicylic acid content via the cytosolic malate dehydrogenase gene in transgenic apple plants. *Plant Biotechnol. J.* **14**, 1986–1997.
36. Yao Y.X., Dong Q.L., Zhai H., You C.X., Hao Y.J. (2011) The functions of an apple cytosolic malate dehydrogenase gene in growth and tolerance to cold and salt stresses. *Plant Physiol. Bioch.* **49**, 257–264.

## Structure and Evolution of the *AqE* Gene in Insects

L. V. Puzakova<sup>1</sup>, \* and M. V. Puzakov<sup>1</sup>

<sup>1</sup>Kovalevsky Institute of Biology of the Southern Seas, Russian Academy of Sciences, Sevastopol, 299011 Russia

\*e-mail: kvluda@yandex.ru

*AqE* gene encodes a sulfolactate dehydrogenase-like enzyme of the LDH2/MDG2 oxidoreductase family. The gene was found in representatives of taxa of bacteria and fungi, as well as animals and plants whose life-style is associated with the aquatic environment. The *AqE* gene is also present in arthropods and, in particular, in the class of insects that are predominantly terrestrial. In our work, we studied the distribution and structure of the *AqE* gene in the class of insects in order to trace its evolutionary fate. We found that the studied gene is not present in all orders/suborders of insects, there is a loss of the gene. In some orders, it is duplicated or multiplied. The variability of the gene both in length and in exon-intron structure was established – from intronless to multi-intron. It was found that the multiplication of the *AqE* gene of insects has an ancient nature, but there are also “young” duplications. It is possible that in connection with the appearance of paralogs, the gene acquired a new function.

**Keywords:** gene evolution, duplication, multiplication, gene loss, insects, sulfolactate dehydrogenase

ГЕНОМИКА.  
ТРАНСКРИПТОМИКА

УДК 616.33-006.6

СТРУКТУРА СОМАТИЧЕСКИХ МУТАЦИЙ В ГЕНАХ *KRAS*, *BRAF*, *PIK3CA*  
И КЛИНИЧЕСКИЕ ХАРАКТЕРИСТИКИ ВЭБ-АССОЦИИРОВАННОГО  
И МИКРОСАТЕЛЛИТНО-НЕСТАБИЛЬНОГО РАКА ЖЕЛУДКА

© 2023 г. А. М. Данишевич<sup>а</sup>, \*, Н. И. Поспехова<sup>б</sup>, А. М. Строганова<sup>б</sup>, Д. А. Головина<sup>б</sup>,  
М. П. Никулин<sup>б</sup>, А. Е. Калинин<sup>б</sup>, С. Э. Николаев<sup>а</sup>, И. С. Стилиди<sup>б</sup>, Л. Н. Любченко<sup>с, д</sup>

<sup>а</sup>Московский клинический научно-практический центр им. А.С. Логинова  
Департамента здравоохранения города Москвы, Москва, 111123 Россия

<sup>б</sup>Национальный медицинский исследовательский центр онкологии им. Н.Н. Блохина  
Министерства здравоохранения России, Москва, 115478 Россия

<sup>с</sup>Национальный медицинский исследовательский центр радиологии Министерства здравоохранения России,  
Москва, 125284 Россия

<sup>д</sup>Научно-исследовательский центр урологии и интервенционной радиологии им. Н.А. Лопаткина  
филиал “Национального медицинского исследовательского центра радиологии”  
Министерства здравоохранения России, Москва, 105425 Россия

\*e-mail: danisham7@gmail.com

Поступила в редакцию 07.06.2022 г.

После доработки 11.07.2022 г.

Принята к публикации 26.07.2022 г.

Персонализация подходов к лечению рака желудка (РЖ) является актуальной проблемой, учитывая клиническую гетерогенность и агрессивное течение заболевания. В 2014 г. на основании молекулярных характеристик исследователи The Cancer Genome Atlas выделили четыре подтипа РЖ: Эпштейн–Барр-позитивный (ВЭБ<sup>+</sup>), микросателлитно-нестабильный (МСН), хромосомно-нестабильный (ХН), геномно-стабильный (ГС). На сегодняшний день не существует унифицированной методики идентификации ХН и ГС подтипов, в то время как определение МСН и ВЭБ применяется в рутинной практике и имеет большое клиническое значение. Мы проанализировали статус МСН, провели качественный анализ наличия ДНК ВЭБ, а также поиск соматических мутаций в кодонах 12–13 (экзон 2), 61 (экзон 3), 146 (экзон 4) гена *KRAS*, кодонах 597–601 (экзон 15) гена *BRAF* и кодонах 542–546 (экзон 9), 1047–1049 (экзон 20) гена *PIK3CA* в 159 образцах РЖ. ВЭБ выявлен в 8.2% образцов, МСН – в 13.2%. ВЭБ<sup>+</sup> и МСН-фенотипы были взаимоисключающими. Средний возраст манифестации ВЭБ<sup>+</sup> и МСН рака составил 54.8 и 62.1 г. соответственно. В 92.3% случаев ВЭБ<sup>+</sup> рак выявлен у мужчин, в 76.2% – в группе старше 50 лет, диффузный рак и аденокарцинома кишечного типа встречались в 6 (46.2%) и 5 (38.5%) случаях соответственно. МСН рак встречался в приблизительно равной доле у мужчин и женщин ( $n = 10$ ; 47.6% /  $n = 11$ ; 52.4%) с превалированием кишечного гистотипа (71.4%) и поражением малой кривизны (28.6%). В одном образце ВЭБ<sup>+</sup> рака обнаружен вариант E545K в гене *PIK3CA*. Во всех образцах МСН опухолей выявлено сочетание клинически значимых вариантов генов *KRAS* и *PIK3CA*. Не обнаружено мутации V600E в гене *BRAF*, специфичной для МСН колоректального рака. Показано, что ВЭБ подтип связан с лучшим прогнозом: общая пятилетняя выживаемость при ВЭБ<sup>+</sup> и МСН раке составила 100.0 и 54.7% соответственно.

**Ключевые слова:** рак желудка, молекулярная классификация, вирус Эпштейна–Барр, микросателлитная нестабильность, *KRAS*, *BRAF*, *PIK3CA*

**DOI:** 10.31857/S0026898423010044, **EDN:** AXIQNP

## ВВЕДЕНИЕ

Рак желудка (РЖ) занимает лидирующие позиции в структуре онкологической заболеваемости

Сокращения: РЖ – рак желудка; ЗНО – злокачественное новообразование; ВЭБ – вирус Эпштейна–Барр; ВЭБ<sup>+</sup> – подтип РЖ, позитивный по ВЭБ; МСН – подтип РЖ с микросателлитной нестабильностью/микросателлитная нестабильность; TCGA – атлас опухолевого генома (The Cancer Genome Atlas); ХН – хромосомно-нестабильный подтип; ГС – геномно-стабильный подтип; MMR – система репарации неспаренных оснований ДНК; dMMR – дефицит MMR; RTK – рецепторные тирозинкиназы.

сти. По данным Globocan IARC в 2020 г. в мире зарегистрировано 1089103 новых случаев РЖ, и это пятое место (6%) среди всех впервые выявленных злокачественных новообразований (ЗНО) [1]. Абсолютное число летальных исходов составило 768793 (7.8%), что также соответствует пятому месту в структуре смертности от ЗНО. Общая пятилетняя выживаемость пациентов с I–IV стадией заболевания колеблется в пределах 10–30%, медиана общей выживаемости – менее 1 г. [2, 3].

В Российской Федерации в 2019 г. на долю опухолей желудка пришлось 5.7% (6 место) от всех впервые зарегистрированных ЗНО, абсолютное число заболевших составило 36171. В большинстве случаев (38.9%) РЖ выявляют на III–IV (60.9%) стадиях заболевания, что объясняется значительной гетерогенностью клинических проявлений РЖ и отсутствием скрининговых программ [4].

Согласно классификации Laugen, выделяют два подтипа аденокарциномы желудка: кишечный (50%) и диффузный (33%) [5]. Остальные 17% относятся к смешанным и неклассифицируемым опухолям. Описанные подтипы имеют биологические и клинические различия, а также широко используются для прогноза заболевания [6].

**Факторы риска РЖ.** Известно, что до 50% случаев РЖ обусловлены социальным поведением и пищевыми привычками с преобладанием в рационе красного мяса, копченых продуктов, обработанного и соленого мяса [7–9]. Значительный вклад в развитие РЖ (до 6.2% случаев) вносит бактерия *Helicobacter pylori*, признанная канцерогеном I класса. [10]. Инфицирование вирусом Эпштейна–Барр (ВЭБ) приводит к 10-кратному увеличению риска РЖ и обуславливает до 10% неоплазм [11, 12].

В качестве немодифицируемого фактора риска РЖ рассматривается наследственная предрасположенность к раку в составе наследственных онкологических синдромов (НОС). Герминальные мутации удается выявить в 1–3% случаев семейных форм РЖ. На сегодняшний день выделяют три основные формы наследственного РЖ: диффузный РЖ (ген *CDH1*), аденокарцинома желудка с проксимальным полипозом (ген *APC*) и семейный РЖ кишечного типа (гены-кандидаты *SMAD4*, *PMS1*, *PRSS1* и *TP53*) [13]. Описаны и другие НОС, ассоциированные с высоким риском РЖ: синдром Линча, синдром наследственного рака молочной железы и яичников, семейный аденоматозный полипоз, МУН-ассоциированный полипоз, ювенильный аденоматозный полипоз, синдром Пейтца–Егерса, синдром Ли-Фраумени и др.

**Молекулярная классификация РЖ.** В результате полногеномного исследования опухолей желудка, оценки экспрессии, а также белковых паттернов в 2014 г. группой исследователей The Cancer Genome Atlas (TCGA) была разработана и опубликована молекулярная классификация РЖ, которая включает четыре подтипа [14]:

- Эпштейн–Барр-позитивный (ВЭБ<sup>+</sup>),
- Микросателлитно-нестабильный (МСН),
- Хромосомно-нестабильный (ХН),
- Геномно-стабильный (ГС).

Каждый из этих подтипов имеет свои молекулярные и клинические характеристики, пред-

ставляющие интерес для прогноза заболевания и перспективы индивидуализации лечения РЖ.

**ВЭБ<sup>+</sup> подтип РЖ.** ВЭБ<sup>+</sup> РЖ отличается значительным разнообразием генетических и эпигенетических нарушений, ответственных за регуляцию энергетического обмена и метаболизма [14]. Инфицирование ВЭБ индуцирует aberrантное гиперметилирование CpG-островков в промоторах 886 генов, поэтому ВЭБ<sup>+</sup> РЖ отнесен к эпигенотипу с чрезвычайно высоким уровнем метилирования [15, 16]. Гиперметилирование промотора гена *CDKN2A* (p16INK4A) обнаружено во всех образцах ВЭБ<sup>+</sup> [17]. Кроме того, выявлена высокая частота соматических мутаций в генах *ARID1A* (55%), *BCOR* (23%) [18] и *PIK3CA* (40–80%) [19]. Как показали Bögerlo и соавт., до 14% первичных ВЭБ<sup>+</sup> опухолей желудка характеризуются значительной внутриопухолевой гетерогенностью и могут содержать от трех до пяти вариантов онкогена *PIK3CA*, включая дикий тип [20]. Этот факт предполагает эффективность лечения ВЭБ<sup>+</sup> РЖ ингибиторами PI3K. Dong и соавт. отметили ассоциацию сверхэкспрессии *PIK3CA* при ВЭБ<sup>+</sup> РЖ с лучшей пятилетней выживаемостью в отличие от ВЭБ-негативных опухолей (57.8 против 33.4%,  $p < 0.001$ ) [21].

Обнаружены также протяженные делеции генов *PTEN*, *SMAD4*, *CDKN2A*, *ARID1A* и вариация числа копий генов *JAK2*, *CD274* и *PDCD1LG2* и др. [8, 22].

Клинически большинство ВЭБ<sup>+</sup> опухолей относятся к диффузному гистологическому типу по Laugen, они в 2–4 раза чаще встречаются у мужчин (79–81%) [23] и в 2 раза чаще поражают кардиальный отдел и тело желудка (62%) [24]. Средний возраст манифестации ВЭБ<sup>+</sup> рака – 53 года, что значительно ниже, чем при других подтипах РЖ [25, 26]. Кроме того, ВЭБ<sup>+</sup> РЖ отличается наиболее благоприятным прогнозом, меньшим количеством пораженных лимфоузлов и лучшими показателями пятилетней выживаемости [26, 27].

**МСН подтип РЖ.** МСН диагностируется в 10–37% опухолей желудка. Опухоли этого подтипа характеризуются многочисленными геномными нарушениями, которые возникают из-за дефекта системы репарации неспаренных оснований ДНК (Mismatch Repair system, MMR) [28–30]. Большинство МСН опухолей желудка имеют спорадический характер [31, 32], до 15% случаев развиваются в результате герминальных мутаций в генах MMR (чаще в *MLH1* и *MSH2*) при синдроме Линча [33, 34]. В случае спорадического рака МСН обусловлена гиперметилированием промотора и сайленсинга гена *MLH1* [35]. МСН опухоли имеют гипермутированный фенотип с предполагаемой частотой 31.6 мутаций/млн.п.н. (в микросателлитно-стабильных опухолях – 3.3 мутации/млн.п.н.) [36]. В ходе исследования

TCGA выявлено 37 наиболее часто мутирующих генов, участвующих в регуляции клеточного цикла (*TP53*, *IGF1R*, *TCF4*), в процессах репарации ДНК (*MSH6*, *MSH3*, *MED1*, *RAD50*, *BLM*, *ATR* и *MRE11*), ремоделировании хроматина, апоптозе (*RIZ*, *BAX*, *CASPASE5*, *FAS*, *BCL10* и *APAF1*), регуляции транскрипции и сигнальных систем клетки [14]. Corso и соавт. описали мутации в генах *EGFR* в 47.6% случаев РЖ, *KRAS* – в 14.9%, *PIK3CA* – в 14.3 и *MLK3* – в 3.2% [37]. При МСН раке, как и при ВЭБ<sup>+</sup> РЖ, отмечена высокая частота нарушений в генах *ARID1A* (83%) и *RNF43* (55%) [38]. Описаны соматические мутации (22%) или потеря экспрессии (35–54%) генов, ответственных за процессинг микроРНК [38, 39], и повышенная экспрессия генов митотической сети *AURKA A/B*, *E2F*, *FOXM1*, *PLK1* и *MYC* [14]. Кроме того, выявлены частые изменения генов главного комплекса гистосовместимости, включая *B2M* и *HLA-B* [40].

Накопление соматических мутаций способствует образованию неоантигенов, способных вызывать иммунный ответ, что делает МСН опухоли чувствительными к терапии ингибиторами иммунных контрольных точек. Результаты рандомизированных исследований KEYNOTE-059, KEYNOTE-061, KEYNOTE-062 показали улучшение показателей выживаемости при применении пембролизумаба в монорежиме и в комбинации с химиотерапией. Таким образом, у пациентов с распространенным РЖ МСН может быть биомаркером эффективности пембролизумаба независимо от линии терапии [41].

Клинически МСН РЖ связан с женским полом, пожилым возрастом (средний возраст манифестации 60 лет) и дистальным расположением ЗНО. Кроме того, отмечено преобладание кишечного гистотипа (кишечный – 10.7%, диффузный – 2.9%, смешанный – 0.9% из всех случаев РЖ [42]) и более ранняя стадия заболевания вследствие низкой локальной инвазии и меньшего поражения лимфатических узлов [43]. В большинстве случаев МСН РЖ ассоциирован с лучшим прогнозом в отличие от ХН и ГС подтипов и наиболее низкой частотой рецидивирования (22%) [14, 44].

**ХН подтип РЖ.** ХН подтип составляет примерно 50% всех случаев РЖ и характеризуется выраженной анеуплоидией, потерей гетерозиготности, делецией и амплификацией ряда генов [45, 46]. Мутации генов рецепторных тирозинкиназ (РТК) *VEGFA*, *EGFR*, *ERBB2*, *ERBB3*, *c-Met*, *FGFR2* найдены в 37% опухолей [46, 47]. Амплификация гена *ERBB2* и иммуногистохимическая сверхэкспрессия белка HER2neu выявлена в 4–24% опухолей желудка [48]. Амплификация гена *KRAS* обнаружена примерно в 10% ХН опухолей желудка и, по-видимому, ее присутствие исключает амплификацию других РТК [49]. В целом, ампли-

фикация РТК показала себя как предиктор плохого прогноза [50].

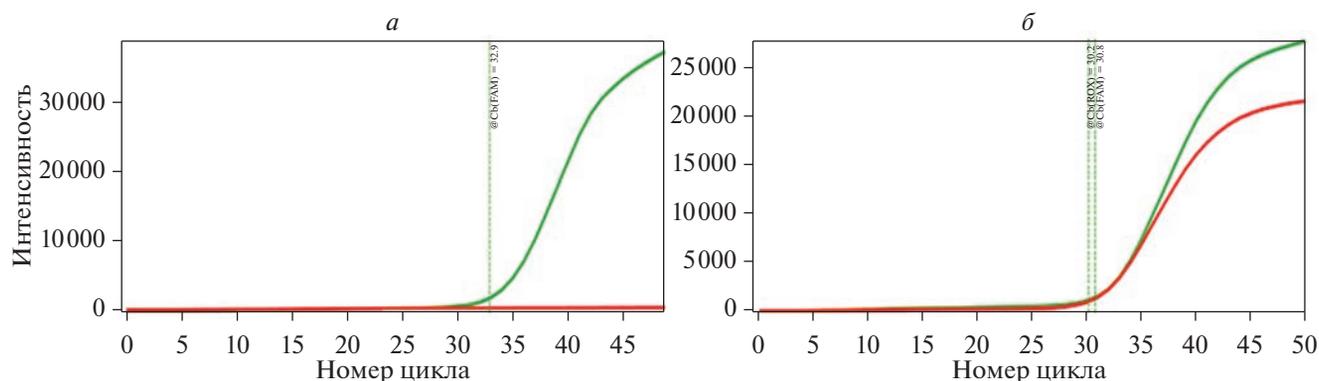
Также ХН РЖ отличается высокой частотой соматических мутаций антионкогена *TP53*; отмечено увеличение числа копий генов-медиаторов клеточного цикла *CCNE1*, *CCND1* и *CDK6* [14, 46, 47]. Кроме того, в ХН опухолях диагностированы редкие слияния генов *AGTRAP-BRAF*, *ERBB2-CDK12*, *ERBB2-NEUROD2*, *SLC34A2-ROS1*, *SNX2-PRDM6* и *MLL3-PRKAG2* [26]. Клиническое значение этих нарушений в контексте таргетной терапии еще предстоит выяснить.

В 50% случаев ХН подтип представлен интестинальным гистотипом с кардиальной/кардиоэзофагеальной локализацией опухоли и характеризуется относительно благоприятным прогнозом [14]. В отличие от других подтипов РЖ применение адъювантной химиотерапии при ХН РЖ показало увеличение трехлетней безрецидивной выживаемости пациентов после лекарственного лечения в сравнении с его отсутствием (58.7 и 33.5%) [26].

**ГС подтип РЖ.** ГС подтип отличается от ХН РЖ низкой степенью анеуплоидии и низкой частотой мутаций в гене *TP53*, к этому подтипу относятся до 15–20% аденокарцином желудка. Гены *CDH1* и *RHOA*, ассоциированные с разрозненной моделью роста клеток и отсутствием межклеточной адгезии, несли патогенные варианты в 37 и 15% опухолей соответственно [14, 52]. Таким образом, до 73% диффузного РЖ можно отнести к ГС подтипу [53]. Кроме точечных мутаций, в ГС РЖ выявлено 170 различных структурных перестроек. В 62% ГС опухолей обнаружены химерные гены *CLDN18-ARHGAP*. Также в образцах ГС РЖ отмечена активация микроРНК miR-21, приводящая к быстрой пролиферации, инвазии и метастазированию [54, 55].

Средний возраст манифестации ГС РЖ – 52 г., этот подтип ассоциирован с поражением всех отделов желудка в равном соотношении и наихудшим прогнозом [14]. Применение адъювантной химиотерапии при ГС РЖ не влияло на прогноз [26].

На основе молекулярных характеристик РЖ разрабатываются различные прогностические модели течения РЖ и риска возникновения рецидива. Кроме того, эта классификация может служить основой для стратификации пациентов с целью индивидуализации лечения в будущем [26]. На данный момент не существует унифицированной методики идентификации ХН и ГС подтипов РЖ, в то время как определение МСН и виртуальной нагрузки ВЭБ широко используется в рутинной практике и имеет большое клиническое значение. Соматические мутации в генах *KRAS*, *BRAF*, *PIK3CA* относятся к наиболее значимым маркерам, которые определяют при различных ЗНО и имеют прогностическое и предиктивное



**Рис. 1.** Качественная оценка наличия ДНК ВЭБ. *а* – Отсутствие ДНК ВЭБ. *б* – Наличие ДНК ВЭБ (зеленая кривая – график амплификации внутреннего контрольного образца, красная кривая – график амплификации ДНК ВЭБ).

значение. На сегодняшний день проходят клинические испытания нескольких ингибиторов компонентов сигнального пути RTK/RAS у больных РЖ [84]. Таким образом, актуальным представляется анализ соматических нарушений этих генов.

В нашей работе оценена частота соматических мутаций в генах *KRAS*, *BRAF*, *PIK3CA* и определены клинико-молекулярные характеристики ВЭБ<sup>+</sup> и МСН опухолей желудка у пациентов российской популяции.

### ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

В исследование включены образцы опухолевой ткани и периферической крови 159 пациентов, проходивших лечение в НМИЦ онкологии им. Н.Н. Блохина Минздрава России по поводу РЖ I–IV стадии в период с 2010 по 2020 гг. включительно. ДНК-диагностика 54 образцов свежей послеоперационной опухолевой ткани и 105 срезов FFPE выполнена в молекулярно-биологической лаборатории отдела морфологической и молекулярно-генетической диагностики опухолей НИИ клинической онкологии им. Н.Н. Трапезникова НМИЦ онкологии им. Н.Н. Блохина. Образцы опухолевой ткани отбирали во время оперативного вмешательства, быстро замораживали и хранили в морозильной камере при температуре  $-80^{\circ}\text{C}$ . Отобранные препараты FFPE содержали не менее 20% опухолевых клеток по площади среза, периферическую кровь отбирали в пробирки с EDTA.

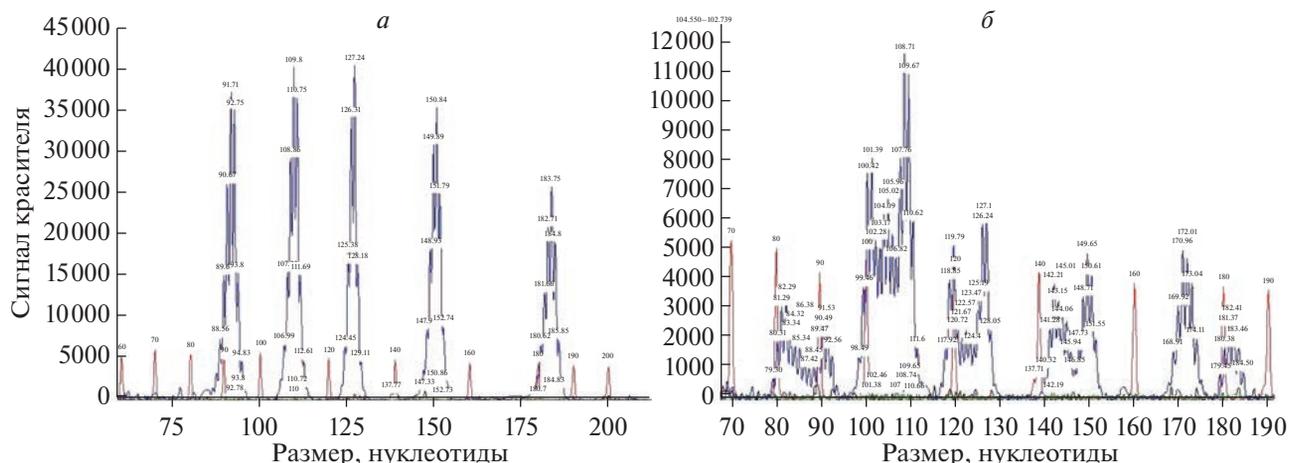
Выделение ДНК из послеоперационных образцов опухолевой ткани и срезов опухоли, заключенной в парафин, проводили с использованием коммерческого набора Cobas® DNA Sample Preparation Kit (“Roche”, Германия) согласно инструкции производителя. Геномную ДНК выделяли из лимфоцитов периферической крови с использованием набора Проба-ГС-Генетика (“ДНК-технология”, Россия) в соответствии с протоко-

лом производителя. Качество ДНК оценивали, измеряя концентрацию и соотношение  $A_{260}/A_{280}$  на спектрофотометре NanoDrop Lite (“Thermo Fisher Scientific”, США). Во всех 159 образцах РЖ определяли ДНК ВЭБ, статус МСН и кодирующую последовательность генов *KRAS*, *BRAF* и *PIK3CA*.

ДНК ВЭБ в образцах выявляли с использованием коммерческого набора реагентов для определения ДНК ВЭБ методом ПЦР в режиме реального времени “РеалБест” ДНК ВЭБ (комплект 1) (“Вектор Бест”, Россия) на приборе ДТ-96 (“НПФ ДНК-Технология”, Россия). Ложноположительные/ложноотрицательные результаты исключали, анализируя все образцы в двух повторностях. Результаты оценивали в соответствии с инструкцией производителя (рис. 1).

Статус МСН определяли методом фрагментного анализа с использованием пяти квазиномономорфных мононуклеотидных маркеров NR21, NR24, NR27, BAT25, BAT26 на платформе Beckman Coulter Genome Lab GeXP (“Beckman Coulter”, США). При отсутствии нарушений или нестабильности одного маркера статус опухоли определяли как микросателлитно-стабильный (МСС), при нестабильности двух или более маркеров – как опухоль с МСН (рис. 2).

Поиск соматических мутаций в кодонах 12–13 (экзон 2), 61 (экзон 3) и 146 (экзон 4) гена *KRAS*, в кодонах 597–601 (экзон 15) гена *BRAF* и кодонах 542–546 (экзон 9), 1047–1049 (экзон 20) гена *PIK3CA* выполнен путем анализа кривых плавления с высоким разрешением (High Resolution Melting, HRM) после проведения ПЦР в режиме реального времени с применением праймеров и зондов на амплификаторе Rotor-Gene (“QIAGEN”, Германия). Последовательности праймеров, зондов и условия ПЦР описаны ранее [65]. Присутствие выявленного варианта подтверждали путем очистки полученного ПЦР-продукта и прямого секвенирования по Сенгеру на платформе Beckman



**Рис. 2.** Оценка статуса микросателлитной нестабильности (МСН). *а* – Микросателлитно-стабильный статус опухоли. *б* – Микросателлитно-нестабильный статус опухоли.

Coulter Genome Lab GeXP (“Beckman Coulter”) согласно инструкции производителя.

Статистический анализ проводили с помощью программного обеспечения GraphPad Prism 8.0.0 (GraphPad Software, США). Общую выживаемость определяли как время от момента постановки диагноза до смерти от любой причины или даты последнего наблюдения пациента и рассчитывали по методу Каплана–Мейера. Различия между общей выживаемостью пациентов трех групп оценивали при помощи теста лог-ранга критерия Мантеля–Кокса. Для попарных сравнений выживаемости группы пациентов использовали тест лог-ранга Мантеля–Кокса и критерий Бреслоу–Гехана–Вилкоксона. Выживаемость на определенных этапах исследования сравнивали с использованием точного теста Фишера.

Обследуемая группа состояла из 159 человек. Клинические характеристики пациентов представлены в табл. 1.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

На первом этапе исследования была проспективно собрана коллекция образцов послеоперационной опухолевой ткани 54 пациентов. В результате ПЦР-диагностики ДНК ВЭБ выявлена в трех образцах опухоли. На втором этапе с целью валидации результатов, полученных для трех ВЭБ<sup>+</sup> и семи случайно отобранных ВЭБ-негативных (ВЭБ<sup>-</sup>) образцов, определяли вирусную нагрузку на срезах FFPE, в результате чего во всех случаях получено совпадение показателей. В дальнейшем присутствие ВЭБ оценивали ретроспективно на 107 срезах FFPE.

В результате проведенного исследования ДНК ВЭБ выявлена в 13 (8.2%) образцах. В большинстве случаев ВЭБ<sup>+</sup> РЖ диагностирован у мужчин

( $n = 12$ ; 92.3%). У 10 (76.9%) пациентов диагноз установлен после 50 лет, средний возраст манифестации РЖ 54.8 лет. Чаще ВЭБ<sup>+</sup> опухоли располагались в теле желудка ( $n = 4$ ; 30.7%), кардиально ( $n = 3$ ; 23.0%) или поражали дно ( $n = 3$ ; 23.0%). Диффузный рак и аденокарцинома кишечного типа встречались в практически равном соотношении – 6 (46.2%) и 5 (38.5%) случаев, в то время как смешанный тип наблюдался только у 2 (15.4%) обследованных. Распространенный рак (III–IV стадии) выявлен у 53.9% ( $n = 7$ ), в остальных случаях ( $n = 6$ ; 46.2%) диагноз установлен на I–II стадии. При III стадии отмечено поражение 3–6 регионарных лимфатических узлов ( $n = 1$ ), 7 и более ( $n = 1$ ), вовлечение висцеральной брюшины ( $n = 1$ ), в двух случаях лимфатические узлы не оценены. Терминальная стадия ( $n = 2$ ) характеризовалась метастатическим поражением в кости.

МСН статус опухоли обнаружен в 13.2% ( $n = 21$ ) образцов РЖ. У мужчин и женщин МСН РЖ диагностирован в 10 (47.6%) и 11 (52.4%) образцах соответственно. В 76.2% ( $n = 16$ ) заболевание манифестировало в возрасте старше 50 лет, средний возраст диагностики  $62.1 \pm 14.5$  лет. У 6 (28.6%) обследованных установлено поражение малой кривизны, в 5 (23.8%) случаях – больше одной топографической области, реже тела желудка ( $n = 4$ ; 19.0%) и преддверия привратника ( $n = 4$ ; 19.0%). Кишечный и диффузный гистотипы были представлены в 71.4% ( $n = 15$ ) и 19.0% ( $n = 4$ ) образцов соответственно. В большинстве случаев диагноз установлен на прогрессирующих стадиях ( $n = 14$ ; 66.7%). На III стадии МСН РЖ ( $n = 8$ ; 38.1%) чаще наблюдалось поражение 1–2 региональных лимфатических узлов ( $n = 3$ ; 37.5%), реже 3–6 ( $n = 2$ ; 25.0%), 7 и более лимфоузлов ( $n = 1$ , 12.5%), в двух (25.0%) случаях III стадия установлена вследствие вовлечения брюшины. Для IV стадии ( $n = 6$ ,

**Таблица 1.** Клинические характеристики обследованных групп с различным статусом ВЭБ и МСН

Клиническая характеристика		ВЭБ <sup>+</sup> -группа	МСН-группа	ВЭБ <sup>-</sup> + МСН-группа
Общее количество случаев		13 (8.2)	21 (13.2)	125 (78.6)
Пол	Мужчины	12 (92.3)	10 (47.6)	54 (43.2)
	Женщины	1 (7.7)	11 (52.4)	71 (56.8)
Возраст манифестации, лет	16–49	10 (76.9)	5 (23.8)	86 (71.2)
	50–83	3 (23.1)	16 (76.2)	39 (28.8)
	Среднее ± SD	54.8 ± 11.9	62.1 ± 14.5	55.0 ± 14.7
Стадия	I	2 (15.4)	1 (4.8)	17 (13.6)
	II	4 (30.8)	6 (28.6)	22 (17.6)
	III	5 (38.5)	8 (38.1)	43 (34.4)
	IV	2 (15.4)	6 (28.6)	43 (34.4)
Гистологический тип по Lauren	Кишечный	5 (38.5)	15 (71.4)	73 (58.4)
	Диффузный	6 (46.2)	4 (19.0)	60 (48.0)
	Смешанный	2 (15.4)	2 (9.5)	12 (9.6)
Локализация опухоли	Кардия	3 (23.1)	2 (9.5)	30 (24.0)
	Дно	3 (23.1)	1 (4.8)	7 (5.6)
	Тело	4 (30.7)	5 (23.8)	38 (30.4)
	Преддверие привратника	1 (7.7)	4 (19.0)	13 (10.4)
	Привратник	0 (0)	1 (4.8)	5 (4.0)
	Малая кривизна	1 (7.7)	3 (14.3)	3 (2.4)
	Большая кривизна	1 (7.7)	0 (0)	5 (4.0)
	Выходит за пределы указанных областей	0 (0)	5 (23.8)	24 (19.2)

Примечание. указано число пациентов и их доля (%) от общего числа пациентов.

28.6%) было характерно вторичное поражение легких ( $n = 1$ ), костей ( $n = 2$ ), печени ( $n = 3$ ), яичников ( $n = 1$ ) и лимфоузлов средостения ( $n = 1$ ). В нашем исследовании ВЭБ<sup>+</sup> и МСН РЖ были взаимоисключающими.

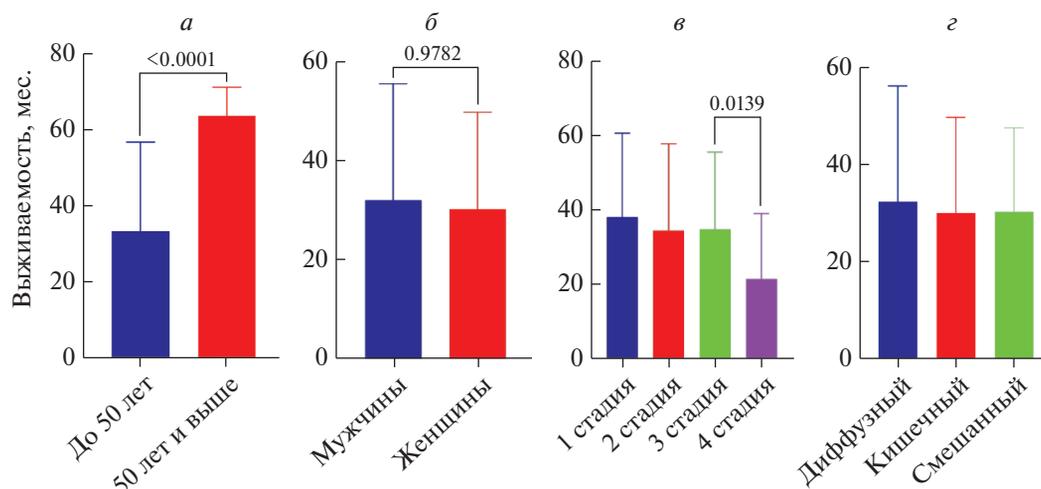
**Оценка клиничко-анамнестических данных.** Анализ исследованных подгрупп по возрасту манифестации рака не выявил значимых различий ( $p = 0.0691$ , критерий Краскела–Уоллиса). Средний возраст манифестации РЖ в группе МСН составил 62.1 лет. У пациентов с ВЭБ<sup>+</sup> и ВЭБ<sup>-</sup> + МСН рак диагностировали в более молодом возрасте – 54.8 и 55.0 лет соответственно. ВЭБ<sup>+</sup> РЖ в 92.3% случаев выявлен у мужчин, в МСН-группе соотношение мужчин и женщин составило 47.6 и 52.4% соответственно. Опухоли ВЭБ<sup>+</sup> в 76.2% выявлены у пациентов 50 лет и старше (критерий Фишера,  $p = 0.7560$ ). Напротив, в случае МСН 76.9% обследованных были моложе 50 лет (критерий Фишера,  $p = 0.4351$ ).

В отличие от I, II и IV стадий в группах ВЭБ<sup>+</sup> и МСН чаще диагностировали III стадию заболевания – 38.5 и 38.1% соответственно. Диффузный

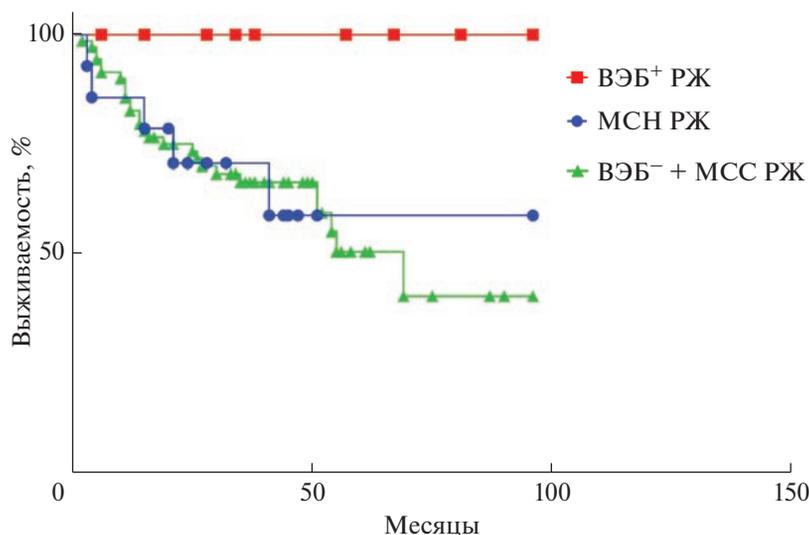
рак при ВЭБ<sup>+</sup> РЖ выявлен в 46.2% случаев, кишечный гистотип – в 38.5%. МСН РЖ чаще ассоциирован с кишечным гистотипом (71.4%). Подавляющее большинство ВЭБ<sup>+</sup> опухолей располагалось в верхних отделах желудка (76.9%) и теле желудка (30.7%). МСН рак наблюдался во всех отделах и, как и ВЭБ<sup>+</sup>, чаще поражал тело желудка (23.8%), но не развивался на большой кривизне. Клинические характеристики обследованных групп в зависимости от статуса ВЭБ и МСН представлены в табл. 1.

**Оценка общей выживаемости.** Среднее время наблюдения за больными в общей группе от момента постановки диагноза составило 31.3 мес. (95%; ДИ: 27.6–35.1 мес.), медиана общей выживаемости (ОВ) – 55.0 мес. ОВ (1-, 3- и 5-летняя) составила 84.6, 69.6 и 54.3 мес. соответственно.

Статистически значимые различия в ОВ выявлены при сравнении групп с учетом возраста манифестации: до и после 50 лет – 33.39 vs. 63.81 мес. ( $p < 0.0001$ , U-критерий Манна–Уитни). Не выявлено значимых различий в ОВ в зависимости от пола пациентов ( $p = 0.9782$ , U-критерий Манна–



**Рис. 3.** Анализ общей выживаемости (ОВ) в обследованной группе. *a* – ОВ в зависимости от возраста манифестации РЖ. *б* – ОВ в зависимости от пола пациента. *в* – ОВ в зависимости от стадии РЖ. *г* – ОВ в зависимости от гистологического подтипа РЖ.



**Рис. 4.** Показатели общей выживаемости пациентов с инфекцией ВЭБ, микросателлитно-нестабильным раком желудка и без них (из анализа исключены пациенты IV стадии заболевания). МСН РЖ – рак желудка с микросателлитной нестабильностью. ВЭБ<sup>+</sup> – ВЭБ-ассоциированный рак желудка. ВЭБ<sup>-</sup> + МСС РЖ – ВЭБ-негативный микросателлитно-стабильный рак желудка.

Уитни). Значимые различия в ОВ выявлены на III и IV стадиях РЖ ( $p = 0.0139$ , критерий Данна). Других статистических отличий между стадиями заболевания не прослежено в связи с малочисленностью обследуемых групп. Также не найдено значимых различий в ОВ в зависимости от гистологического типа опухоли ( $p = 0.9641$ , критерий Краскела–Уоллиса) (рис. 3).

Как показано на графиках (рис. 4), у пациентов с ВЭБ<sup>+</sup> подтипом прогноз существенно лучше, чем в группах МСН ( $p = 0.0598$ , лог-ранг тест) и ВЭБ<sup>-</sup> + МСС ( $p = 0.0446$ , лог-ранг тест).

Статистически значимой разницы в показателях ОВ пациентов с МСН и ВЭБ<sup>-</sup> + МСС не наблюдалось ( $p = 0.6572$ , лог-ранг-тест). 50%-Выживаемость достигнута в группе больных ВЭБ<sup>-</sup> + МСС. Результаты анализа ОВ представлены в табл. 2.

**Исследование соматических мутаций в генах *KRAS*, *BRAF* и *PIK3CA*.** В ходе изучения статуса генов *KRAS* и *PIK3CA* выявлено 12 (7.5%) и 5 (2.5%) клинически значимых вариантов соответственно (рис. 5).

В кодонах 597–601 (экзон 15) гена *BRAF* мутаций не выявлено.

**Таблица 2.** 12-, 36-, 60-месячная выживаемость больных I–III стадиями рака желудка (%) и медиана выживаемости

N = 92	12 мес.	36 мес.	60 мес.	96 мес.	Медиана, мес.
ВЭБ <sup>+</sup>	100	100	100	100	Не достигнута
МСН	84.6	68.4	54.7	54.7	Не достигнута
ВЭБ <sup>-</sup> + МСС	82.6	66.5	50.6	40.5	69
Парные сравнения, 96-месячная выживаемость					
	ВЭБ <sup>+</sup>	МСН	ВЭБ <sup>-</sup> + МСС		
ВЭБ <sup>+</sup>		0.0598	0.0446		
МСН	0.0598		0.6572		
ВЭБ <sup>-</sup> + МСС	0.0446	0.6572			
Парные сравнения, 12-месячная выживаемость					
ВЭБ <sup>+</sup>		0.4935	0.3408		
МСН	0.4935		>0.9999		
ВЭБ <sup>-</sup> + МСС	0.3408	>0.9999			
Парные сравнения, 36-месячная выживаемость					
ВЭБ <sup>+</sup>		0.1150	0.0557		
МСН	0.1150		>0.9999		
ВЭБ <sup>-</sup> + МСС	0.0557	>0.9999			
Парные сравнения, 60-месячная выживаемость					
ВЭБ <sup>+</sup>		0.0537	0.0264		
МСН	0.0537		>0.9999		
ВЭБ <sup>-</sup> + МСС	0.0264	>0.9999			

Примечание. Приведены значения *p*-критерия для парных сравнений.

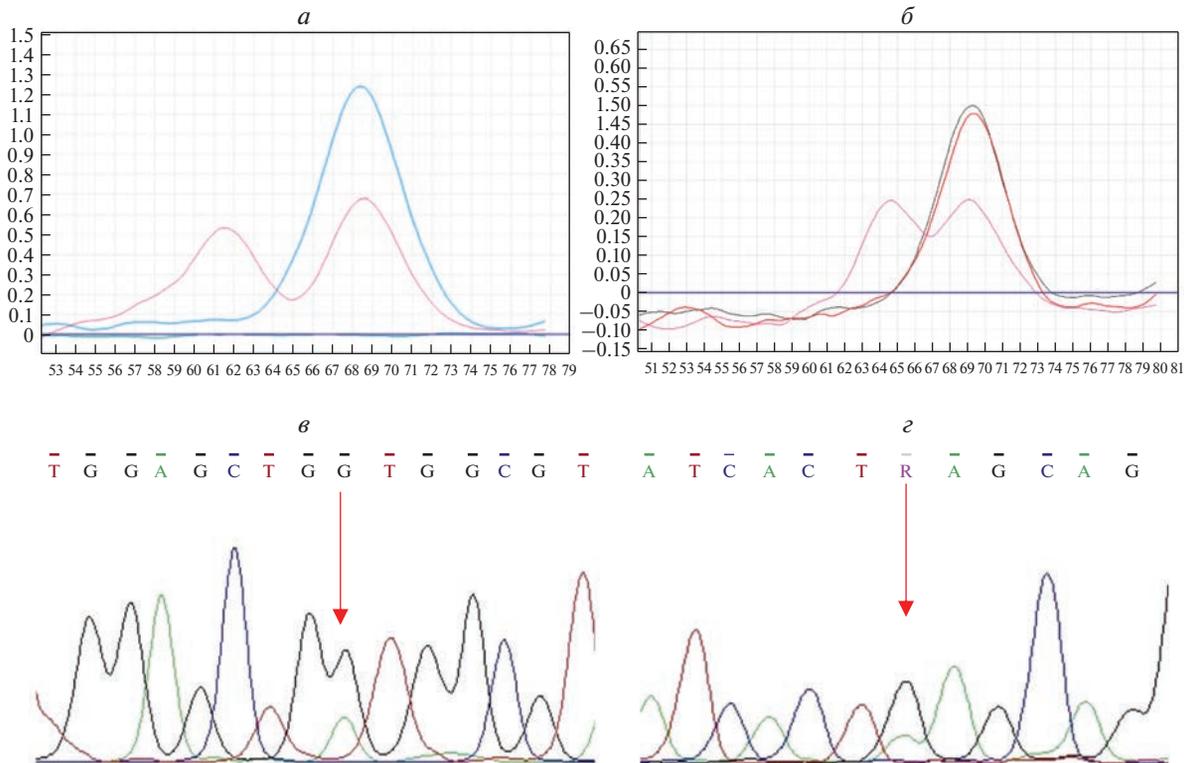
Все 13 (100%) образцов ВЭБ<sup>+</sup> содержали ген *KRAS* дикого типа, лишь в одном случае (7.7%) в экзоне 20 гена *PIK3CA* выявлен вариант E545K. В отличие от ВЭБ<sup>+</sup> опухолей, МСН-фенотип характеризовался большим количеством соматических мутаций: 4 (19.0%) варианта в кодонах 12–13 гена *KRAS* и 7 (33.3%) – в экзонах 9 и 20 гена *PIK3CA* (рис. 6).

Кроме того, во всех 4 (100%) образцах МСН опухолей с мутированным *KRAS* обнаружены варианты гена *PIK3CA* (табл. 3).

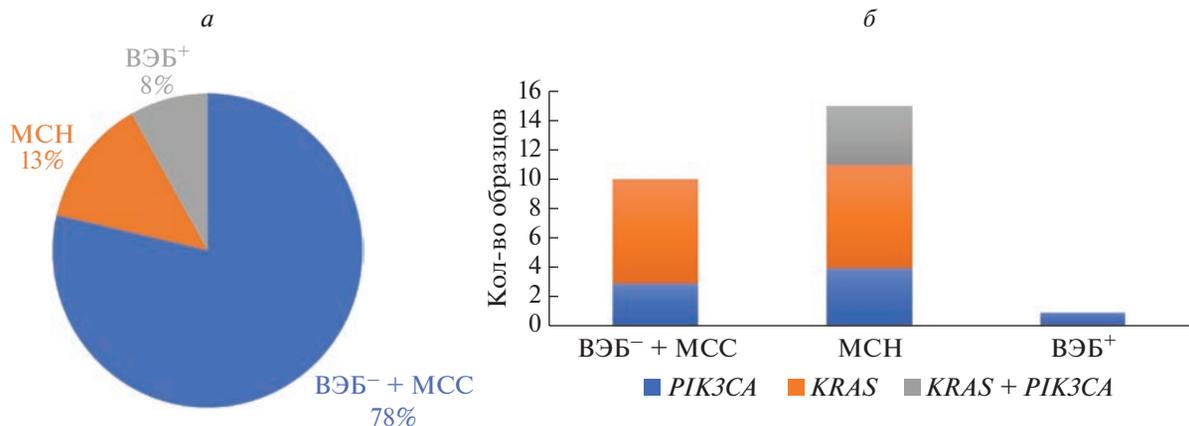
## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Экономичный и доступный метод ПЦР широко используется для выявления ВЭБ-инфекции в сыворотке, плазме и цельной крови. Однако, как сообщалось Deyhimi и соавт., при ПЦР-диагностике ВЭБ в солидных опухолях существует возможность получения ложноположительных результатов. Низкую специфичность ПЦР можно объяснить тем, что клетки памяти и/или неопухолевые лимфоциты, инфицированные ВЭБ, также могут попасть в образец для анализа. Таким образом, методы на основе ПЦР более чувствительны, но менее специфичны, чем метод гибридизации

*in situ* (ISH) – золотой стандарт диагностики ВЭБ в гистопатологических образцах [56]. С целью исключения из анализа нормальной ткани желудка мы проводили забор опухолевого материала с использованием метода макродиссекции под контролем патоморфолога. С той же целью ПЦР проведена в двух повторностях, в результате чего частота ВЭБ<sup>+</sup> РЖ в нашей выборке составила 8.2%. Tavakoli и соавт. провели метаанализ данных 71 статьи и 2226 пациентов с РЖ из 26 стран и показали, что совокупная распространенность ВЭБ среди больных РЖ составила 8.77% [57]. По данным метаанализа, проведенного Samargo и соавт. (13 исследований с участием 4599 пациентов), средний возраст постановки диагноза ВЭБ<sup>+</sup> РЖ составлял 58 лет с преобладанием мужчин (71%) [23]. Tavakoli и соавт. также отметили высокую частоту возникновения ВЭБ<sup>+</sup> РЖ у лиц мужского пола. В нашем исследовании 92.3% ВЭБ<sup>+</sup> РЖ также диагностирован у мужчин. Однако в нашей выборке отмечен более ранний возраст манифестации ВЭБ<sup>+</sup> РЖ – 54.8 лет. Кроме того, Tavakoli и соавт. в 57% случаев отнесли ВЭБ<sup>+</sup> РЖ к кишечному типу, тогда как в нашей группе наблюдалась тенденция к преобладанию диффузного типа над кишечным (46.2 vs. 38.5%).



**Рис. 5.** HRM-анализ соматических мутаций в генах *KRAS*, *PIK3CA*, и результаты прямого секвенирования по Сенгеру. *a* – Кривая плавления при анализе кодонов 12–13 гена *KRAS* (синий график – дикий тип, розовый – мутантный тип). *б* – Кривая плавления при анализе кодонов 1047–1049 гена *PIK3CA* (красный и зеленый графики – дикий тип, розовый – мутантный тип). *в* – Гетерозиготный вариант G12D в гене *KRAS*. *з* – Гетерозиготный вариант E545K в гене *PIK3CA*.



**Рис. 6.** Генотипические характеристики образцов. *a* – Распределение больных раком желудка с учетом инфицирования ВЭБ и статуса микросателлитной нестабильности. *б* – Количество соматических мутаций в генах *KRAS* и *PIK3CA* в зависимости от статуса ВЭБ и микросателлитной нестабильности. МСН – рак желудка с микросателлитной нестабильностью. ВЭБ<sup>+</sup> – ВЭБ-ассоциированный рак желудка. ВЭБ<sup>-</sup> + МСС – ВЭБ-негативный микросателлитно-стабильный рак желудка.

По данным различных источников МСН РЖ диагностируется в 10–37% опухолей желудка [28–30], что согласуется с частотой МСН опухолей в нашем исследовании (13.2%). Поскольку для оценки статуса МСН мы использовали эта-

лонную панель Bethesda [58], разработанную преимущественно для диагностики статуса МСН при колоректальном раке, можно предположить, что частота МСН РЖ может быть несколько выше. Так, Ronald и соавт. оценили МСН аденокарци-

**Таблица 3.** Спектр выявленных клинически значимых вариантов в генах *KRAS* и *PIK3CA*

Ген	ВЭБ <sup>+</sup>	МСН	ВЭБ <sup>-</sup> + МСС
<i>KRAS</i>	—	—	G12A, 1* G12V, 1* G12D, 3* Q61E, 1*
<i>PIK3CA</i>	Q546K, 1*	Q546K, 1* H1047R, 2*	E545K, 1* H1047R, 2*
<i>KRAS</i> + <i>PIK3CA</i>	—	<i>KRAS</i> G13D + <i>PIK3CA</i> E542K, 2* <i>KRAS</i> G13D + <i>PIK3CA</i> H1047R, 1* <i>KRAS</i> G12V + <i>PIK3CA</i> H1047R, 1*	—

Примечание. При каждом варианте приведено число случаев, в которых он выявлен.

номы желудка методом NGS. Частота МСН РЖ составила 19.0%, уступая раку эндометрия (30.0%) и колоректальному раку (19.0%). Отмечена высокая частота МСН в широком спектре локусов, часто совпадающих с МСН локусами при раке толстой кишки [44]. Несмотря на то, что Su-Jin Shin и соавт. сообщили о большей частоте МСН-опухолей у женщин, в нашей группе 92.3% случаев диагностировано у пациентов мужского пола. Как и в других исследованиях, МСН фенотип был представлен кишечным гистотипом [31, 32].

Согласно TCGA, мутации в гене *PIK3CA* могут содержать до 80% ВЭБ<sup>+</sup> опухолей, тогда как в исследованной нами выборке лишь в одном случае выявлен вариант E545K в экзоне 20. В исследовании TCGA отмечено, что в отличие от других подтипов РЖ, где наибольшая частота aberrаций зарегистрирована в киназном домене гена *PIK3CA* (экзон 20), в ВЭБ<sup>+</sup> опухолях мутации распределены по всему гену [14]. Таким образом, низкая частота соматических мутаций в гене *PIK3CA* при ВЭБ<sup>+</sup> РЖ может быть связана с недостаточным количеством тестируемых экзонов и/или недостаточной чувствительностью HRM. Напротив, МСН РЖ отличается высокой частотой соматических мутаций (33.3%) в экзонах 9 и 20 гена *PIK3CA*, что значительно превышает показатели, полученные Polom и соавт. (8.1%) [59] и Corso и соавт. (14.3%) [37]. Расхождение показателей может быть связано с недостаточным объемом выборки, использованной нами.

Кроме того, в 100% наблюдений в нашей выборке с МСН отмечена высокая частота соматических мутаций в гене *KRAS* (19.0%) и их сочетание с мутациями в гене *PIK3CA*, что объясняется гипермутированным фенотипом МСН опухолей. Крупное международное многоцентровое исследование, изучающее статус генов *KRAS* и MMR у пациентов с местно-распространенным операбельным РЖ, подтвердило корреляцию между мутациями *KRAS* и механизмом dMMR [60]. По нашим результатам, а также по данным Varbi и

соавт., мутация H1047R в гене *PIK3CA* обнаружена только в образцах МСН РЖ и связана со статусом МСН [61].

Отличительной особенностью нашего исследования стало отсутствие мутаций в кодонах 597–601 гена *BRAF*, патогномоничных для МСН колоректального рака, которые служат маркером исключения синдрома Линча [53]. Согласно проведенным ранее исследованиям, частота мутаций в гене *BRAF* колеблется в диапазоне 0–11% [62].

Как значения ОВ, так и результаты опубликованных ранее исследований подтверждают связь ВЭБ<sup>+</sup> подтипа с наилучшим прогнозом — 100% ОВ в течение 96 мес. наблюдения. Различия в ОВ при МСН и ВЭБ<sup>-</sup> + МСС раке не были статистически значимыми, что, вероятно, связано с малым количеством пациентов в группе МСН ( $n = 8$ , 5.3%). Polom и соавт. отметили более низкую пятилетнюю выживаемость пациентов с МСН и мутациями, чем у пациентов с геном *PIK3CA* дикого типа — 40.0 vs. 70.4% соответственно. В том же исследовании оценили различия в выживаемости пациентов с МСН и различными мутациями в *PIK3CA*. Оказалось, что пятилетняя выживаемость при мутации в экзоне 9 составила 0 и 80% — при мутации в экзоне 20 [42].

На сегодняшний день в России стандарт молекулярной диагностики при РЖ включает исследование статуса HER2/neu, МСН и экспрессии PD-L1 в случае местнораспространенной неоперабельной или диссеминированной аденокарциномы. Оценка статуса МСН возможна также при операбельных формах РЖ для определения показаний к адьювантной химиотерапии [63]. Таким образом, индивидуализированное лечение доступно преимущественно пациентам с прогрессирующими стадиями РЖ. Ограниченный спектр рекомендованных ДНК-тестов определяет актуальность поиска и внедрения в клиническую практику новых молекулярных маркеров РЖ. Это позволит прогнозировать течение заболевания при первично операбельном РЖ, что в конечном

счете сможет благоприятно отразиться на показателях выживаемости. Результаты многочисленных исследований показали, что ВЭБ<sup>+</sup> и МСН являются прогностически благоприятными маркерами при РЖ. Вероятно, системы анализа статуса ВЭБ и МСН в гистопатологических образцах с помощью искусственного интеллекта, разрабатываемые в настоящий момент, сделают оценку данных нарушений менее затратной и более доступной [64]. Таким образом, дальнейшее изучение влияния статуса МСН и ВЭБ-инфицирования на возникновение и течение РЖ является актуальной задачей. Расхождение в данных о роли соматических мутаций в генах *KRAS*, *PIK3CA* и *BRAF* делают необходимым дальнейшее изучение прогностического и предиктивного значения этих маркеров.

Проведение исследования одобрено комитетом по биомедицинской этике НМИЦ онкологии им. Н.Н. Блохина Минздрава России, все пациенты подписали информированное согласие

Авторы сообщают об отсутствии конфликта интересов.

#### СПИСОК ЛИТЕРАТУРЫ

- Sung H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A., Bray F. (2021) Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J. Clinicians*. **71**, 209–249.
- Oba K., Paoletti X., Alberts S., Bang Y.J., Benedetti J., Bleiberg H., Catalano P., Lordick F., Michiels S., Morita S., Ohashi Y., Pignon J.P., Rougier P., Sasako M., Sakamoto J., Sargent D., Shitara K., Van Cutsem E., Buysse M., Byrzykowski T., GASTRIC group (2013) Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. *J. Natl. Cancer Inst.* **105**, 1600–1607.
- Dicken B.J., Bigam D.L., Cass C., Mackey J.R., Joy A.A., Hamilton S.M. (2005) Gastric adenocarcinoma: review and considerations for future directions. *Ann. Surgery*. **241**, 27–39.
- Каприн А.Д., Старинский В.В., Шахзадова А.О. (2020) *Злокачественные новообразования в России в 2019 году*. Москва: МНИОИ им. П.А. Герцена – филиал ФГБУ “НМИЦ радиологии” Минздрава России.
- Lauren P. (1965) The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol. Microbiol. Scandinavica*. **64**, 31–49.
- Tan P., Yeoh K.G. (2015) Genetics and molecular pathogenesis of gastric adenocarcinoma. *Gastroenterology*. **149**, 1153–1162.
- Bonequi P., Meneses-González F., Correa P., Rabkin C.S., Camargo M.C. (2013) Risk factors for gastric cancer in Latin-America: a meta-analysis. *Cancer Causes Control*. **24**, 217–231.
- Strumylaite L., Zickute J., Dudzevicius J., Dregval L. (2006) Salt-preserved foods and risk of gastric cancer. *Medicina* (Kaunas, Lithuania), **42**, 164–170.
- González C.A., Pera G., Agudo A., Palli D., Krogh V., Vineis P., Tumino R., Panico S., Berglund G., Simán H., Nyrén O., Agren A., Martinez C., Dorronsoro M., Barriarte A., Tormo M.J., Quiros J.R., Allen N., Bingham S., Day N., Miller A., Nagel G., Boeing H., Overland K., Tjonneland A., Bueno-De-Mesquita H.B., Boshuizen H.C., Peeters P., Numans M., Clavel-Chapelon F., Helen I., Agapitos E., Lund E., Fahey M., Saracci R., Kaaks R., Riboli E. (2003) Smoking and the risk of gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Internat. J. Cancer*. **107**, 629–634.
- Plummer M., Franceschi S., Vignat J., Forman D., De Martel C. (2015) Global burden of gastric cancer attributable to *Helicobacter pylori*. *Internat. J. Cancer*. **136**, 487–490.
- Uozaki H., Fukayama M. (2008) Epstein–Barr virus and gastric carcinoma – viral carcinogenesis through epigenetic mechanisms. *Int. J. Clin. Exp. Pathol.* **1**, 198–216.
- Bae J.M., Kim E.H. (2016) Epstein–Barr Virus and gastric cancer risk: a meta-analysis with meta-regression of case-control studies. *J. Prev. Med. Public Health*. **49**, 97–107.
- Assumpção P., Araújo T., Khayat A., Santos S., Barra W., Ishak G., Santos S., Barra W., Felipe Acioli J., Rossi B., Assumpção P. (2020) Hereditary gastric cancer: three rules to reduce missed diagnoses. *World J. Gastroenterol.* **26**, 1382–1393.
- The Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. **513**, 202–209.
- Matsusaka K., Kaneda A., Nagae G., Ushiku T., Kikuchi Y., Hino R., Uozaki H., Seto Y., Takada K., Aburatani H., Fukayama M. (2011) Classification of Epstein–Barr virus-positive gastric cancers by definition of DNA methylation epigenotypes. *Cancer Res*. **71**, 7187–7197.
- Zhao J., Liang Q., Cheung K.F., Kang W., Lung R.W.M., Tong J.H.M., To K.F., Sung J.J.Y., Yu J. (2013) Genome-wide identification of Epstein–Barr virus-driven promoter methylation profiles of human genes in gastric cancer cells. *Cancer*. **119**, 304–312.
- Kandath C., McLellan M.D., Vandin F., Ye K., Niu B., Lu C., Xie M., Zhang Q., McMichael J.F., Wyczalkowski M.A., Leiserson M.D.M., Miller C.A., Welch J.S., Walter M.J., Wendl M.C., Ley T.J., Wilson R.K., Raphael B.J., Ding L. (2013) Mutational landscape and significance across 12 major cancer types. *Nature*. **502**, 333–339.
- Astolfi A., Fiore M., Melchionda F., Indio V., Bertuccio S.N., Pession A. (2019) BCOR involvement in cancer. *Epigenomics*. **11**, 835–855.
- Hoadley K.A., Yau C., Wolf D.M., Cherniack A.D., Tamborero D., Zou L., Ng S., Leiserson M. D. M., Niu B., McLellan M.D., Uzunangelov V., Zhang J., Kandath C., Akbani R., Shen H., Omberg L., Chu A., Margolin A.A., Veer L.A.J., Lopez-Bigas N., Laird P.W., Raphael B.J., Ding L., Robertson A. G., Lauren A.B., Mills G.B.,

- Weinstein J.N., Waes C.V., Chen Z., Collisson E.A., Cancer Genome Atlas Research Network, Benz C.C., Perou C.M., Stuart J.M. (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. **158**, 929–944.
20. Böger C., Krüger S., Behrens H.M., Bock S., Haag J., Kalthoff H., Röcken C. (2017) Epstein–Barr virus-associated gastric cancer reveals intratumoral heterogeneity of PIK3CA mutations. *Ann. Oncol.* **28**, 1005–1014.
  21. Dong M., Wang H.Y., Zhao X.X., Chen J.N., Zhang Y.W., Huang Y., Xue L., Li H.G., Du H., Wu X.Y., Shao C.K. (2016) Expression and prognostic roles of PIK3CA, JAK2, PD-L1, and PD-L2 in Epstein–Barr virus-associated gastric carcinoma. *Hum. Pathol.* **53**, 25–34.
  22. Derks S., Liao X., Chiaravalli A.M., Xu X., Camargo M.C., Solcia E., Sessa F., Fleitas T., Freeman G.J., Rodig S.J., Rabkin C.S., Bass A.J. (2016) Abundant PD-L1 expression in Epstein–Barr virus-infected gastric cancers. *Oncotarget*. **7**, 32925–32932.
  23. Camargo M.C., Kim K.M., Matsuo K., Torres J., Liao L.M., Morgan D.R., Michel A., Waterboer T., Zabaleta J., Dominguez R.L., Yatabe Y., Kim S., Rocha-Guevara E.R., Lissowska J., Pawlita M., Rabkin C.S. (2016) Anti-*Helicobacter pylori* antibody profiles in Epstein–Barr virus (EBV)-positive and EBV-negative gastric cancer. *Helicobacter*. **21**, 153–157.
  24. Yanagi A., Nishikawa J., Shimokuri K., Shuto T., Takagi T., Takagi F., Kobayashi Y., Yamamoto M., Miura O., Yanai H., Suehiro Y., Yamasaki T., Yoshiyama H., Sakaida I. (2019) Clinicopathologic characteristics of Epstein–Barr virus-associated gastric cancer over the past decade in Japan. *Microorganisms*. **7**, 305.
  25. van Beek J., zur Hausen A., Kranenbarg E.K., van de Velde C.J.H., Middeldorp J.M., van den Brule A.J.C., Meijer C.J.L.M., Bloemena E. (2004) EBV-positive gastric adenocarcinomas: a distinct clinicopathologic entity with a low frequency of lymph node involvement. *J. Clin. Oncol.: Official J. Am. Soc. Clin. Oncol.* **22**, 664–670.
  26. Sohn B.H., Hwang J.E., Jang H.J., Lee H.S., Oh S.C., Shim J.J., Lee K.W., Kim E.H., Yim S.Y., Lee S.H., Cheong J.H., Jeong W., Cho J.Y., Kim J., Chae J., Lee J., Kang W.K., Kim S., Noh S.H., Ajani J.A., Lee J.S. (2017) Clinical significance of four molecular subtypes of gastric cancer identified by The Cancer Genome Atlas Project. *Clin. Cancer Res.* **23**, 4441–4449.
  27. Lee H.S., Chang M.S., Yang H.K., Lee B.L., Kim W.H. (2004) Epstein–Barr virus-positive gastric carcinoma has a distinct protein expression profile in comparison with Epstein–Barr virus-negative carcinoma. *Clin. Cancer Res.* **10**, 1698–1705.
  28. Choi Y.Y., Bae J.M., An J.Y., Kwon I.G., Cho I., Shin H.B., Eiji T., Aburahmah M., Kim H. I., Cheong J.H., Hyung W.J., Noh S.H. (2014) Is microsatellite instability a prognostic marker in gastric cancer? A systematic review with meta-analysis. *J. Surg. Oncol.* **110**, 129–135.
  29. Smyth E.C., Wotherspoon A., Peckitt C., Gonzalez D., Hulkki-Wilson S., Eltahir Z., Fassan M., Rugge M., Valeri N., Okines A., Hewish M., Allum W., Stenning S., Nankivell M., Langley R., Cunningham D. (2017) Mismatch repair deficiency, microsatellite instability, and survival: an exploratory analysis of the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) trial. *JAMA Oncol.* **3**, 1197–1203.
  30. Choi Y.Y., Cheong J.H. (2019) Comment on ‘To Treat, or Not to Treat, That is the Question: biomarker-guided adjuvant chemotherapy for stage ii and iii gastric cancer’. *Ann. Surgery*. **270**, e40–e41.
  31. Shin S.J., Kim S.Y., Choi Y.Y., Son T., Cheong J.-H., Hyung W.J., Noh S.H., Park C.-G., Kim H.-I. (2019) Mismatch repair status of gastric cancer and its association with the local and systemic immune response. *Oncologist*. **24**, 835–844.
  32. Zhu L., Li Z., Wang Y., Zhang C., Liu Y., Qu X. (2015) Microsatellite instability and survival in gastric cancer: a systematic review and meta-analysis. *Mol. Clin. Oncol.* **3**, 699–705.
  33. Latham A., Srinivasan P., Kemel Y., Shia J., Bandlamudi C., Mandelker D., Middha S., Hechtman J., Zehir A., Dubard-Gault M., Tran C., Stewart C., Sheehan M., Penson A., DeLair D., Yaeger R., Vijai J., Mukherjee S., Galle J., Dickson M.A., Janjigian Y., O’Reilly E.M., Segal N., Saltz L.B., Reidy-Lagunes D., Varghese A.M., Bajorin D., Carlo M.I., Cadoo K., Walsh M.F., Weiser M., Aguilar J.G., Klimstra D.S., Diaz L.A., Baselga J., Zhang L., Ladanyi M., Hyman D.M., Solit D.B., Robson M.E., Taylor B.S., Offit K., Berger M.F., Stadler Z.K. (2019) Microsatellite instability is associated with the presence of Lynch syndrome pan-cancer. *J. Clin. Oncol.* **37**, 286–295.
  34. Wu M.S., Sheu J.C., Shun C.T., Lee W.J., Wang J.T., Wang T.H., Cheng A.L., Lin J.T. (1997) Infrequent hMSH2 mutations in sporadic gastric adenocarcinoma with microsatellite instability. *Cancer Lett.* **112**, 161–166.
  35. Leite M., Corso G., Sousa S., Milanezi F., Afonso L.P., Henrique R., Soares J.M., Castedo S., Carneiro F., Roviello F., Oliveira C., Seruca R. (2011) MSI phenotype and MMR alterations in familial and sporadic gastric cancer. *Internat. J. Cancer.* **128**, 1606–1613.
  36. Wang K., Kan J., Yuen S.T., Shi S.T., Chu K.M., Law S., Chan T.L., Kan Z., Chan A.S.Y., Tsui W.Y., Lee S.P., Ho S.L., Chan A.K.W., Cheng G.H.W., Roberts P.C., Rejto P.A., Gibson N.W., Pocalyko D.J., Mao M., Xu J., Leung S.Y. (2011) Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat. Genet.* **43**, 1219–1223.
  37. Corso G., Velho S., Paredes J., Pedrazzani C., Martins D., Milanezi F., Pascale V., Vindigni C., Pinheiro H., Leite M., Marrelli D., Sousa S., Carneiro F., Oliveira C., Roviello F., Seruca R. (2011) Oncogenic mutations in gastric cancer with microsatellite instability. *Eur. J. Cancer.* **47**, 443–451.
  38. Wang K., Yuen S.T., Xu J., Lee S.P., Yan H.H.N., Shi S.T., Siu H.C., Deng S., Chu K.M., Law S., Chan K.H., Chan A.S.Y., Tsui W.Y., Ho S.L., Chan A.K.W., Man J.L.K., Foglizzo V., Ng M.K., Chan A.S., Ching Y.P., Leung S.Y. (2014) Whole-genome sequencing and comprehensive molecular profiling iden-

- tify new driver mutations in gastric cancer. *Nat. Genet.* **46**, 573–582.
39. Kim M.S., Oh J.E., Kim Y.R., Park S.W., Kang M.R., Kim S.S., Ahn C.H., Yoo N.J., Lee S.H. (2010) Somatic mutations and losses of expression of microRNA regulation-related genes *AGO2* and *TNRC6A* in gastric and colorectal cancers. *J. Pathol.* **221**, 139–146.
  40. Bernal M., Ruiz-Cabello F., Concha A., Paschen A., Garrido F. (2012) Implication of the  $\beta$ 2-microglobulin gene in the generation of tumor escape phenotypes. *Cancer Immunol., Immunother.* **61**, 1359–1371.
  41. Chao J., Fuchs C.S., Shitara K., Tabernero J., Muro K., Van Cutsem E., Bang Y.J., De Vita F., Landers G., Yen C.J., Chau I., Elme A., Lee J., Özgüroğlu M., Catenacci D., Yoon H.H., Chen E., Adelberg D., Shih C.S., Shah S., Bhagia P., Waingberg Z.A. (2021) Assessment of pembrolizumab therapy for the treatment of microsatellite instability-high gastric or gastroesophageal junction cancer among patients in the KEYNOTE-059, KEYNOTE-061, and KEYNOTE-062 clinical trials. *JAMA Oncol.* **7**, 895–902.
  42. Polom K., Marano L., Marrelli D., De Luca R., Roviello G., Savelli V., Tan P., Roviello F. (2018) Meta-analysis of microsatellite instability in relation to clinicopathological characteristics and overall survival in gastric cancer. *Br. J. Surgery.* **105**, 159–167.
  43. Mathiak M., Warneke V.S., Behrens H.M., Haag J., Böger C., Krüger S., Röcken C. (2017) Clinicopathologic characteristics of microsatellite instable gastric carcinomas revisited: urgent need for standardization. *Appl. Immunohistochem. Mol. Morphol.* **25**, 12–24.
  44. Hause R.J., Pritchard C.C., Shendure J., Salipante S.J. (2016) Classification and characterization of microsatellite instability across 18 cancer types. *Nat. Med.* **22**, 1342–1350.
  45. Lengauer C., Kinzler K.W., Vogelstein B. (1998) Genetic instabilities in human cancers. *Nature.* **396**, 643–649.
  46. Ottini L., Falchetti M., Lupi R., Rizzolo P., Agnese V., Colucci G., Bazan V., Russo A. (2006) Patterns of genomic instability in gastric cancer: clinical implications and perspectives. *Ann. Oncol.* **17**(Suppl. 7), vii97–vii102.
  47. Jung E.J., Jung E.J., Min S.Y., Kim M.A., Kim W.H. (2012) Fibroblast growth factor receptor 2 gene amplification status and its clinicopathologic significance in gastric carcinoma. *Hum. Pathol.* **43**, 1559–1566.
  48. Xie L., Su X., Zhang Lin, Yin X., Tang L., Zhang Xiuhua, Xu Y., Gao Z., Liu K., Zhou M., Gao B., Shen D., Zhang L., Ji A., Gavine P.R., Zhang J., Kilgour E., Zhang X., Ji Q. (2013) *FGFR2* gene amplification in gastric cancer predicts sensitivity to the selective *FGFR* inhibitor AZD4547. *Clin. Cancer Res.* **19**, 2572–2583.
  49. Dulak A.M., Schumacher S.E., Van Lieshout J., Imamura Y., Fox C., Shim B., Ramos A.H., Saksena G., Baca S.C., Baselga J., Tabernero J., Barretina J., Enzinger P.C., Corso G., Roviello F., Lin L., Bandla S., Luketich J.D., Pennathur A., Meyerson M., Ogino S., Shivasani R.A., Beer D.G., Godfrey T.E., Beroukhim R., Bass A.J. (2012) Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res.* **72**, 4383–4393.
  50. Deng N., Goh L.K., Wang H., Das K., Tao J., Tan I.B., Zhang S., Lee M., Wu J., Lim K.H., Lei Z., Goh G., Lim Q.Y., Tan A.L.K., Poh D.Y.S., Riahi S., Bell S., Shi M.M., Linnartz R., Zhu F., Yeoh K.G., Toh H.C., Yong W.P., Cheong H.C., Rha S.Y., Boussioutas A., Grabsch H., Rozen S., Tan P. (2012) A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets. *Gut.* **61**, 673–684.
  51. Palanisamy N., Ateeq B., Kalyana-Sundaram S., Pflueger D., Ramnarayanan K., Shankar S., Han B., Cao X., Cao X., Suleman K., Kumar-Sinha C., Dhanasekaran S.M., Chen Y.B., Esgueva R., Banerjee S., Lafargue C.J., Siddiqui J., Demichelis F., Moeller P., Bismar T.A., Kuefer R., Fullen D.R., Johnson T.M., Greenon J.K., Giordano T.J., Tan P., Tomlins S.A., Varambally S., Rubin M.A., Maher C.A., Chinnaiyan A.M. (2010) Rearrangements of the *RAF* kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat. Med.* **16**, 793–798.
  52. Yu H., Jove R. (2004) The STATs of cancer – new molecular targets come of age. *Nat. Rev. Cancer.* **4**, 97–105.
  53. The Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* **487**, 330–337.
  54. Medina P.P., Nolde M., Slack F.J. (2010) OncomiR addiction in an *in vivo* model of microRNA-21-induced pre-B-cell lymphoma. *Nature.* **467**, 86–90.
  55. Pfeffer S.R., Yang C.H., Pfeffer L.M. (2015) The role of miR-21 in cancer. *Drug Dev. Res.* **76**, 270–277.
  56. Deyhimi P., Kalantari M. (2014) Study of Epstein-Barr virus expression in Burkitt's lymphoma by polymerase chain reaction and *in situ* hybridization: a study in Iran. *Dental Res. J.* **11**, 380–385.
  57. Tavakoli A., Monavari S.H., Solaymani Mohammadi F., Kiani S.J., Armat S., Farahmand M. (2020) Association between Epstein-Barr virus infection and gastric cancer: a systematic review and meta-analysis. *BMC Cancer.* **20**, 493.
  58. Suraweera N., Duval A., Reperant M., Vaury C., Furlan D., Leroy K., Seruca R., Iacopetta B., Hamelin R. (2002) Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. *Gastroenterology.* **123**, 1804–1811.
  59. Polom K., Marrelli D., Roviello G., Pascale V., Voglino C., Vindigni C., Generali D., Roviello F. (2018) *PIK3CA* mutation in gastric cancer and the role of microsatellite instability status in mutations of exons 9 and 20 of the *PIK3CA* gene. *Adv. Clin. Exp. Med.* **7**, 963–969.
  60. Van Grieken N.C.T., Aoyama T., Chambers P.A., Bottomley D., Ward L.C., Inam I., Buffart T.E., Das K., Lim T., Pang B., Zhang S.L., Tan I.B., Carvalho B., Heideman D.A.M., Miyagi Y., Kameda Y., Arai T., Meijer G.A., Tsuburaya A., Tan P., Yoshikawa T., Grabsch H.I. (2013) *KRAS* and *BRAF* mutations are rare and related to DNA mismatch repair deficiency in gastric cancer from the East and the West: results from

- a large international multicentre study. *Br. J. Cancer*. **108**, 1495–1501.
61. Barbi S., Cataldo I., De Manzoni G., Bersani S., Lamba S., Mattuzzi S., Bardelli A., Scarpa A. (2010) The analysis of *PIK3CA* mutations in gastric carcinoma and metanalysis of literature suggest that exon-selectivity is a signature of cancer type. *J. Exp. Clin. Cancer Res.* **29**, 32.
62. Lu W., Wei H., Li M., Wang H., Liu L., Zhang Q., Liu L., Lu S. (2015) Identification of *KRAS* and *PIK3CA* but not *BRAF* mutations in patients with gastric cancer. *Mol. Med. Rep.* **1**, 1219–1224.
63. Бесова Н. С., Трякин А. А., Артамонова Е. В., Болотина Л. В., Калинин А. Е., Кононец П. В., Лядов В. К., Малихова О. А., Неред С. Н., Проценко С. А., Рябов А. Б., Стилиди И. С., Тер-Ованесов М. Д., Хомяков В. М. (2020) Практические рекомендации по лекарственному лечению рака желудка. *Злокачественные опухоли*. **10**, 339–354.
64. Muti H.S., Heij L.R., Keller G., Kohlruss M., Langer R., Dislich B., Cheong J.H., Kim Y.W., Kim H., Kook M.C., Cunningham D., Allum W.H., Langley R.E., Nankivell M.G., Quirke P., Hayden J.D., West N.P., Irvine A.J., Yoshikawa T., Oshima T., Huss R., Grosser M., Roviello F., d'Ignazio A., Quaas A., Alakus H., Tan X., Pearson A.T., Luedde., Ebbert M.P., Jager D., Trautwein C., Gaisa N.T., Grabsch H.I., Kather J.N. (2021) Development and validation of deep learning classifiers to detect Epstein–Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. *Lancet. Digital Health.* **3**, e654–e664.
65. Ботезату И. В., Панчук И. О., Строганова А. М., Сендерович А. И., Кондратова В. Н., Шелепов В. П., Лихтенштейн А. В. (2017) Сканирование мутаций *KRAS*, *NRAS*, *BRAF* и *PIK3CA* методом плавления ДНК с использованием зондов TaqMan. *Молекуляр. биология*. **51**, 50–58.

## Landscape of *KRAS*, *BRAF*, *PIK3CA* Genes Mutations and Clinical Features of EBV-Associated and MSI Gastric Cancer

A. M. Danishevich<sup>1</sup>\*, N. I. Posphehova<sup>2</sup>, A. M. Stroganova<sup>2</sup>, D. A. Golovina<sup>2</sup>, M. P. Nikulin<sup>2</sup>, A. E. Kalinin<sup>2</sup>, S. E. Nikolaev<sup>1</sup>, I. S. Stilidi<sup>2</sup>, and L. N. Lyubchenko<sup>3,4</sup>

<sup>1</sup>Moscow Clinical Scientific and Practical Center named after A.S. Loginov of Moscow Department of Health, Moscow, 111123 Russia

<sup>2</sup>Blokhin National Medical Research Center of Oncology of the Ministry of Health of the Russian Federation, Moscow, 115478 Russia

<sup>3</sup>National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation, Moscow, 125284 Russia

<sup>4</sup>Lopatkin Scientific Research Institute of Urology and Interventional Radiology – Branch of the National Medical Research Radiological Centre of the Ministry of Health of the Russian Federation, Moscow, 105425 Russia

\*e-mail: danisham7@gmail.com

Personalization of gastric cancer treatment is an urgent problem due to clinical heterogeneity and aggressive course of the disease. In 2014, Cancer Genome Atlas researchers divided gastric cancers into four subtypes based on molecular characteristics: Epstein–Barr virus positive (EBV+), microsatellite instability (MSI), chromosomal instability, genomically stable. To date, there is no single method for detecting chromosomal instability and genomically stable subtypes, while MSI analysis and EBV assessment are used in routine practice and are of the greatest clinical importance. We analyzed 159 gastric cancer samples for the presence of MSI, EBV DNA, and somatic mutations in codons 12–13 (exon 2), 61 (exon 3), and 146 (exon 4) of the *KRAS* gene, codons 597–601 (exon 15) of the *BRAF* gene and codons 542–546 (exon 9), 1047–1049 (exon 20) of the *PIK3CA* gene. As a result, the EBV+ gastric cancer was detected in 8.2% of samples, MSI – in 13.2%. MSI and EBV+ were found to be mutually exclusive. The mean age of patients with EBV+ and MSI cancers was 54.8 and 62.1 years, respectively. In 92.3% of EBV+ cancer was detected in men, of which 76.2% were older than 50 years. Diffuse and intestinal adenocarcinomas in EBV+ cancer accounted for 6 (46.2%) and 5 (38.5%) cases, respectively. MSI occurred in almost equal proportions in men and women ( $n = 10$ ; 47.6%,  $n = 11$ ; 52.4%), with a predominance of intestinal histological type (71.4%) and lesion of the lesser curvature (28.6%). One case of EBV+ cancer was diagnosed with the E545K variant in the *PIK3CA* gene. A combination of variants in the *KRAS* and *PIK3CA* genes was found in all MSI cases. The EBV+ subtype was associated with a better prognosis. Overall five-year survival rates for MSI and EBV+ cancers were 100.0 and 54.7%, respectively.

**Keywords:** gastric cancer, molecular classification, Epstein–Barr virus, microsatellite instability, *KRAS*, *BRAF*, *PIK3CA*

ГЕНОМИКА.  
ТРАНСКРИПТОМИКА

УДК 577.2+577.29

ЛИНИЯ ТРАНСГЕННЫХ МЫШЕЙ APP<sup>swe</sup>/PS1dE9/Blg  
ДЛЯ МОДЕЛИРОВАНИЯ ЦЕРЕБРАЛЬНОГО АМИЛОИДОЗА  
ПРИ БОЛЕЗНИ АЛЬЦГЕЙМЕРА

© 2023 г. Е. А. Лысикова<sup>а, \*</sup>, Е. В. Кузубова<sup>б</sup>, А. И. Радченко<sup>б</sup>, Е. А. Патраханов<sup>б</sup>,  
К. Д. Чапров<sup>а, б</sup>, М. В. Корокин<sup>б</sup>, А. В. Дейкин<sup>б</sup>, О. С. Гудырев<sup>б</sup>, М. В. Покровский<sup>б</sup>

<sup>а</sup>Институт физиологически активных веществ Российской академии наук, Черноголовка, 142432 Россия

<sup>б</sup>Белгородский государственный национальный исследовательский университет, Белгород, 308015 Россия

\*e-mail: lysikova.ipac@gmail.com

Поступила в редакцию 08.06.2022 г.

После доработки 30.07.2022 г.

Принята к публикации 30.07.2022 г.

Болезнь Альцгеймера – наиболее распространенная протеинопатия, при которой у пациента наблюдается неуклонное снижение когнитивных функций с одновременным накоплением в тканях мозга амилоидных бляшек – внеклеточных агрегатов амилоида  $\beta$  (A $\beta$ ), ассоциированных с нейровоспалением и нейродегенерацией. Крысы и мыши, A $\beta$  которых отличаются от A $\beta$  человека и других млекопитающих тремя аминокислотными заменами, не подвержены патологии альцгеймеровского типа. Однако в мозге трансгенных мышей со сверхэкспрессией A $\beta$  человека наблюдается появление амилоидных бляшек, что позволяет широко использовать этих животных в биомедицине для моделирования ключевых аспектов болезни Альцгеймера. Трансгенные мыши линии APP<sup>swe</sup>/PS1dE9 часто используются для исследования молекулярных механизмов болезни Альцгеймера. В настоящей работе нами охарактеризована линия мышей APP<sup>swe</sup>/PS1dE9/Blg, полученная в результате скрещивания линии APP<sup>swe</sup>/PS1dE9 на генетическом фоне CH3 с животными C57Bl6/Chg. Нами показано, что показатели фертильности и выживаемости потомства у животных данной линии не отличаются от показателей у контрольных животных дикого типа. Гистологический анализ мозга животных APP<sup>swe</sup>/PS1dE9/Blg подтвердил сохранение нейроморфологических признаков БА с увеличением количества и размеров амилоидных бляшек в мозге по мере старения животных. Таким образом животные линии APP<sup>swe</sup>/PS1dE9/Blg являются удобной моделью при поиске терапевтических стратегий замедления прогрессии БА.

**Ключевые слова:** болезнь Альцгеймера, трансгенные мыши, бета-амилоид, APP, PSEN1

**DOI:** 10.31857/S0026898423010081, **EDN:** AXGEVN

## ВВЕДЕНИЕ

Болезнь Альцгеймера (БА) – нейродегенеративное заболевание и самая распространенная форма деменции [1]. БА относится к протеинопатиям – группе заболеваний, в патогенезе которых важная роль отводится изменениям структуры и/или нарушениям метаболизма склонных к агрегации белков. БА характеризуется медленной прогрессией патологического процесса и длительной продромальной стадией [2]. На сегодняшний день еще не создано эффективных методов терапии БА, а широкая распространенность делает это заболевание актуальной социальной проблемой.

В начале 20 века немецкий психиатр Алоис Альцгеймер обнаружил в патогистологических

препаратах мозга пациентов с деменцией включения – сенильные бляшки. Позже было установлено, что одним из основных компонентов сенильных бляшек является агрегированный бета-амилоидный (A $\beta$ ) пептид. A $\beta$ -пептид представляет собой продукт гидролиза белка-предшественника амилоида APP (amyloid precursor protein) [3]. Согласно принятой на сегодняшний день гипотезе амилоидного каскада, церебральный амилоидоз запускает остальные патологические процессы при прогрессии БА [4].

APP – широко экспрессируемый трансмембранный белок первого типа имеет три основные изоформы: APP695, APP751 и APP770. Изоформа APP695 является основной в нейронах, тогда как APP751 и APP770 экспрессируются преимущественно в других типах клеток [5]. Белок APP играет важную роль в разрастании нейритов и ветв-

Сокращения: БА – болезнь Альцгеймера; APP – белок-предшественник амилоида; PSEN1 – пресенилин-1.

лении аксонов в процессе развития мозга, он участвует в синаптогенезе и обеспечении синаптической пластичности [5]. Белок APP в норме подвергается неамилоидогенному процессингу с последовательным расщеплением  $\alpha$ - и  $\gamma$ -секретазами внутри домена A $\beta$  с образованием непатогенных фрагментов (sAPP $\alpha$  и C-концевых фрагментов (CTF)) [6]. Растворимая форма sAPP $\alpha$  обладает нейротрофической активностью и участвует в защите нейронов от эксайтотоксичности, регулируя гомеостаз кальция [5]. Последовательное протеолитическое расщепление APP  $\beta$ - и  $\gamma$ -секретазами приводит к образованию A $\beta$ -пептида, способного к агрегации, а также непатогенных фрагментов [7]. Наиболее распространены две формы A $\beta$ -пептида, состоящие из 40 (A $\beta$ 40) и 42 аминокислотных остатков (A $\beta$ 42), при этом форма A $\beta$ 42 более токсична для клетки [8].

БА считается многофакторным заболеванием, наиболее распространены его спорадические формы с поздним дебютом (старше 65 лет). Исследование семейных форм БА с более ранним дебютом (менее 65 лет) позволило выявить генетические факторы, ассоциированные с развитием нейродегенеративных процессов в мозге больных. На сегодняшний день известно несколько мутаций в гене *APP*, приводящих к существенно повышению уровня A $\beta$ 42 и ассоциированных с ранним дебютом БА [9]. Описаны также мутации в генах пресенилина 1 (*PSEN1*) и пресенилина 2 (*PSEN2*), кодирующих трансмембранные белки эндоплазматического ретикула. Геномные исследования выявили более 170 различных мутаций в гене *PSEN1* и 13 мутаций в гене *PSEN2*, связанных с наиболее распространенными семейными формами БА с ранним началом [7]. Белки *PSEN1* и *PSEN2*, входящие в состав комплекса  $\gamma$ -секретазы, оказывают влияние на процессинг APP [10]. Мутации в гене *PSEN1* увеличивают специфичность расщепления белка APP  $\gamma$ -секретазой до токсичного пептида A $\beta$ 42 [11], что подтверждается повышением соотношения A $\beta$ 42/A $\beta$ 40 в плазме пациентов с мутантным геном *PSEN1* по сравнению с плазмой крови пациентов со спорадическими формами БА [12].

С целью выявления молекулярных механизмов, лежащих в основе развития и прогрессии БА, созданы животные модели. К наиболее адекватным и широко используемым моделям относятся линии трансгенных мышей, в которых воспроизводится прогрессирующий церебральный амилоидоз, характерный для БА. В геном мышей были введены трансгенные кассеты, кодирующие наиболее часто встречающиеся мутации в генах, ассоциированных с развитием наследственных форм БА. Животные модели первого поколения, основанные преимущественно на сверхэкспрессии кДНК мутантных генов *APP* или *PSEN1* человека, воспроизводят основные патологические

признаки заболевания. В модельных животных второго поколения проведена замена участков (knock-in) гена *App* мыши на участки, кодирующие шведскую, иберийскую и/или арктическую мутации (*App*<sup>NL-G-F</sup>). Трансгенная линия мышей третьего поколения содержала knock-in мутации в генах *App* и *Psen1* (*App*<sup>NL-F</sup> *Psen1*<sup>P117L/WT</sup>) [13]. Трансгенные животные, воспроизводящие формирование амилоидных включений — основной тип патологии при БА, представлены в табл. 1.

Каждая линия отличалась вариабельностью патологического фенотипа, зависящего от генетического фона, условий содержания и длительности поддержания обособленных колоний в условиях одного вивария. В нашей работе подробно охарактеризована сублиния APP<sup>swe</sup>/PS1<sup>dE9</sup>.

Геном трансгенных мышей линии APP<sup>swe</sup>/PS1<sup>dE9</sup> содержит две трансгенные кассеты, встроенные в один локус на хромосоме 9. Первая трансгенная кассета кодирует кДНК химерного гуманизированного в домене A $\beta$  и содержащего шведскую мутацию (K670N/M671L) гена *APP* под контролем прионного белка мыши. Вторая трансгенная кассета кодирует кДНК гена *PSEN1* человека с делецией экзона 9 (PS1 $\Delta$ E9) под контролем прионного белка мыши. Бигенная линия животных APP/PS1 создана Borchelt и соавт. путем коинъекции линейаризованной и очищенной плазмидной ДНК в пронуклеус зиготы мыши [19]. У 10% животных, полученных после коинъекции трансгенных кассет, выявлено одновременное встраивание обоих трансгенов [31]. Далее отобрали одну линию с наибольшим количеством копий трансгенной кассеты, которую использовали в дальнейших экспериментах. Высокий уровень трансгенного белка *PSEN1* и трансгенного APP в мозге модельных мышей подтвержден методом иммуноблоттинга [19].

При создании трансгенных животных внесли модификацию в фрагмент гена прионоподобного белка (PrP) мыши в векторе Bluescript KS<sup>+</sup>. Обозначенный phgPrP фрагмент в составе гена *PrP* содержал последовательность размером примерно 6 т.п.н. до сайта инициации транскрипции, состоящую из первого кодирующего экзона, первого интрона, кодирующую область *PrP* мыши, слитую со вторым экзоном, и примерно 3 т.п.н. 3'-нетранслируемой области. Открытую рамку считывания во фрагменте phgPrP заменили на сайт рестрикции XhoI. Полученный модифицированный вектор назвали MoPrP.Xho. В вектор MoPrP.Xho по сайтам рестрикции XhoI между экзонами 2 и 3 гена *PrP* мыши встроили полную кДНК гена *PSEN1* человека, кодирующую белок с делецией аминокислотных остатков 290–319, фланкированную сайтами XhoI. кДНК белка-предшественника амилоида, гуманизированного в домене A $\beta$ , содержащую шведскую му-

**Таблица 1.** Наиболее часто используемые линии трансгенных мышей, моделирующие церебральный амилоидоз

Линия (JAX mice)	Мутация	Промотор	Амилоидные бляшки (возраст)	Ссылка
PDAPP	hAPP (V717F)	Промотор гена фактора роста тромбоцитов- $\beta$ (PDGF- $\beta$ )	6–9 мес.: кора, гиппокамп, мозолистое тело	[14]
APP23 (B6.Cg-Tg(Thy1-APP)3Somm/J)	hAPP <sub>751</sub> (K670N/M671L)	Промотор <i>Thy1</i> мыши	6–8 мес.: кора 12–24 мес.: кора, гиппокамп, таламус, миндаля	[15]
Tg2576 (B6;SJL-Tg(APP <sup>SWE</sup> )2576Kha)	hAPP <sub>695</sub> (K670N/M671L)	Промотор гена прионного белка хомяка	7–8 мес.: кора 11–13 мес.: гиппокамп, мозжечок	[16]
Tg-SwDI (C57BL/6-Tg(Thy1-APP <sup>SwDutIowa</sup> )BWevn/Mmjax)	hAPP <sub>770</sub> (K670N/M671L) hAPP <sub>770</sub> (E693Q) hAPP <sub>770</sub> (D694N)	Промотор <i>Thy1</i> мыши	3 мес.: кора, гиппокамп 6 мес.: обонятельные луковицы, таламус	[17]
3xTg (B6;129-Tg(APP <sup>Swe</sup> ,tauP301L)1LfaPsen1 <sup>tm1Mpm</sup> /Mmjax)	hAPP (K670N/M671L) hМАРТ (P301L) mPsen1 (M146V)	Промотор <i>Thy1</i> мыши (APP, МАРТ), промотор <i>Psen1</i> мыши	6 мес.: кора, гиппокамп 12 мес.: гиппокамп	[18]
APP <sup>Swe</sup> /PS1 $\Delta$ E9 (B6;C3-Tg(APP <sup>Swe</sup> ,PSEN1 $\Delta$ E9)85Dbo/Mmjax)	Mo/HuAPP (K670N/M671L) hPSEN1 $\Delta$ E9	Промотор гена прионного белка мыши	5–6 мес.: кора, гиппокамп	[19]
5xFAD (B6SJL-Tg(APP <sup>SwFILon</sup> ,PSEN1* <sup>M146L</sup> * <sup>L286V</sup> )6799Vas/Mmjax)	hAPP (K670N/M671L) hAPP (V717I) hAPP (I716V) hPSEN1 (M146L) hPSEN1(L286V)	Промотор <i>Thy1</i> мыши	2 мес.: кора, основание гиппокампа 4 мес.: спинной мозг 6 мес.: гиппокамп, таламус, обонятельные луковицы, основание мозга	[20]
J20 (B6.Cg-Zbtb <sup>20Tg</sup> (PDGF <sup>B</sup> -APP <sup>SwInd</sup> )20Lms/2Mmjax)	hAPP (K670N/M671L) hAPP (V717F)	Промотор гена фактора роста тромбоцитов- $\beta$ (PDGF- $\beta$ )	5–8 мес.: кора, гиппокамп	[21]
APPPS1 (B6.Cg-Tg(Thy1-APP <sup>Sw</sup> ,Thy1-PSEN1* <sup>L166P</sup> )21Jckr)	hAPP (K670N/M671L) hPSEN1 (L166P)	Промотор <i>Thy1</i> мыши	6 недель: кора 2–3 мес.: гиппокамп 3–5 мес.: стриатум, таламус	[22]
PS/APP	hAPP (K670N/M671L) hPSEN1 (M146L)	Промотор гена прионного белка хомяка, промотор гена фактора роста тромбоцитов- $\beta$ (PDGF- $\beta$ )	6 мес.: кора, гиппокамп 12 мес.: стриатум, таламус, ствол мозга	[23–25]
PS2Tg2576	hAPP (K670N/M671L) hPSEN2 (N141I)	Промотор гена прионного белка хомяка, промотор гена $\beta$ -актина курицы	2–3 мес.: кора, гиппокамп	[26]

Таблица 1. Окончание

Линия (JAX mice)	Мутация	Промотор	Амилоидные бляшки (возраст)	Ссылка
mThy1-hAβPP751	hAPP (K670N/M671L) hAPP (V717I)	Промотор <i>Thy1</i> мыши	3–6 мес.: кора 5–7 мес.: гиппокамп, таламус, обонятельные луковицы	[27]
<i>PLB1-triple</i>	hAPP (K670N/M671L) hAPP (V717I) hМАРТ (P301L) hМАРТ (R406W) hPSEN1 (A246E)	Промотор <i>CaMKII-α</i> мышь (APP, МАРТ), промотор гена прионного белка мыши (PSEN1)	5–6 мес.: кора, гиппокамп	[28]
<i>knock-in APP</i> (App <sup>NL-G-F</sup> )	Mo/HuAPP (K670N/M671L) Mo/HuAPP (E693G) Mo/HuAPP (I716F)	Промотор гена <i>APP</i> мыши	2 мес.: кора 4 мес.: гиппокамп	[29]
<i>knock-in App</i> <sup>NL-F</sup> Psen1 <sup>P117L/WT</sup>	Mo/HuAPP (K670N/M671L) Mo/HuAPP (I716F) mPsen1 (P117L)	Промотор гена <i>APP</i> мышь, промотор <i>Psen1</i> мыши	3 мес.: кора 12 мес.: гиппокамп	[30]

тацию (K670N/M671L), также клонировали в вектор MoPrP.Xho между экзонами 2 и 3 *PrP* [19]. Каждый трансген экспрессируется под контролем промотора прионоподобного белка мыши. Этот промотор активен в эмбриональном развитии, у взрослых животных он обеспечивает экспрессию трансгена преимущественно в астроцитах и нейронах центральной нервной системы [31].

### ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

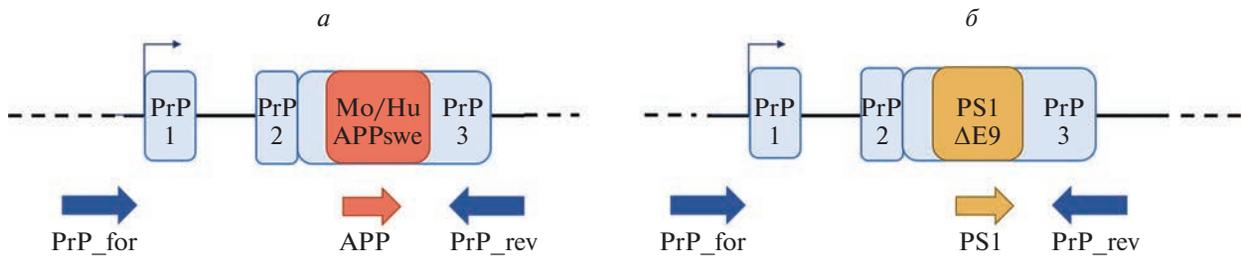
**Лабораторные животные.** В работе использовали сублинию мышей APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg, полученную от скрещивания мышей B6;C3-Tg(APP<sup>swe</sup>,PSEN1<sup>dE9</sup>)85Dbo/Mmjax (#034829-JAX) с мышами дикого типа C57Bl6J/ChG. Линия C57Bl6J/ChG получена из питомника Charles River UK и поддерживалась в качестве обособленной колонии в условиях беспатогенного вивария ИФАВ РАН в течение 7 лет. Поддержание и размножение колонии проводили путем скрещивания трансгенных животных с животными дикого типа из тех же пометов.

Экспериментальных и контрольных животных содержали в условиях беспатогенного вивария Белгородского государственного национального исследовательского университета (НИУ БелГУ) в условиях искусственно регулируемого светового дня (12 ч день и 12 ч ночь) при температуре +22–26°C со свободным доступом к корму и воде.

**Генотипирование трансгенных животных.** Геномную ДНК получали из биопсийного материала уха (около 30 мг ткани). Материал помещали в

лизирующий раствор, содержащий 100 мМ NaCl, 50 мМ Трис-НСl (рН 8.0), 2 мМ EDTA и 2 мг/мл протеиназы К, и инкубировали при 55°C в течение 12–16 ч после чего реакционную смесь прогревали при 85°C в течение 40 мин. Лизат центрифугировали в течение 1 мин при 10000 g, 1 мкл супернатанта использовали в качестве матрицы в реакции ПЦР [32].

**Праймеры.** Использовали праймеры, предложенные Jankowsky и соавт. [19], амплифицирующие одновременно участок трансгенной кассеты (при ее наличии) и участок геномной ДНК мыши. В реакции использовали смесь из трех праймеров: общий обратный праймер PrP\_rev (5'-GTG GAT ACC CCC TCC CCC AGC CTA GAC C), гомологичный последовательности гена *PrP* в геноме мыши и в составе трансгенной кассеты; прямой праймер PrP\_for (5'-CCT CTT TGT GAC TAT GTG GAC TGA TGT CGG), гомологичный участку геномного *PrP* мыши, удаленного из вектора MoPrP.Xho; специфические прямые праймеры PS1 (5'-CAG GTG GTG GAG CAA GAT G) и APP (5'-CCG AGA TCT CTG AAG TGA AGA TGG ATG), гомологичные последовательностям генов мутантных белков в составе трансгенной кассеты. Реакционная смесь содержала 1× Taq Turbo буфер (“Евроген”, Россия), по 0.2 мМ каждого dNTP (“Евроген”), 0.5 мкМ каждого прямого праймера и 1 мкМ общего обратного праймера, 2 ед. HS Taq-ДНК-полимеразы (“Евроген”). Программа амплификации: 1 цикл – 95°C, 3 мин; 30 циклов – 95°C, 20 с; 55°C (APP) или 65°C (PS1), 20 с; 72°C, 20 с; 1 цикл – 72°C, 2 мин.



**Рис. 1.** Схема положения праймеров для детекции трансгенных кассет, содержащих кДНК гена белка-предшественника амилоида со шведской мутацией (Mo/Hu APP<sub>swe</sub>) (а) и кДНК гена пресенилина 1 человека с делецией экзона 9 (PS1ΔE9) в геноме мышей APP<sub>swe</sub>/PS1ΔE9 (б). Обе кассеты расположены в одном локусе на хромосоме 9 и экспрессируются каждая под собственным промотором гена прионного белка мыши (*PrP*). Прямой праймер PrP\_for гомологичен участку эндогенного прионного белка мыши. Обратный праймер PrP\_rev гомологичен последовательности эндогенного *PrP* и последовательности гена *PrP* в составе трансгенной кассеты. Прямые праймеры APP и PS1 специфично узнают последовательность внутри соответствующих трансгенных кассет. Обозначения: PrP1 – экзон 1, PrP2 – экзон 2, PrP3 – экзон 3, стрелками обозначены прямые и обратные праймеры.

Продукты амплификации разделяли с помощью электрофореза в 1.5%-ном агарозном геле в 1× TAE буфере (40 мМ Трис-НСl, 20 мМ уксусная кислота, 1 мМ EDTA, 0.5 мкг/мл бромистого этидия) при 120 В в течение 30 мин.

#### Приготовление гистологических препаратов.

Животных подвергали терминальной анестезии, мозг диссектировали и фиксировали в растворе Карнуа (6 частей 96%-ного этилового спирта, 3 части хлороформа, 1 часть ледяной уксусной кислоты) в течение ночи. Ткань дегидратировали, используя последовательное проведение в растворах этилового спирта с увеличивающейся концентрацией: 75% 1 ч; 96% (I) 5 мин; 96% (II) 45 мин; 100% (I) 5 мин; 100% (II) 10 мин. Далее последовательно инкубировали в течение 30 мин в смеси 100%-ный этиловый спирт : хлороформ (1 : 1), 1 ч в хлороформе (I), оставляли на ночь в хлороформе (II), после чего ткани пропитывали парафином (три смены по 1 ч) при 60°C. Парафиновые блоки подготавливали на станции для заливки Leica EG 1160 (“Leica Biosystems”, ФРГ). Парафиновые срезы толщиной 8 мкм монтировали на предметные стекла с полилизинным покрытием. Срезы депарафинизировали в течение 20 мин в ксилоле, регидратировали, инкубируя последовательно в этиловом спирте: 100% – 10 мин; 95% – 5 мин; 50% – 5 мин; далее трижды промывали в деионизованной воде по 5 мин. Срезы окрашивали раствором красителя Конго красный (0.5% Конго красный в 50%-ном этиловом спирте) в течение 5 мин и дифференцировали в растворе 0.2%-ного КОН в 80%-ном этиловом спирте в течение 1 мин, промывали 3 раза по 5 мин в деионизованной воде, заключали в среду на водной основе Immu-Mount™ (“Thermo Scientific”, США).

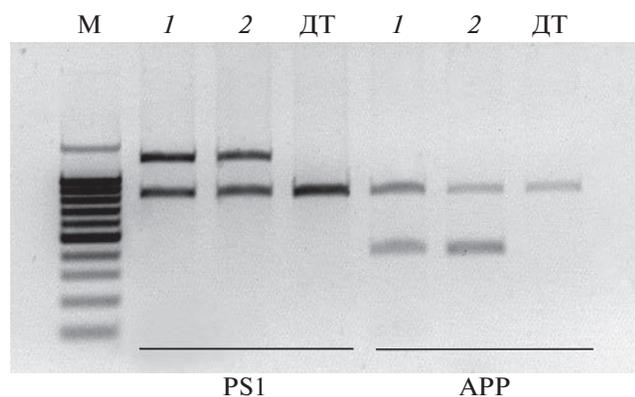
## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Сублиния мышей APP<sub>swe</sub>/PS1ΔE9/Blg получена от скрещивания животных B6;C3-

Tg(APP<sub>swe</sub>,PSEN1ΔE9)85Dbo/Mmjax на смешанном генетическом фоне C3H/C57Bl6 с животными C57Bl6J/ChG. Для формирования репрезентативных групп экспериментальных и контрольных животных, синхронизированных по возрасту, сначала получали нужное количество родственных производителей. Гемизиготных по трансгенной кассете мышей скрещивали с мышами дикого типа на том же смешанном генетическом фоне, полученными в предыдущих скрещиваниях. Потомство генотипировали методом конвенционной ПЦР с использованием одной пары специфичных праймеров – либо на *PSEN1*, либо на *APP*, поскольку в геноме мыши обе трансгенные кассеты локализованы в одном локусе на хромосоме 9. Схематическое изображение положения праймеров для амплификации трансгенной кассеты представлено на рис. 1. Пример генотипирования по каждой из трансгенных кассет приведен на рис. 2.

Встраивание двух трансгенных кассет в один локус в геноме мышей, показанное в [19], позволяет проводить генотипирование лишь по одной из трансгенных кассет. Результаты проведенного нами генотипирования подтверждают присутствие обеих трансгенных кассет в нескольких поколениях животных. Таким образом, для генотипирования мышей линии APP<sub>swe</sub>/PS1ΔE9/Blg можно проводить ПЦР с праймерами к одному из трансгенных векторов. При этом мы рекомендуем подтверждать присутствие обеих трансгенных кассет в генотипах мышей после окончательного формирования экспериментальных и контрольных групп животных в каждом третьем–четвертом поколении.

Для формирования экспериментальных групп мышей APP<sub>swe</sub>/PS1ΔE9/Blg скрещивали с особями дикого типа из предыдущих пометов. Нами не выявлено значительных различий в выживаемости трансгенных животных и животных дикого типа в тех же пометах. Так, анализ результатов ге-



**Рис. 2.** Детекция трансгенных кассет в геноме мышей APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg методом ПЦР. Все образцы содержат продукт амплификации (750 п.н.) гена прионного белка мыши в качестве эндогенного контроля. Животные дикого типа (ДТ) содержат только фрагмент 750 п.н. У трансгенных животных APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg (1 и 2) появляется дополнительный фрагмент размером 1.3 т.п.н. после амплификации последовательности мутантного гена *PSEN1* или фрагмент размером 400 п.н. после амплификации последовательности мутантного *APP*. Маркер 100+ п.н. (“Евроген”).

нотипирования 315 мышей показал, что трансгенную кассету имели 52.4% потомков, 47.6% не имели модификации генома.

Нами сформированы пары производителей, состоящие из трансгенных самок и самцов дикого типа и пары из трансгенных самцов и самок дикого типа. Возраст животных-производителей 2–4 мес. Анализ фертильности самок APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg показал, что из общего числа ссаженных самок фертильными оказались 74%, тогда как среди самок дикого типа фертильными были 84%. Пересчет общего количества рожденных потомков на одну самку показал, что и самки APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg, и самки дикого типа рожают одинаковое число потомков (табл. 2).

Нами показано, что показатели выживаемости и фертильности у животных линии APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg не отличаются существенно от показателей у контрольных животных дикого типа. Полученные нами данные позволяют планировать формирование пар производителей для составления экспериментальных групп животных и получения расчетного числа трансгенных животных в экспериментальных когортах.

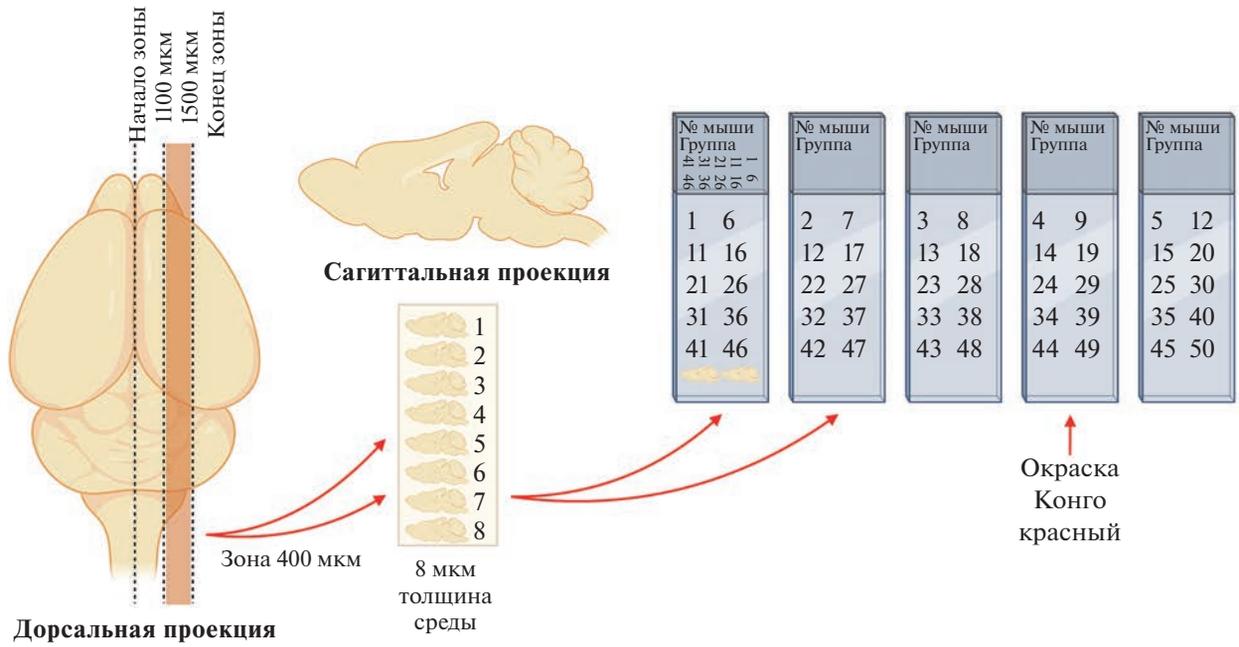
Отличительной особенностью включений амилоидного типа является кросс- $\beta$ -структура белка в их составе, обеспечивающая связывание со специфическими красителями, к которым относятся Конго красный и тиофлавин S [2]. Сроки регистрации дебюта патогистологических признаков церебрального амилоидоза в мозге мышей APP<sup>swe</sup>/PS1<sup>dE9</sup> и динамика прогрессии нейродегенеративного процесса, полученные в разных лабораториях, несколько различаются. Наиболее часто патологию, уже широко распространенную в гиппокампе и коре головного мозга, выявляют в возрасте 6 мес. [33]. Данные о прогрессии амилоидоза различаются в зависимости от метода анализа, размера амилоидных включений (подсчет только крупных или средних и мелких включений), использования поправки Аберкромби для нивелирования возможности подсчета одного и того же включения на нескольких серийных срезах, от способов окрашивания тканей – использование специальных красителей (Конго красный или тиофлавин S) или антител к бета-амилоиду, а также от способа анализа окрашенных срезов – детекции включений только в диапазоне определенных длин волн или с дополнительным использованием поляризованного света [4, 34, 35]. Исключительно важной представляется задача подробной характеристики данной сублинии для ее использования в качестве модельной системы церебрального амилоидоза.

Нами проведен гистологический анализ мозга мышей APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg в возрасте 5.5 и 10 мес. Принцип формирования групп срезов для морфометрического анализа включений представлен на рис. 3.

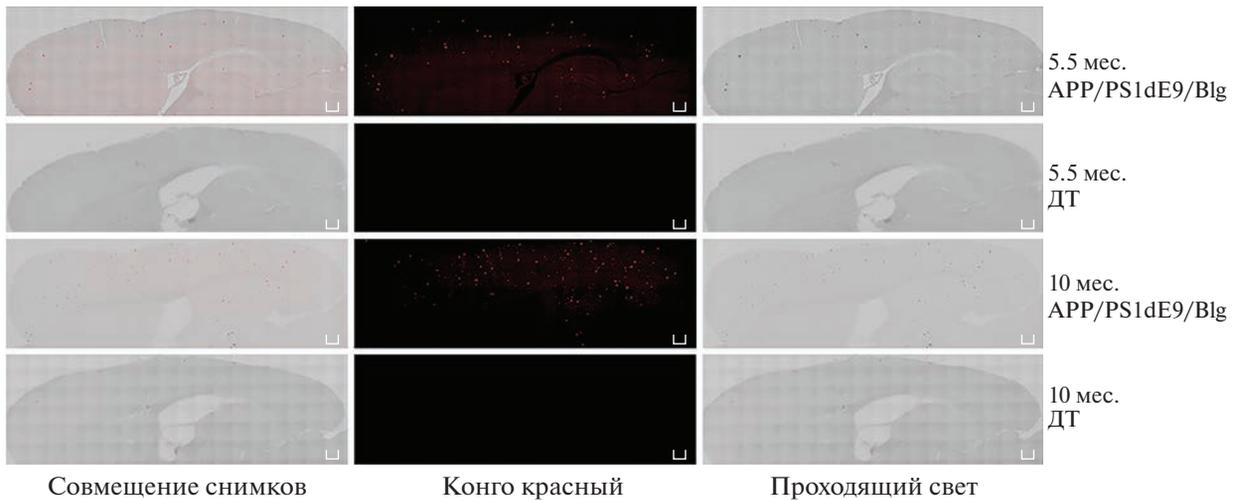
Для оценки числа амилоидных включений в гиппокампе и коре мозга мышей APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg срезы полушария головного мозга окрашивали красителем Конго красным (рис. 4). Показано, что уже в возрасте 5.5 мес. в коре и гиппокампе трансгенных животных обнаруживаются как мелкие и средние, так и крупные амилоидные включения. В возрасте 10 мес. наблюдается увеличение числа и размера агрегатов в мозге трансгенных мышей. В мозге контрольных животных дикого типа включения амилоида не обнаруживаются. Согласно опубликованным данным, самыми старыми животными APP<sup>swe</sup>/PS1<sup>dE9</sup>, у ко-

**Таблица 2.** Фертильность самок при получении трансгенных животных

Самки-производители	Ссаженные самки, всего	Не забеременевшие самки	Потомки	Среднее число потомков в одном помете
APP <sup>swe</sup> /PS1 <sup>dE9</sup> /blg	35	9	140	5.38
Дикий тип	38	6	175	5.47



**Рис. 3.** Схема раскладки срезов мозга при подготовке гистологических препаратов. Из зоны мозга толщиной 400 мкм формировали пять стекол, состоящих из 10 срезов, на каждое предметное стекло помещали каждый пятый срез мозга.



**Рис. 4.** Гистологический анализ включений Аβ на срезах мозга трансгенных животных APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg и контрольных животных дикого типа в возрасте 5.5 и 10 мес.; окраска Конго красным. Шкала измерений 200 мкм.

торых изучено образование амилоидных включений, были животные в возрасте 24 мес. [33].

Нами определена скорость образования патологических агрегатов Аβ в коре и гиппокампе трансгенных APP/PS1 и контрольных мышей в возрасте 5.5 и 10 мес. (табл. 3).

БА у человека может развиваться в течение десятилетий до момента появления первых симптомов. Нами оценена прогрессия агрегации Аβ в гиппокампе и коре у молодых трансгенных мышей (5.5 мес.) и у стареющих животных (10 мес.). Число включений в гиппокампе стареющих жи-

вотных увеличивается почти в 5 раз и в 4 раза возрастает число агрегатов в коре ( $p < 0.0001$ ,  $t$ -критерий Стьюдента).

Полученные нами данные соотносятся с проведенными ранее экспериментами по подсчету амилоидных бляшек в мозге трансгенных самцов линии APP<sup>swe</sup>/PS1<sup>dE9</sup> – на срезе животных в возрасте 4 мес. в среднем обнаруживалось около 20 бляшек в коре и примерно пять бляшек в гиппокампе, тогда как у животных в возрасте 10 мес. количество включений увеличивалось примерно в 4 раза в коре и в 2 раза в гиппокампе [4].

**Таблица 3.** Морфометрический анализ включений у трансгенных животных APP<sup>swe</sup>/PS1dE9/Blg в возрасте 5.5 и 10 мес.\*

Возраст, мес.	Агрегаты в гиппокампе	Агрегаты в коре
5.5	5.010 ± 0.886	30.11 ± 4.714
10	24.37 ± 2.929	125.1 ± 13.50

\* Представлены средние данные анализа 10 животных.

Свою эффективность в качестве терапевтического средства, снижающего уровень агрегированной формы Аβ *in vivo*, показали инъекции синтетических тетрапептидов НАЕЕ и RADD. Эти тетрапептиды специфически связываются с сайтом EVNH в пептиде Аβ, что приводит к подавлению Zn-опосредованной димеризации и позволяет снизить уровень амилоидных бляшек, обнаруживаемых в коре и гиппокампе трансгенных животных APP<sup>swe</sup>/PS1dE9 в возрасте 7 мес. [36]. Также экспериментально показано, что изомеризация остатка Asp7 и фосфорилирование Ser8 в пептиде Аβ42 может препятствовать Zn-зависимой агрегации. Инъекции синтетического пептида isoD7-pS8-Aβ42 с такими модификациями приводили к статистически значимому снижению числа амилоидных включений более чем в 3 раза в мозге 8-месячных животных APP<sup>swe</sup>/PS1dE9 (7.4 ± 2.8) по сравнению с группой трансгенных животных без инъекции (28.7 ± 4.6) [37]. Инъекции синтетического пептида pS8-Aβ42 с фосфорилированным остатком Ser8 также снижали число агрегатов амилоидного типа в 1.5 раза в гиппокампе 8-месячных трансгенных животных APP<sup>swe</sup>/PS1dE9 (18.3 ± 0.94) по сравнению с контрольными животными без инъекций (28.7 ± 4.6) [38].

Таким образом, число агрегатов Аβ в мозге трансгенных мышей линии APP<sup>swe</sup>/PS1dE9 является важным параметром, который следует учитывать при поиске терапевтических стратегий замедления прогрессии БА.

Работа выполнена при поддержке Министерства науки и высшего образования РФ, (соглашение № 075-15-2021-1346). Работы с животными финансировались из средств Государственного задания лаборатории генетических технологий и геномного редактирования для биомедицины и ветеринарии FZWG-2021-0016.

Работу с животными проводили в соответствии с “Правилами лабораторной практики в Российской Федерации” от 1.04.2016 № 199н. Проведение эксперимента утверждено на заседании этического комитета НИУ БелГУ от 8.02.2021, протокол № 02.21-4.

Авторы сообщают об отсутствии конфликта интересов.

## СПИСОК ЛИТЕРАТУРЫ

- Konttinen H., Cabral-da-Silva M.E.C., Ohtonen S., Wojciechowski S., Shakirzyanova A., Caligola S., Gigugno R., Ishchenko Y., Hernandez D., Fazaludeen M.F., Eamen S., Budia M.G., Fagerlund I., Scoyni F., Korhonen P., Huber N., Naapasalo A., Hewitt A.W., Vickers J., Smith G.C., Oksanen M., Graff C., Kanninen K.M., Lehtonen S., Propson N., Schwartz M.P., Pebay A., Koistinaho J., Ooi L., Malm T. (2019) PSEN1DeltaE9, APP<sup>swe</sup>, and APOE4 confer disparate phenotypes in human iPSC-derived microglia. *Stem. Cell Rep.* **13**, 669–683.
- Шелковникова Т.А., Куликова А.А., Цветков Ф.О., Петерс О., Бачурин С.О., Бухман В.Л., Нинкина Н.Н. (2012) Протеинопатии – формы нейродегенеративных заболеваний, в основе которых лежит патологическая агрегация белков. *Молекуляр. биология.* **46**, 402–415.
- Masters C.L., Bateman R., Blennow K., Rowe C.C., Sperling R.A., Cummings J.L. (2015) Alzheimer’s disease. *Nat. Rev. Dis. Primers.* **1**, 15056.
- Kozin S.A., Cheglakov I.B., Ovsepyan A.A., Telegin G.B., Tsvetkov P.O., Lisitsa A.V., Makarov A.A. (2013) Peripherally applied synthetic peptide isoAsp7-Abeta(1-42) triggers cerebral beta-amyloidosis. *Neurotoxicity Res.* **24**, 370–376.
- Evin G., Li Q.X. (2012) Platelets and Alzheimer’s disease: potential of APP as a biomarker. *W. J. Psychiatry.* **2**, 102–113.
- Vetrivel K.S., Thinakaran G. (2006) Amyloidogenic processing of beta-amyloid precursor protein in intracellular compartments. *Neurology.* **66**, S69–73.
- De Strooper B., Annaert W. (2010) Novel research horizons for presenilins and gamma-secretases in cell biology and disease. *Annu. Rev. Cell. Dev. Biol.* **26**, 235–260.
- Manczak M., Kandimalla R., Yin X., Reddy P.H. (2018) Hippocampal mutant APP and amyloid beta-induced cognitive decline, dendritic spine loss, defective autophagy, mitophagy and mitochondrial abnormalities in a mouse model of Alzheimer’s disease. *Hum. Mol. Genet.* **27**, 1332–1342.
- Pan J.X., Sun D., Lee D., Xiong L., Ren X., Guo H.H., Yao L.L., Lu Y., Jung C., Xiong W.C. (2021) Osteoblastic Swedish mutant APP expedites brain deficits by inducing endoplasmic reticulum stress-driven senescence. *Commun. Biol.* **4**, 1326.
- Татарникова О.Г., Орлов М.А., Бобкова Н.В. (2015) Бета-амилоид и тау-белок: структура, взаимодействие и прионоподобные свойства. *Успехи биол. химии.* **55**, 351–390.

11. Armstrong R.A. (2019) Risk factors for Alzheimer's disease. *Folia Neuropathol.* **57**, 87–105.
12. Jankowsky J.L., Fadale D.J., Anderson J., Xu G.M., Gonzales V., Jenkins N.A., Copeland N.G., Lee M.K., Younkin L.H., Wagner S.L., Younkin S.G., Borchelt D.R. (2004) Mutant presenilins specifically elevate the levels of the 42 residue beta-amyloid peptide *in vivo*: evidence for augmentation of a 42-specific gamma secretase. *Hum. Mol. Genet.* **13**, 159–170.
13. Sasaguri H., Hashimoto S., Watamura N., Sato K., Takamura R., Nagata K., Tsubuki S., Ohshima T., Yoshiki A., Sato K., Kumita W., Sasaki E., Kitazume S., Nilsson P., Winblad B., Saito T., Iwata N., Saido T.C. (2022) Recent advances in the modeling of Alzheimer's disease. *Front. Neurosci.* **16**, 807473.
14. Games D., Adams D., Alessandrini R., Barbour R., Borthette P., Blackwell C., Carr T., Clemens J., Donaldson T., Gillespie F., Guido T., Hagopian S., Johnson-Wood K., Khan K., Lee M., Leibowitz P., Lieberburg I., Little S., Masliah E., McConlogue L., Montoya-Zavala M., Mucke L., Paganini L., Penniman E., Power M., Schenk D., Seubert P., Snyder B., Soriano F., Tan H., Vitale J., Wadsworth S., Wolozin B., Zhao J. (1995) Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein. *Nature.* **373**, 523–527.
15. Kuo Y.M., Beach T.G., Sue L.I., Scott S., Layne K.J., Kokjohn T.A., Kalback W.M., Luehrs D.C., Vishnivetskaya T.A., Abramowski D., Sturchler-Pierrat C., Staufenbiel M., Weller R.O., Roher A.E. (2001) The evolution of A beta peptide burden in the APP23 transgenic mice: implications for A beta deposition in Alzheimer disease. *Mol. Med.* **7**, 609–618.
16. Hsiao K., Chapman P., Nilsen S., Eckman C., Harigaya Y., Younkin S., Yang F., Cole G. (1996) Correlative memory deficits, A beta elevation, and amyloid plaques in transgenic mice. *Science.* **274**, 99–102.
17. Davis J., Xu F., Deane R., Romanov G., Previti M.L., Zeigler K., Zlokovic B.V., Van Nostrand W.E. (2004) Early-onset and robust cerebral microvascular accumulation of amyloid beta-protein in transgenic mice expressing low levels of a vasculotropic Dutch/Iowa mutant form of amyloid beta-protein precursor. *J. Biol. Chem.* **279**, 20296–20306.
18. Oddo S., Caccamo A., Shepherd J.D., Murphy M.P., Golde T.E., Kaye R., Metherate R., Mattson M.P., Akbari Y., LaFerla F.M. (2003) Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular A beta and synaptic dysfunction. *Neuron.* **39**, 409–421.
19. Jankowsky J.L., Slunt H.H., Ratovitski T., Jenkins N.A., Copeland N.G., Borchelt D.R. (2001) Co-expression of multiple transgenes in mouse CNS: a comparison of strategies. *Biomol. Engin.* **17**, 157–165.
20. Oakley H., Cole S.L., Logan S., Maus E., Shao P., Craft J., Guillozet-Bongaarts A., Ohno M., Disterhoft J., Van Eldik L., Berry R., Vassar R. (2006) Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* **26**, 10129–10140.
21. Mucke L., Masliah E., Yu G.Q., Mallory M., Rockenstein E.M., Tatsuno G., Hu K., Kholodenko D., Johnson-Wood K., McConlogue L. (2000) High-level neuronal expression of abeta 1–42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050–4058.
22. Radde R., Bolmont T., Kaeser S.A., Coomaraswamy J., Lindau D., Stoltze L., Calhoun M.E., Jaggi F., Wolburg H., Gengler S., Haass C., Ghetti B., Czech C., Holscher C., Mathews P.M., Jucker M. (2006) Abeta42-driven cerebral amyloidosis in transgenic mice reveals early and robust pathology. *EMBO Rep.* **7**, 940–946.
23. Holcomb L., Gordon M.N., McGowan E., Yu X., Benkovic S., Jantzen P., Wright K., Saad I., Mueller R., Morgan D., Sanders S., Zehr C., O'Campo K., Hardy J., Prada C.M., Eckman C., Younkin S., Hsiao K., Duff K. (1998) Accelerated Alzheimer-type phenotype in transgenic mice carrying both mutant amyloid precursor protein and presenilin 1 transgenes. *Nat. Med.* **4**, 97–100.
24. Kurt M.A., Davies D.C., Kidd M., Duff K., Rolph S.C., Jennings K.H., Howlett D.R. (2001) Neurodegenerative changes associated with beta-amyloid deposition in the brains of mice carrying mutant amyloid precursor protein and mutant presenilin-1 transgenes. *Exp. Neurol.* **171**, 59–71.
25. McGowan E., Sanders S., Iwatsubo T., Takeuchi A., Saido T., Zehr C., Yu X., Uljon S., Wang R., Mann D., Dickson D., Duff K. (1999) Amyloid phenotype characterization of transgenic mice overexpressing both mutant amyloid precursor protein and mutant presenilin 1 transgenes. *Neurobiol. Dis.* **6**, 231–244.
26. Toda T., Noda Y., Ito G., Maeda M., Shimizu T. (2011) Presenilin-2 mutation causes early amyloid accumulation and memory impairment in a transgenic mouse model of Alzheimer's disease. *J. Biomed. Biotechnol.* **2011**, 617974.
27. Rockenstein E., Mallory M., Mante M., Sisk A., Masliah E. (2001) Early formation of mature amyloid-beta protein deposits in a mutant APP transgenic model depends on levels of Abeta(1-42). *J. Neurosci. Res.* **66**, 573–582.
28. Platt B., Drever B., Koss D., Stoppelkamp S., Jyoti A., Plano A., Utan A., Merrick G., Ryan D., Melis V., Wan H., Mingarelli M., Porcu E., Scrocchi L., Welch A., Riedel G. (2011) Abnormal cognition, sleep, EEG and brain metabolism in a novel knock-in Alzheimer mouse, PLB1. *PLoS One.* **6**, e27068.
29. Saito T., Matsuba Y., Mihira N., Takano J., Nilsson P., Itohara S., Iwata N., Saido T.C. (2014) Single App knock-in mouse models of Alzheimer's disease. *Nat. Neurosci.* **17**, 661–663.
30. Sato K., Watamura N., Fujioka R., Mihira N., Sekiguchi M., Nagata K., Ohshima T., Saito T., Saido T.C., Sasaguri H. (2021) A third-generation mouse model of Alzheimer's disease shows early and increased cored plaque pathology composed of wild-type human amyloid beta peptide. *J. Biol. Chem.* **297**, 101004.
31. Jankowsky J.L., Xu G., Fromholt D., Gonzales V., Borchelt D.R. (2003) Environmental enrichment exacerbates amyloid plaque formation in a transgenic mouse model of Alzheimer disease. *J. Neuropathol. Exp. Neurol.* **62**, 1220–1227.
32. Lysikova E.A., Kukharsky M.S., Chaprov K.D., Vasilieva N.A., Roman A.Y., Ovchinnikov R.K., Dey-

- kin A.V., Ninkina N., Buchman V.L. (2019) Behavioural impairments in mice of a novel FUS transgenic line recapitulate features of frontotemporal lobar degeneration. *Genes Brain Behavior*. **18**, e12607.
33. D'Angelo C., Costantini E., Salvador N., Marchioni M., Di Nicola M., Greig N.H., Reale M. (2021) nAChRs gene expression and neuroinflammation in APP<sup>swe</sup>/PS1<sup>dE9</sup> transgenic mouse. *Sci. Rep.* **11**, 9711.
34. Kuhla A., Ruhlmann C., Lindner T., Polei S., Hadlich S., Krause B.J., Vollmar B., Teipel S.J. (2017) APP<sup>swe</sup>/PS1<sup>dE9</sup> mice with cortical amyloid pathology show a reduced NAA/Cr ratio without apparent brain atrophy: a MRS and MRI study. *Neuroimage Clin.* **15**, 581–586.
35. Pezzini A., Del Zotto E., Volonghi I., Giossi A., Costa P., Padovani A. (2009) Cerebral amyloid angiopathy: a common cause of cerebral hemorrhage. *Curr. Med. Chem.* **16**, 2498–2513.
36. Tsvetkov P.O., Cheglakov I.B., Ovsepyan A.A., Mediannikov O.Y., Morozov A.O., Telegin G.B., Kozin S.A. (2015) Peripherally applied synthetic tetrapeptides HAEE and RADD slow down the development of cerebral beta-amyloidosis in AbetaPP/PS1 transgenic mice. *J. Alzheimer's Dis.* **46**, 849–853.
37. Kozin S.A., Barykin E.P., Telegin G.B., Chernov A.S., Adzhubei A.A., Radko S.P., Mitkevich V.A., Makarov A.A. (2018) Intravenously injected amyloid-beta peptide with isomerized Asp7 and phosphorylated Ser8 residues inhibits cerebral beta-amyloidosis in AbetaPP/PS1 transgenic mice model of Alzheimer's disease. *Front. Neurosci.* **12**, 518.
38. Barykin E.P., Petrushanko I.Y., Kozin S.A., Telegin G.B., Chernov A.S., Lopina O.D., Radko S.P., Mitkevich V.A., Makarov A.A. (2018) Phosphorylation of the amyloid-beta peptide inhibits zinc-dependent aggregation, prevents Na,K-ATPase inhibition, and reduces cerebral plaque deposition. *Front. Mol. Neurosci.* **11**, 302.

### APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg Transgenic Mouse Line for Modeling Cerebral Amyloid Angiopathy in Alzheimer's Disease

E. A. Lysikova<sup>1, \*</sup>, E. V. Kuzubova<sup>2</sup>, A. I. Radchenko<sup>2</sup>, E. A. Patrakhanov<sup>2</sup>, K. D. Chaprov<sup>1, 2</sup>,  
M. V. Korokin<sup>2</sup>, A. V. Deykin<sup>2</sup>, O. S. Gudyrev<sup>2</sup>, and M. V. Pokrovskii<sup>2</sup>

<sup>1</sup>*Institute of Physiologically Active Compounds, Russian Academy of Sciences, Chernogolovka, 142432 Russia*

<sup>2</sup>*Belgorod State University, Belgorod, 308015 Russia*

\*e-mail: lysikova.ipac@gmail.com

Alzheimer's disease (AD) is the most common proteinopathy, which is accompanied by a steady decrease in the patient's cognitive functions with simultaneous accumulation of extracellular aggregates of amyloid  $\beta$  (A $\beta$ ) – amyloid plaques in the brain tissues and associated with neuroinflammation and neurodegeneration. Unlike humans and all other mammals, rats and mice have three amino acid substitutions in A $\beta$  and do not reproduce the Alzheimer's pathology. However, the appearance of amyloid plaques is observed in the brains of transgenic mice with the overexpression of human A $\beta$ , which makes it possible to widely use these transgenic animals in biomedicine for the manifestation of AD. Transgenic mouse line APP<sup>swe</sup>/PS1<sup>dE9</sup> is a widely used animal model for the study of the molecular mechanisms of AD. In this paper we provide a detailed description of the APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg subline of animals obtained by crossing APP<sup>swe</sup>/PS1<sup>dE9</sup> mice on a CH3 genetic background with C57Bl6/Chg animals. We have shown no difference in parameters of offspring's survival and fertility of this line compared to wild-type control animals. Histological analysis of the brain of APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg line confirmed the main neuromorphological feature of AD with the progression in number and size of amyloid plaques during aging. Thus, APP<sup>swe</sup>/PS1<sup>dE9</sup>/Blg line is a convenient model in the search for therapeutic strategies for AD.

**Keywords:** Alzheimer disease, transgenic mice, amyloid beta, cerebral amyloid angiopathy

ГЕНОМИКА.  
ТРАНСКРИПТОМИКА

УДК 58.084.1:575.2.084

ИДЕНТИФИКАЦИЯ ФРАГМЕНТА СТИГМАСПЕЦИФИЧНОЙ  
ЭКСПРЕССИИ В ПРОМОТОРЕ ГЕНА ХИТИНАЗЫ КЛАССА I СОИ<sup>1</sup>

© 2023 г. С. М. Zhao<sup>a</sup>, Н. Hou<sup>a</sup>, М. G. Xing<sup>b</sup>, R.-G. Xue<sup>a</sup>, \*

<sup>a</sup>College of Life Sciences, Qingdao Agricultural University, Qingdao, 266109 China

<sup>b</sup>Ubrigene (Jinan) Biosciences Co., Ltd, Jinan, 250000 China

\*e-mail: xuerengao@163.com

Поступила в редакцию 26.04.2022 г.

После доработки 01.08.2022 г.

Принята к публикации 23.08.2022 г.

Уровень экспрессии гетерологичных генов в трансгенных растениях служит важным показателем эффективности работы генов. Тонкая настройка экспрессии трансгенов ограничена небольшим репертуаром известных на данный момент эффективных промоторов. Нами клонирован и охарактеризован тканеспецифичный фрагмент промотора гена хитиназы класса I сои (*GmChi1*). Промотор *GmChi1* (*GmChi1P*) клонировали из сои сорта Jungery. Последовательность промотора содержит ряд предполагаемых *cis*-действующих элементов, включая тканеспецифичные мотивы и мотивы, регулируемые в условиях стресса. По результатам гистохимического анализа самая высокая активность репортерного фермента  $\beta$ -глюкуронидазы (GUS), находящегося под контролем *GmChi1P*, обнаружена в корнях трансгенного растения, *Nicotiana tabacum* cv. NC89, на стадии формирования ростка с четырьмя листьями. Интересно, что обработка салициловой кислотой эффективно подавляла высокую активность GUS в корнях трансгенного табака. Делеционным анализом *GmChi1P* установлено, что последовательности, расположенные между позициями –719 и –382, содержат ключевые *cis*-элементы, ответственные за экспрессию репортерного гена *uidA* (кодирующего GUS) в листьях, корнях и местах поранения *Nicotiana tabacum*. Кроме того, согласно результатам флуориметрического анализа, активность укороченных промоторов: от *ChiP*(–1292) до *ChiP*(–719) – в корнях трансгенного табака значительно подавляется абсцизовой кислотой и полностью – салициловой. Обнаружено также, что промотор *ChiP*(–382) экспрессируется исключительно в рыльце цветков трансгенного растения. С использованием репортерного фермента GUS не обнаружено окрашивания в других органах цветка трансгенного *Nicotiana tabacum*, включая чашелистики, лепестки, пыльники, нити и завязи, а также ни в одной из вегетативных тканей. Полученные результаты свидетельствуют о том, что фрагмент промотора *ChiP*(–382) может быть использован в тканеспецифичной регуляции экспрессии генов и генной инженерии растений.

**Ключевые слова:** *GmChi1*, соя, промотор, трансгенные растения, *Nicotiana tabacum*, стигмаспецифичный промотор

**DOI:** 10.31857/S0026898423010172, **EDN:** AXAYSA

СПИСОК ЛИТЕРАТУРЫ

1. Li W., Yu D., Yu J., Zhu D., Zhao Q. (2018) Functional analysis of maize silk-specific *ZmbZIP25* promoter. *Int. J. Mol. Sci.* **19**, 822–835.
2. Bowles D.J. (1990) Defense related proteins in higher plants. *Annu. Rev. Biochem.* **59**, 873–907.
3. Kenton P., Darby R.M., Shelley G., Draper J. (2000) A *PR5* gene promoter from *Asparagus officinalis* (AoPRT-L) is not induced by abiotic stress, but is activated around sites of pathogen challenge and salicylate in transgenic tobacco. *Mol. Plant Pathol.* **1**, 367–378.
4. Brown R.L., Kazan K., McGrath K.C., Maclean D.J., Manners J.M. (2003) A role for the GCC-box in jasmonate-mediated activation of the *PDF1.2* gene of *Arabidopsis*. *Plant Physiol.* **132**, 1020–1032.
5. Collinge D.B., Kragh K.M., Mikkelsen J.D., Rasmussen U., Vad K. (1993) Plant chitinases. *Plant J.* **3**, 31–40.
6. Sticken M.B., Graham L.S. (1994) Plant chitinases. *Can. J. Bot.* **72**, 1057–1083.
7. Cletus J., Balasubramanian V., Vashisht D., Sakthivel N. (2013) Transgenic expression of plant chitinases to en-

<sup>1</sup> Статья представлена авторами на английском языке.

- hance disease resistance. *Biotechnol. Lett.* **35**, 1719–1732.
8. Abd El-Rahman S.S., Mazen M.M., Mohamed H.I., Mahmoud N.M. (2012) Induction of defence related enzymes and phenolic compounds in lupin (*Lupinus albus* L.) and their effects on host resistance against *Fusarium* wilt. *Eur. J. Plant Pathol.* **134**, 105–116.
  9. Acharya K., Chakraborty N., Dutta A.K., Sarkar S., Acharya R. (2011) Signaling role of nitric oxide in the induction of plant defense by exogenous application of abiotic inducers. *Arch. Phytopath. Plant Protect.* **44**, 1501–1511.
  10. Arlorio M., Ludwig A., Boller T., Bonfante P. (1992) Inhibition of fungal growth by plant chitinases and  $\beta$ -1,3-glucanases. *Protoplasma.* **171**, 34–43.
  11. Veluthakkal R., Dasgupta M.G. (2012) Isolation and characterization of pathogen defence-related class I chitinase from the actinorhizal tree *Casuarina equisetifolia*. *For. Path.* **42**, 467–480.
  12. Yeboah N.A., Arahira M., Nong V.H., Zhang D., Kado-kura K., Watanabe A., Fukazawa C. (1998) A class III acidic endochitinase is specifically expressed in the developing seeds of soybean (*Glycine max* [L.] Merr.). *Plant Mol. Biol.* **36**, 407–415.
  13. Gijzen M., Kuflu K., Qutob D., Chernys J.T. (2001) A class I chitinase from soybean seed coat. *J. Exp. Bot.* **52**, 2283–2289.
  14. Zhou G. (2017) Cloning and functional analysis of specific promoter p51408 in stigma exsertion of wild rice (*Oryza rufipogon* Griff.) from yuanjiang city. *Southwest China J. Agricult. Sci.* **11**, 2393–2397.
  15. Yu L., Ma T., Zhang Y., Hu Y., Yu K., Chen Y., Ma H., Zhao J. (2017) Identification and analysis of the stigma and embryo sac-preferential/specific genes in rice pistils. *BMC Plant Biology.* **17**, 60–81.
  16. Stein J.C., Dixit R., Nasrallah M.E., Nasrallah J.B. (1996) SRK, the stigma-specific S locus receptor kinase of Brassica, is targeted to the plasma membrane in transgenic tobacco. *Plant Cell.* **8**, 429–445.
  17. Hackett R.M., Cadwallader G., Franklin F.C. (1996) Functional analysis of a *Brassica oleracea* SLR1 gene promoter. *Plant Physiol.* **112**, 1601–1607.
  18. Alvim F.C., Carolino S.M., Cascardo J.C., Nunes C.C., Martinez C.A., Otoni W.C., Fontes E.P. (2001) Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress. *Plant Physiol.* **126**, 1042–1054.
  19. Higo K., Ugawa Y., Iwamoto M., Korenaga T. (1999) Plant cis-acting regulatory DNA element (PLACE) database. *Nucleic Acids Res.* **27**, 297–300.
  20. Jefferson R.A., Kavanagh T.A., Bevan M.W. (1987) GUS fusions:  $\beta$ -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.
  21. Xu X., Chen C., Fan B., Chen Z. (2006) Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell.* **18**, 1310–1326.
  22. Nishiuchi T., Shinshi H., Suzuki K. (2004) Rapid and transient activation of transcription of the *ERF3* gene by wounding in tobacco leaves: possible involvement of NtWRKYs and autorepression. *J Biol Chem.* **279**, 55355–55361.
  23. Chen C., Chen Z. (2002) Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced *Arabidopsis* transcription factor. *Plant Physiol.* **129**, 706–716.
  24. Zhang Z.L., Xie Z., Zou X., Casaretto J., Ho T.H., Shen Q.J. (2004) A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. *Plant Physiol.* **134**, 1500–1513.
  25. Xie Z., Zhang Z.L., Zou X., Huang J., Ruas P., Thompson D., Shen Q.J. (2005) Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol.* **137**, 176–189.
  26. Urao T., Yamaguchi-Shinozaki K., Urao S., Shinozaki K. (1993) An *Arabidopsis myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell.* **5**, 1529–1539.
  27. Solano R., Nieto C., Avila J., Canas L., Diaz I., Paz-Ares J. Dual. (1995) DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB.Ph3) from *Petunia* hybrid. *EMBO J.* **14**, 1773–1784.
  28. Agarwal M., Hao Y., Kapoor A., Dong C.H., Fujii H., Zheng X., Zhu J.K. (2006) A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J. Biol. Chem.* **281**, 37636–37645.
  29. Hartmann U., Sagasser M., Mehrtens F., Stracke R., Weisshaar B. (2005) Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. *Plant Mol. Biol.* **57**, 155–171.
  30. Park H.C., Kim M.L., Kang Y.H., Jeon J.M., Yoo J.H., Kim M.C., Park C.Y., Jeong J.C., Moon B.C., Lee J.H., Yoon H.W., Lee S.H., Chung W.S., Lim C.O., Lee S.Y., Hong J.C., Cho M.J. (2004) Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. *Plant Physiol.* **135**, 2150–2161.
  31. Luo H., Song F., Goodman R.M., Zheng Z. (2005) Up-regulation of *OsBIHDI1*, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses. *Plant Biol.* (Stuttg). **7**, 459–468.
  32. Nakashima K., Fujita Y., Katsura K., Maruyama K., Narusaka Y., Seki M., Shinozaki K., Yamaguchi-Shinozaki K. (2006) Transcriptional regulation of ABI3- and ABA-responsive genes including *RD29B* and *RD29A* in seeds, germinating embryos, and seedlings of *Arabidopsis*. *Plant Mol. Biol.* **60**, 51–68.
  33. Simpson S.D., Nakashima K., Narusaka Y., Seki M., Shinozaki K., Yamaguchi-Shinozaki K. (2003) Two different novel cis-acting elements of *erd1*, a *clpA* homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. *Plant J.* **33**, 259–270.
  34. Fehlberg V., Vieweg M.F., Dohmann E.M., Hohnjec N., Puhler A., Perlick A.M., Kuster H. (2005) The promoter of the leghaemoglobin gene *VfLb29*: functional analysis and identification of modules necessary for its acti-

- vation in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots. *J. Exp. Bot.* **56**, 799–806
35. Vieweg M.F., Frühling M., Quandt H.J., Heim U., Bäumlein H., Pühler A., Küster H., Perlick A.M. (2004) The promoter of the *Vicia faba* L. leghemoglobin gene *VfLb29* is specifically activated in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots from different legume and nonlegume plants. *MPMI.* **17**, 62–69.
  36. Kagaya Y., Ohmiya K., Hattori T. (1999) RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* **27**, 470–478.
  37. Gowik U., Burscheidt J., Akyildiz M., Schlue U., Koczor M., Streubel M., Westhoff P. (2004) *cis*-Regulatory elements for mesophyll-specific gene expression in the C<sub>4</sub> plant *Flaveria trinervia*, the promoter of the C<sub>4</sub> phosphoenolpyruvate carboxylase gene. *Plant Cell.* **16**, 1077–1090.
  38. Filichkin S.A., Leonard J.M., Monteros A., Liu P.P., Nonogaki H. (2004) A novel endo- $\beta$ -mannanase gene in tomato LeMAN5 is associated with anther and pollen development. *Plant Physiol.* **134**, 1080–1087.
  39. Yanagisawa S. (2000) Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *Plant J.* **21**, 281–288.
  40. Niki T., Mitsuhashi I., Seo S., Ohtsubo N., Ohashi Y. (1998) Antagonistic effect of salicylic acid and jasmonic acid on the expression of pathogenesis-related (PR) protein genes in wounded mature tobacco leaves. *Plant Cell Physiol.* **39**, 500–507.
  41. Chen L., Jiang B., Wu C., Sun S., Hou W., Han T. (2014) *GmPRP2* promoter drives root-preferential expression in transgenic *Arabidopsis* and soybean hairy roots. *BMC Plant Biol.* **14**, 245–257.
  42. Lalonde B.A., Nasrallah M.E., Dwyer K.G., Chen C.H., Barlow B., Nasrallah J. (1989) A highly conserved *Brassica* gene with homology to the S-locus specific glycoprotein structural gene. *Plant Cell.* **1**, 249–258.
  43. Ke L., Zheng T., Wu X., Chen J., Zhu S. (2007) Isolation and sequence analysis of the stigma-specific promoter from *Brassica napus*. *J. Agricult. Biotechnol.* **15**, 257–262.

## Identification of Stigma-Specific Expression Fragment in the Promoter of the Soybean Chitinase Class I Gene

C. M. Zhao<sup>1</sup>, H. Hou<sup>1</sup>, M. G. Xing<sup>2</sup>, and R.-G. Xue<sup>1</sup> \*

<sup>1</sup>College of Life Sciences, Qingdao Agricultural University, Qingdao, 266109 China

<sup>2</sup>Ubrigene (Jinan) Biosciences Co., Ltd, Jinan, 250000 China

\*e-mail: xuerengao@163.com

The expression level of heterologous genes in transgenic plants serves as an important indicator of gene efficiency. The small number of currently known effective promoters, limits the possibilities in fine-tuning the expression of transgenes. We cloned and characterized a tissue-specific promoter fragment of the soybean chitinase class I gene (*GmChi1*). The *GmChi1* promoter (*GmChi1P*) was cloned from Jungery soybean. The promoter sequence contains a number of putative *cis*-acting elements, including tissue-specific and stress-regulated motifs. By histochemical analysis, the *GmChi1P*-controlled  $\beta$ -glucuronidase (GUS) reporter enzyme activity was shown to be highest in the roots of transgenic *Nicotiana tabacum* cv. NC89 at the four-leaf sprout formation stage. Interestingly, the high GUS activity in transgenic tobacco roots was effectively suppressed by salicylic acid (SA) treatment. Deletion analysis of *GmChi1P* revealed that the sequences located between positions –719 and –382 contain key *cis*-elements responsible for the reporter *uidA* gene expression (encoding GUS) in leaves, roots, and wounds of *Nicotiana tabacum*. In addition, fluorometric analysis showed that the activity of the shortened *ChiP*(–1292) to *ChiP*(–719) promoters in the roots of transgenic tobacco was significantly suppressed by abscisic acid and completely suppressed by SA. The *ChiP*(–382) promoter was also found to be expressed exclusively in the stigma of transgenic tobacco flowers. Using the GUS reporter enzyme, no staining was detected in other flower organs in transgenic *Nicotiana tabacum*, including sepals, petals, anthers, filaments, and ovaries, or in any vegetative tissues. The results indicate that the promoter fragment *ChiP*(–382) can be used in tissue-specific regulation of gene expression and plant genetic engineering.

**Keywords:** *GmChi1*, soybean, promoter, transgenic plant, *Nicotiana tabacum*, stigma-specific promoter

ГЕНОМИКА.  
ТРАНСКРИПТОМИКА

УДК 575.852'113

НЕОБЫЧНАЯ ЗАВИСИМОСТЬ МЕЖДУ ЭКСПРЕССИЕЙ ГЕНОВ  
И ОТРИЦАТЕЛЬНЫМ ОТБОРОМ У ИНФУЗОРИЙ РОДА *Euplotes*<sup>1</sup>

© 2023 г. М. А. Молдован<sup>а</sup> \*, С. А. Гайдукова<sup>б</sup>

<sup>а</sup>Сколковский институт науки и технологий, Москва, 121205 Россия

<sup>б</sup>Факультет биоинженерии и биоинформатики, Московский государственный университет им. М.В. Ломоносова, Москва, 199911 Россия

\*e-mail: mika.moldovan@gmail.com

Поступила в редакцию 13.09.2022 г.

После доработки 13.09.2022 г.

Принята к публикации 13.09.2022 г.

У большинства исследованных организмов экспрессия генов связана с рядом эволюционных особенностей, относящихся к белоккодирующим последовательностям. В частности, экспрессия генов положительно коррелирует со средней интенсивностью отрицательного отбора и влияет на использование кодонов. Нами изучена связь между экспрессией генов и паттернами отбора у двух видов ресничных простейших рода *Euplotes*. Мы обнаружили, что на использование кодонов влияет экспрессия генов в этих организмах, что указывает на дополнительные эволюционные ограничения на возникновение мутаций в сильно экспрессируемых генах по сравнению с генами, экспрессируемыми с меньшей скоростью. В то же время, на уровне синонимичных и несинонимичных замен мы наблюдаем более сильное ограничение на гены, экспрессируемые с более низкой скоростью, по сравнению с генами с более высокой скоростью экспрессии. Наше исследование дополняет дискуссию об общих закономерностях эволюции и ставит новые вопросы о механизмах контроля экспрессии генов у инфузорий.

**Ключевые слова:** экспрессия генов, отрицательный отбор, инфузории, ресничные простейшие, *Euplotes*, эволюция белоккодирующих последовательностей

DOI: 10.31857/S002689842301010X, EDN: AWYZOB

REFERENCES

1. O'Connor L.J., Schoech A.P., Hormozdiari F., Gazal S., Patterson N., Price A.L. (2019) Extreme polygenicity of complex traits is explained by negative selection. *Am. J. Hum. Genet.* **105**, 456–476.
2. Sella G., Barton N.H. (2019) Thinking About the evolution of complex traits in the era of genome-wide association studies. *Annu. Rev. Genom. Hum. Genet.* **20**, 461–493.
3. Vitsios D., Dhindsa R.S., Middleton L., Gussow A.B., Petrovski S. (2021) Prioritizing non-coding regions based on human genomic constraint and sequence context with deep learning. *Nat. Commun.* **12**, 1504.
4. Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A., Bork P., Kondrashov A.S., Sunyaev S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods.* **7**, 248–249.
5. Crick F.H.C., Barnett L., Brenner S., Watts-Tobin R.J. (1961) General nature of the genetic code for proteins. *Nature.* **192**, 1227–1232.
6. Kimura M. (1983) The neutral theory of molecular evolution. *Cambridge University Press.*
7. Kimura M. (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature.* **267**, 275–276.
8. Kryazhimskiy S., Plotkin J.B. (2008) The population genetics of dN/dS. *PLoS Genet.* **4**, e1000304.
9. Yang Z., Bielawski J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503.
10. Zhang J., Yang J.R. (2015) Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* **16**, 409–420.
11. Jansen R. (2003) Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res.* **31**, 2242–2251.

<sup>1</sup> Статья представлена авторами на английском языке.  
Статья рекомендована к печати М.С. Гельфандом.

12. Sharp P.M., Li W.-H. (1987) The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295.
13. Seplyarskiy V.B., Soldatov R.A., Koch E., McGinty R.J., Goldmann J.M., Hernandez R.D., Barnes K., Correa A., Burchard E.G., Ellinor P.T., McGarvey S.T., Mitchell B.D., Vasan R.S., Redline S., Silverman E., Weiss S.T., Arnett D.K., Blangero J., Boerwinkle E., He J., Montgomery C., Rao D.C., Rotter J.I., Taylor K.D., Brody J.A., Chen Y.-D.I., de Las Fuentes L., Hwu C.-M., Rich S.S., Manichaikul A.W., Mychaleckyj J.C., Palmer N.D., Smith J.A., Kardia S.L.R., Peysner P.A., Bielak L.F., D O'Connor T., Emery L.S., NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Population Genetics Working Group, Gilissen C., Wong W.S.W., Kharchenko P.V., Sunyaev S.R. (2021) Population sequencing data reveal a compendium of mutational processes in the human germ line. *Science.* **373**, 1030–1035.
14. Pál C., Papp B., Hurst L.D. (2001) Highly expressed genes in yeast evolve slowly. *Genetics.* **158**, 927–931.
15. Rocha E.P.C., Danchin A. (2004) An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol. Biol. Evol.* **21**, 108–116.
16. Subramanian S., Kumar S. (2004) Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics.* **168**, 373–381.
17. Tuller T., Kupiec M., Ruppin E. (2008) Evolutionary rate and gene expression across different brain regions. *Genome Biol.* **9**, R142.
18. Marais G., Domazet-Lošo T., Tautz D., Charlesworth B. (2004) Correlated evolution of synonymous and nonsynonymous sites in drosophila. *J. Mol. Evol.* **59**, 771–779.
19. Hodgins K.A., Yeaman S., Nurkowski K.A., Rieseberg L.H., Aitken S.N. (2016) Expression divergence is correlated with sequence evolution but not positive selection in conifers. *Mol. Biol. Evol.* **33**, 1502–1516.
20. Popescu C.E., Borza T., Bielawski J.P., Lee R.W. (2006) Evolutionary rates and expression level in *Chlamydomonas*. *Genetics.* **172**, 1567–1576.
21. Gout J.-F., Kahn D., Duret L. (2010) The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* **6**, e1000944.
22. Cherry J.L. (2010) Expression level, evolutionary rate, and the cost of expression. *Genome Biol. Evol.* **2**, 757–769.
23. Wei C., Chen Y.-M., Chen Y., Qian W. (2021) The missing expression level—evolutionary rate anticorrelation in viruses does not support protein function as a main constraint on sequence evolution. *Genome Biol. Evol.* **13**(4), evab049.
24. Feyertag F., Berninsone P.M., Alvarez-Ponce D. (2016) Secreted proteins defy the expression level—evolutionary rate anticorrelation. *Mol. Biol. Evol.* **34**(3), 692–706.
25. Feyertag F., Berninsone P.M., Alvarez-Ponce D. (2019) N-glycoproteins exhibit a positive expression level—evolutionary rate correlation. *J. Evol. Biol.* **32**, 390–394.
26. Rzeszutek I., Maurer-Alcalá X.X., Nowacki M. (2020) Programmed genome rearrangements in ciliates. *Cell. Mol. Life Sci.* **77**, 4615–4629.
27. Prescott D.M. (1994) The DNA of ciliated protozoa. *Microbiol. Rev.* **58**, 233–267.
28. Mozzicafreddo M., Pucciarelli S., Swart E.C., Piersanti A., Emmerich C., Migliorelli G., Ballarini P., Miceli C. (2021) The macronuclear genome of the Antarctic psychrophilic marine ciliate *Euplotes focardii* reveals new insights on molecular cold adaptation. *Sci. Rep.* **11**(1), 18782.
29. Lobanov A.V., Heaphy S.M., Turanov A.A., Gerashchenko M.V., Pucciarelli S., Devaraj R.R., Xie F., Petyuk V.A., Smith R.D., Klobutcher L.A., Atkins J.F., Miceli C., Hatfield D.L., Baranov P.V., Gladyshev V.N. (2016) Position-dependent termination and widespread obligatory frameshifting in *Euplotes* translation. *Nat. Struct. Mol. Biol.* **24**, 61–68.
30. Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Muceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652.
31. Bray N.L., Pimentel H., Melsted P., Pachter L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* **34**, 525–527.
32. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
33. Abascal F., Zardoya R., Telford M.J. (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–13.
34. Nei M., Gojobori T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **5**, 418–426.
35. Nielsen R., Yang Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics.* **148**, 929–936.
36. Yang Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.
37. Dana A., Tuller T. (2014) The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* **42**, 9171–9181.
38. Zhou Z., Dang Y., Zhou M., Li L., Yu C., Fu J., Chen S., Liu Y. (2016) Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. USA.* **113**(41), E6117–E6125.
39. Akashi H., Gojobori T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA.* **99**, 3695–3700.
40. Lawrence J.R., Snyder R.A. (1998) Feeding behaviour and grazing impacts of a *Euplotes* sp. on attached bacteria. *Can. J. Microbiol.* **44**, 623–629.
41. Wiackowski K., Fyda J., Cieccko A. (2004) The behaviour of an omnivorous protozoan affects the extent of induced morphological defence in a protozoan prey. *Freshwater Biol.* **49**, 801–809.

## Unusual Dependence between Gene Expression and Negative Selection in *Euplotes*

M. A. Moldovan<sup>1, \*</sup> and S. A. Gaydukova<sup>2</sup>

<sup>1</sup>*Skolkovo Institute of Science and Technology, Moscow, 121205 Russia*

<sup>2</sup>*Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, 199911 Russia*

*\*e-mail: mika.moldovan@gmail.com*

In most of the studied organisms, gene expression is associated with a number of evolutionary features pertaining to the protein-coding sequences. In particular, gene expression positively correlates with the average intensity of negative selection and influences codon usage. Here, we study the connection between gene expression and selection patterns in two species of ciliate protists of the genus *Euplotes*. We find that codon usage is influenced by gene expression in these organisms, pointing at additional evolutionary constraints on mutations in heavily expressed genes relative to the genes expressed at lower rates. At the same time, at the level of synonymous vs. non-synonymous substitutions we observe a stronger constraint on the genes expressed at lower rates relative to those with higher rates of expression. Our study adds to the discussion about the general evolutionary patterns and opens new questions about the mechanisms of control of gene expression in ciliates.

**Keywords:** gene expression, negative selection, infusoria, ciliates, *Euplotes*, coding constraint

УДК 577.21

## АКТИВНОСТЬ АНТИРЕСТРИКЦИОННОГО БЕЛКА ArdB В ОТНОШЕНИИ ЭНДОНУКЛЕАЗЫ EcoAI

© 2023 г. А. А. Кудрявцева<sup>а, \*</sup>, В. А. Алексин<sup>а</sup>, М. Д. Лебедева<sup>б</sup>, Е. Csefalvay<sup>с</sup>,  
М. Weiserova<sup>д</sup>, И. В. Манухов<sup>а, е</sup>

<sup>а</sup>Московский физико-технический институт (Национальный исследовательский университет),  
Долгопрудный, Московская обл., 141707 Россия

<sup>б</sup>Автономная некоммерческая общеобразовательная организация “Физтех-лицей” им. П.Л. Капицы,  
Долгопрудный, Московская обл., 141707 Россия

<sup>с</sup>Laboratory of Structural Biology and Bioinformatics, Institute of Microbiology, Academy of Sciences of the Czech Republic,  
Zamek 136, Nove Hradu, 37333 Czech Republic

<sup>д</sup>Institute of Microbiology, Academy of Sciences of the Czech Republic,  
Videňská 1083, Praha 4, 14220 Czech Republic

<sup>е</sup>Московский государственный университет пищевых производств, Москва, 125080 Россия

\*e-mail: kudryavtseva@phystech.edu

Поступила в редакцию 30.07.2022 г.

После доработки 08.08.2022 г.

Принята к публикации 08.08.2022 г.

Белки семейства ArdB подавляют активность белков системы рестрикции–модификации I типа (RM-I). Механизм действия ArdB к настоящему моменту остается неизвестным; спектр ингибируемых ими мишеней исследован недостаточно. Нами показано, что присутствие гена *ardB* из конъюгативной плазмиды R64 подавляет активность EcoAI – представителя семейства IB системы рестрикции–модификации. Ввиду обнаруженного отсутствия специфичности ArdB к определенному семейству рестриктаз системы RM-I (ингибирует как IA-, так и IB-семейство) можно предполагать, что механизм антирестрикторной активности этого белка не зависит от последовательности как ДНК в сайте узнавания, так и структуры рестриктазы системы RM-I.

**Ключевые слова:** антирестрикция, RM-I, ArdB, EcoAI

**DOI:** 10.31857/S0026898423010056, **EDN:** AWVMVU

Для защиты генетического материала от гидролиза системами рестрикции–модификации вирусы, плазмиды и транспозоны используют ряд стратегий, одна из которых – продукция антирестрикторных белков. К ним относятся такие ингибиторы рестриктаз, как Ocr, ArdA и ArdB [1–3]. Некоторые антирестриктазы (Ocr и ArdA) относятся к ДНК-миметиками, т.е. структурно и электростатически имитируют В-форму ДНК и за счет этого функционируют как конкурентные ингибиторы ферментов рестрикции [4, 5]. Белок ArdB не относится к ДНК-миметикам, так как его структура не имеет ничего общего со структурой ДНК [6]. Ранее была высказана гипотеза, согласно которой механизм действия белков семейства ArdB состоит в том, что они неспецифически связываются с ДНК [7], поэтому рестриктаза I типа,

транслоцируя ДНК перед ее гидролизом, “сталкивается” с молекулой антирестрикторного белка. В результате R-субъединица рестриктазы образует комплекс с ArdB, что блокирует транслокацию и соответственно рестриксию [8]. Однако пока это только неподтвержденная гипотеза.

Антирестрикторная активность ArdB описана А. Белогуровым (А. Belogurov) и др. [9] для представителя семейства IA – EcoKI. А. D. Serfiotis-Mitsa и соавт. [10] показали, что ArdB ингибирует рестриктазы из семейств IB, IC и ID. Однако авторы не сравнивали активность ArdB против семейств IA и IB системы рестрикции–модификации I типа (RM-I), так как гены исследованных рестриктаз были локализованы в хромосомах бактерий неизогенных штаммов.

В представленной работе проведено сравнение антирестрикторной активности ArdB по отношению к ферментам EcoKI и EcoAI системы RM-I в штамме *Escherichia coli* TG1.

Сокращения: EOP (efficiency of plating) – эффективность посева; RM-I (restriction-modification system, type I) – система рестрикции–модификации I типа.

**Таблица 1.** Плазмиды, использованные в работе

Плаزمида	Описание	Источник
pAM35	В вектор pACYC184 вставлены гены субъединиц S, M и R рестриктазы EcoAI под контролем собственного промотора, Cm <sup>r</sup>	Эта работа
pACYCEcoKI	В вектор pACYC184 вставлены гены субъединиц S, M и R рестриктазы EcoKI под контролем собственного промотора, Cm <sup>r</sup>	Эта работа
pTZArdB	Вектор pTZ57R содержит ген <i>ardB</i> из плазмиды R64 под контролем промотора <i>Plac</i> , Ap <sup>r</sup>	[12]
pArdBRam	Вектор pRhamhIL-10LT содержит ген <i>ardB</i> из плазмиды R64 под контролем сильного, индуцируемого рамнозой промотора гена <i>rhaB</i> , Km <sup>r</sup>	Эта работа

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Бактериальные штаммы и условия культивирования.** Штамм *Escherichia coli* TG1 (*thi relA supE44 hsdR17 hsdM Δ(lac-proAB) [F' traD36 proAB lacIqZ ΔM15]*) был получен из Всероссийской коллекции промышленных микроорганизмов (ВКПМ) НИЦ “Курчатовский институт” (Москва, Россия). Клетки культивировали в среде LB (“Диа-М”, Россия) с добавлением антибиотиков: ампициллин (400 мкг/мл), канамицин (80 мкг/мл) или хлорамфеникол (20 мкг/мл). Агаризованная среда содержала 1.5% Бактоагара (“Диа-М”). Верхний слой содержал 0.7% Бактоагара и при индукции промотора гена *PrhaB* – 2 г/л рамнозы (“Диа-М”).

**Конструирование плазмид.** В качестве источника гена *ardB* использовали конъюгативную плазмиду R64, в качестве источника генов, определяющих экспрессию EcoKI, – хромосому штамма *E. coli* MG1655 (ВКПМ НИЦ “Курчатовский институт”). В качестве источника генов, определяющих экспрессию EcoAI, использовали хромосому *E. coli* NK354 [10].

ПЦР-амплификацию клонированных генов, эндонуклеазную обработку, электрофорез в агарозном геле, изолирование фрагментов ДНК проводили по общепринятой методике [11].

Наборы последовательно расположенных генов *hsd* (*hsdS*, *hsdR*, *hsdM*) с конститутивными промоторами, определяющими экспрессию систем рестрикции–модификации EcoKI и EcoAI, были клонированы в вектор pACYC184 по сайтам рестрикции BamHI, HindIII.

Полученные конструкции представлены в табл. 1.

Трансформацию бактерий плазмидами проводили кальциевым методом [11].

**Анализ ферментативной активности рестриктаз.** Эндонуклеазную активность ферментов опреде-

ляли с использованием фаговой методики [13] – по эффективности посева (ЕОР – efficiency of plating) фага λ.0 – и выражали как отношение числа бляшек на газоне штамма *E. coli* TG1 с тем или иным набором плазмид к их числу на газоне того же штамма, не несущего плазмид. Активность рестриктаз оценивали по снижению ЕОР по отношению к штамму TG1 без плазмид.

**Экспрессия и анализ антирестрикционной активности ArdB.** Для проверки антирестрикционной активности клетки инкубировали в среде LB в течение 4 ч с аэрацией 200 об/мин с последующим добавлением фага λ (любезно предоставлен проф. R. Devoret, Laboratoire d'Enzymologie, Centre National de la Recherche Scientifique, Франция) и высевом на агаризованную среду. Величину ЕОР оценивали как отношение числа полученных негативных колоний на газоне с исследуемыми клетками к числу негативных колоний, образованных *E. coli* TG1.

Для индукции экспрессии гена, определяющего синтез ArdB(R64) под контролем промотора *PrhaB*, в среду LB добавляли рамнозу (2 мг/мл) через 2 ч после засева (OD<sub>600</sub> ~ 0.1).

Уровень экспрессии ArdB контролировали методом электрофореза в 15%-ном ПААГ в денатурирующих условиях (SDS-PAGE) с окрашиванием Кумасси G-250 [14].

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для оценки антирестрикционной активности ArdB клетки *E. coli* TG1 трансформировали плазмидами, содержащими гены системы RM-I: pACYCEcoKI (семейство IA) или pAM35 (семейство IB), – и плазмидами с геном *ardB*: pTZArdB или pArdBRam. Их исследовали по отдельности или в комбинации (рестриктаза + антирестриктаза). В результате отслеживали влияние трех раз-

**Таблица 2.** Антирестрикционная активность ArdB

Штамм <i>E. coli</i>	ЕОР <sup>а</sup>	Штамм <i>E. coli</i>	ЕОР
	ЕсоАI (семейство IB)		ЕсоКI (семейство IA) <sup>б</sup>
TG1	1.00 ± 0.11	TG1	1.0 ± 0.11
TG1-pAM35	(5.5 ± 1.78) × 10 <sup>-2</sup>	TG1-pACYCEcoKI	(4.7 ± 2.12) × 10 <sup>-4</sup>
TG1-pAM35-pTZardB	(1.6 ± 0.23) × 10 <sup>-1</sup>	TG1-pACYCEcoKI-pTZardB	(4.7 ± 0.52) × 10 <sup>-2</sup>
TG1-pAM35-pArdBRam (без индукции)	(2.9 ± 0.15) × 10 <sup>-1</sup>	TG1-pACYCEcoKI-pArdBRam + индукция	(3.9 ± 0.89) × 10 <sup>-1</sup>
TG1-pAM35-ArdBRam + индукция	(5.7 ± 0.95) × 10 <sup>-1</sup>		

<sup>а</sup> Эффективность посева фага λ.0 определяли как отношение числа бляшек, образованных трансформированными клетками *E. coli*, к числу бляшек, образованных штаммом TG1; чем меньше это соотношение, тем выше эндонуклеазная активность исследуемого фермента. Представлены средние результаты трех независимых экспериментов.

<sup>б</sup> Положительный контроль.

ных концентраций ArdB в клетке: с плазмидой pTZardB (*ardB* под промотором гена *Plac*, но без индукции ИПТГ), pArdBRam с промотором гена *PrhaB*, который сильнее *Plac*, и pArdBRam с добавлением рамнозы для индукции экспрессии с промотора *PrhaB*.

На полученные трансформанты высевали фаг λ.0 для определения ЕОР, как описано выше. Результаты проведенных экспериментов представлены в табл. 2.

Как видно из данных, представленных в табл. 2, рестрикционная активность эндонуклеазы ЕсоАI, экспрессируемой трансформантом *E. coli* TG1-pAM35, значительно ниже, чем ЕсоКI в клетках *E. coli* TG1-pACYCEcoKI: снижение ЕОР на 1.5 и на 4 порядка соответственно. По всей видимости, это обусловлено разным числом сайтов узнавания этих рестриктаз в геноме фага λ (5 для ЕсоКI и 1 для ЕсоАI).

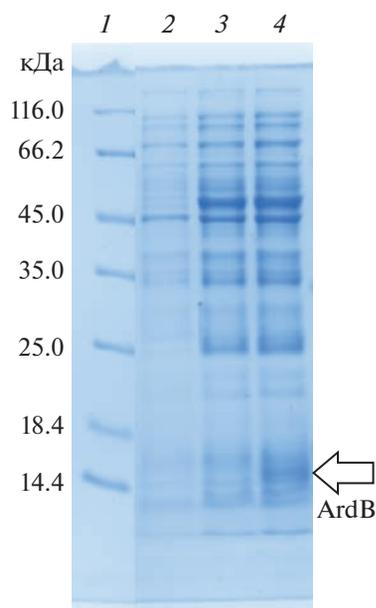
По результатам экспериментов можно заключить, что белок ArdB активен в отношении ЕсоАI – типowego представителя семейства IB системы RM-I. Даже экспрессия ArdB с относительно слабого лактозного промотора вызывала частичное восстановление уровня ЕОР. При экспрессии ArdB с более сильного промотора *PrhaB* выявлено повышение антирестрикционной активности ArdB, связанное с увеличением количества белка в клетке (рис. 1).

Полученные результаты можно рассматривать как важное подтверждение упомянутых выше результатов D. Serfiotis-Mitsa и др. [10] о способности ArgD ингибировать активность ЕсоАI.

Сравнение эффективности ингибирования активности двух различных RM-систем белком ArdB позволяет выдвинуть гипотезу о неспецифичности его работы, то есть способности инги-

бировать RM-I независимо от последовательности сайта узнавания.

На основании данных, представленных в табл. 2, можно говорить о проявлении антирестрикционной активности ArdB как против эндонуклеаз семейств IA, так и IB (ЕсоКI и ЕсоАI соответственно). Отметим, что для этих экспериментов ЕсоКI и ЕсоАI экспрессировали, используя изогенные штаммы *E. coli* TG1. Эффективность ингибирования ЕсоКI и ЕсоАI с помощью ArdB практически одинакова и для обеих эндонуклеаз составляет более 98% (табл. 2). Эндонуклеазы



**Рис. 1.** Электрофоретический анализ лизатов клеток *E. coli* TG1. Дорожки: 1 – маркер молекулярной массы белков PageRuler Unstained Protein Ladder (“Thermo Fisher Scientific”, США); 2 – *E. coli* TG1; 3 – *E. coli* TG1-pAM35-pArdBRam без индукции; 4 – *E. coli* TG1-pAM35-pArdBRam + рамноза.

EcoAI и EcoKI узнают разные последовательности ДНК [15, 16], поэтому одинаковую эффективность ингибирования их активности можно рассматривать как подтверждение высказанной гипотезы о неспецифичности ArdV относительно нуклеотидной последовательности ДНК, узнаваемой S-субъединицей комплекса RM-I. Скорее всего, ArdV распознает пространственную структуру ДНК в сайте расщепления рестриктазой или вблизи него.

Ранее в работах Г.Б. Завильгельского и И.В. Манухова [7, 8] был предположен механизм работы ArdV, объясняющий природу его антирестрикционной активности. В частности предполагалось, что ArdV ингибирует транслокацию ДНК R-субъединицами комплекса RM-I и, как следствие, независимо от работы S-субъединицы комплекса, а значит и от последовательности узнаваемого ею сайта. Теперь нами подтверждено одно из основных следствий этой гипотезы – универсальность механизма действия ArdV. Таким образом, вся накопленная к настоящему времени информация по особенностям структуры, каталитической активности и другим важным свойствам ArdV вполне согласуется с гипотезой, хотя ее доказательство нуждается в дальнейшей проверке.

Работа выполнена при финансовой поддержке Российского научного фонда (22-74-00027). Теоретическая оценка применимости результатов работы в биотехнологии, обзор литературы и редактирование текста публикации были выполнены Мануховым Ильей Владимировичем при финансировании Министерства науки и высшего образования Российской Федерации (проект 1022060200069-0-1.6.2; 1.6.4; 1.6.5; 1.6.10; 1.6.19 по теме “Разработка технологии рационального и высокопродуктивного использования агро- и биоресурсов, их эффективной переработки и получения безопасных и качественных источников пищевых и не пищевых продуктов”).

В работе не использовались животные в качестве объектов исследования.

Авторы заявляют об отсутствии конфликта интересов.

#### СПИСОК ЛИТЕРАТУРЫ

- Samson J.E., Magadán A.H., Moineau S., Sabri M. (2013) Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687.
- Tock M.R., Dryden D.T. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472.
- Liang W., Xie Y., Lu Y., Xiong W., Tang Y., Li G., Jiang X., Lu Y. (2017) Anti-restriction protein, KlcAHS, promotes dissemination of carbapenem resistance. *Front. Cell. Infect. Microbiol.* **7**, 150.
- McMahon S.A., Roberts G.A., Johnson K.A., Cooper L.P., Liu H., White J.H., Carter L.G., Sanghvi B., Oke M., Walkinshaw M.D., Blakely G.W., Naismith J.H., Dryden D.T. (2009) Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. *Nucleic Acids Res.* **37**, 4887–4897.
- Roberts G.A., Stephanou A.S., Kanwar N., Dawson A., Cooper L.P., Chen K., Nutley M., Cooper A., Blakely G.W., Dryden D.T. (2012) Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Res.* **40**, 8129–8143.
- Goryanin I.I., Kudryavtseva A.A., Balabanov V.P., Biryukova V.S., Manukhov I.V., Zavilgelsky G.B. (2018) Antirestriction activities of KlcA (RP4) and ArdB (R64) proteins. *FEMS Microbiol. Lett.* **365**, fny227.
- Кудрявцева А.А., Охрименко И.С., Дидина В.С., Завильгельский Г.Б., Манухов И.В. (2020) Антирестрикционный белок ArdB (R64) взаимодействует с ДНК. *Биохимия.* **85**, 369–377.
- Balabanov V.P., Kudryavtseva A.A., Melkina O.E., Pustovoi K.S., Khrulnova S.A., Zavilgelsky G.B. (2019) ArdB protective activity for unmodified  $\lambda$  phage against EcoKI restriction decreases in UV-treated *Escherichia coli*. *Curr. Microbiol.* **76**, 1374–1378.
- Belogurov A.A., Delver E.P., Rodzevich O.V. (1993) Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *J. Bacteriol.* **175**, 4843–4850.
- Serfiotis-Mitsa D., Herbert A.P., Roberts G.A., Soares D.C., White J.H., Blakely G.W., Uhrin D., Dryden D.T. (2010) The structure of the KlcA and ArdB proteins reveals a novel fold and antirestriction activity against type I DNA restriction systems *in vivo* but not *in vitro*. *Nucleic Acids Res.* **38**, 1723–1737.
- Маниатис Т., Фрич Э., Сэмбрук Дж. (1984) *Молекулярное клонирование*. Москва: Мир.
- Кудрявцева А.А., Осетрова М.С., Ливинюк В.Я., Манухов И.В., Завильгельский Г.Б. (2017) С-концевой остаток аспарагиновой кислоты (D141) необходим для антирестрикционной активности белка ArdB (R64). *Молекуляр. биология.* **51**(5), 831–835.
- Завильгельский Г.Б., Мершавка В.Ю., Юсифов Т.Н., Белогуров А.А. (1984) Ослабление рестрикции EcoK ДНК бактериофага в присутствии плазмиды рKM101 ard<sup>+</sup>. I. Общая характеристика и генетическая локализация. *Молекуляр. биология.* **18**(6), 1590–1596.
- Манухов И.В., Коноплева М.Н., Баженов С.В. (2016) *Практикум по генетической инженерии*. Черноголовка: Редакционно-издательский отдел ИПХФ РАН.
- Kan N.C., Lautenberger J.A., Edgell M.H., Hutchison C.A. (1979) The nucleotide sequence recognized by the *Escherichia coli* K12 restriction and modification enzymes. *J. Mol. Biol.* **130**(2), 191–209. [https://doi.org/10.1016/0022-2836\(79\)90426-1](https://doi.org/10.1016/0022-2836(79)90426-1)
- Kröger M., Hobom G. (1984) The nucleotide sequence recognized by the *Escherichia coli* A restriction and modification enzyme. *Nucleic Acids Res.* **12**(2), 887–899. <https://doi.org/10.1093/nar/12.2.887>

**Anti-Restriction Activity of ArdB Protein against EcoAI Endonuclease****A. A. Kudryavtseva<sup>1, \*</sup>, V. A. Alekhin<sup>2</sup>, M. D. Lebedeva<sup>2</sup>, E. Csefalvay<sup>3</sup>,  
M. Weiserova<sup>4</sup>, and I. V. Manukhov<sup>1, 5</sup>**<sup>1</sup>*Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, 141707 Russia*<sup>2</sup>*“Biophystech” LLC, Dolgoprudny, Moscow Region, 141701 Russia*<sup>3</sup>*Laboratory of Structural Biology and Bioinformatics, Institute of Microbiology, Academy of Sciences of the Czech Republic, Zamek 136, Nove Hrad, 37333 Czech Republic*<sup>4</sup>*Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, Praha 4, 14220 Czech Republic*<sup>5</sup>*Moscow State University of Food Production, Moscow, 125080 Russia**\*e-mail: kudryavtseva@phystech.edu*

ArdB proteins are known to inhibit the activity of type I restriction–modification (RM-I) system, in particular EcoKI (family IA). The mechanism of ArdB’s activity still remains unknown; the spectrum of targets inhibited by them has been poorly studied. In this work, it was shown that the presence of the *ardB* gene from R64 plasmid could suppress the activity of EcoAI endonuclease (IB family) in *Escherichia coli* TG1 cells. The absence of specificity of ArdB to a certain RM-I system (it inhibits both the IA- and IB-family), it can be assumed that the mechanism of the anti-restriction activity of this protein does not depend on both the sequence DNA at the recognition site and the structure of the restrictase of the RM-I systems.

**Keywords:** antirestriction, RM-I, ArdB, EcoAI

УДК 577.25,577.19:57.042.2

## МЕТИЛОФИОПОГОНАНОН А ИНГИБИРУЕТ LPS/АТР-ИНДУЦИРОВАННЫЙ ПИРОПТОЗ МАКРОФАГОВ ЧЕРЕЗ ПУТЬ ROS/NLRP3<sup>1</sup>

© 2023 г. Н. В. Zeng<sup>a</sup>, Л. Н. Zhang<sup>a</sup>, Д. Р. Yuan<sup>a</sup>, В. Wang<sup>b</sup>, Х. М. Su<sup>a</sup>, В. Х. Weng<sup>a</sup>, Р. Miao<sup>a</sup>,  
Ж. Ю. Xu<sup>a</sup>, Ж. Long<sup>a, \*</sup>, Я. Н. Song<sup>c, \*\*</sup>

<sup>a</sup>Jiangsu Key Laboratory for Pharmacology and Safety Evaluation of Chinese Materia Medica,  
School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

<sup>b</sup>Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

<sup>c</sup>Department of Cardiology, Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine,  
Nanjing, Jiangsu, China

\*e-mail: longjun@njucm.edu.cn

\*\*e-mail: sfy003@njucm.edu.cn

Поступила в редакцию 05.05.2022 г.

После доработки 18.07.2022 г.

Принята к публикации 27.07.2022 г.

Как побочный продукт митохондриального дыхания или метаболизма активные формы кислорода (ROS) могут действовать как сигнальные молекулы для активации инфламмосомы NLRP3 (NLR family pyrin domain containing 3), тем самым индуцируя иммунный ответ. NLRP3 инфламмосома действует как сенсор различных опасных сигналов и играет центральную роль в контроле возникновения пироптоза. Пироптоз макрофагов тесно связан с атеросклерозом, артритом, фиброзом легких и другими воспалительными заболеваниями. Метилофиопогонанон А (МО-А) — основной гомоизофлавоноид, входящий в состав китайского лекарственного растения *Ophiopogonis Radix* и проявляющий антиоксидантную активность. Однако неясно, может ли МО-А ослаблять пироптоз макрофагов путем ингибирования окислительного стресса. Нами показано, что МО-А усиливает активность супероксиддисмутазы и каталазы, ингибирует образование ROS, снижает активацию NLRP3-инфламмосомы и высвобождение лактатдегидрогеназы, а также подавляет пироптоз в макрофагах, индуцированный липополисахаридами (LPS) и аденозинтрифосфатом (АТР). Эти эффекты могут быть отменены активатором ROS — пероксидом водорода (H<sub>2</sub>O<sub>2</sub>). Таким образом, МО-А может ингибировать пироптоз макрофагов через путь ROS/NLRP3 и, следовательно, рассматриваться в качестве кандидата для разработки противовоспалительного лекарственного средства.

**Ключевые слова:** инфламмосома NLRP3, пироптоз, макрофаги, активные формы кислорода, метилофиопогонанон А

**DOI:** 10.31857/S0026898423010196, **EDN:** AWTQTV

### СПИСОК ЛИТЕРАТУРЫ

1. Cookson B.T., Brennan M.A. (2001) Pro-inflammatory programmed cell death. *Trends Microbiol.* **9**(3), 113–114.
2. Kovacs S.B., Miao E.A. (2017) Gasdermins: effectors of pyroptosis. *Trends Cell Biol.* **27**(9), 673–684.
3. Shi J.J., Gao W.Q., Shao F. (2017) Pyroptosis: gasdermin-mediated programmed necrotic cell death. *Trends Biochem. Sci.* **42**(4), 245–254.
4. Takeuchi O., Akira S. (2010) Pattern recognition receptors and inflammation. *Cell.* **140**(6), 805–820.
5. Franchi L., Muñoz-Planillo R., Núñez G. (2012) Sensing and reacting to microbes through the inflammasomes. *Nat. Immunol.* **13**(4), 325–332.
6. Liu X., Zhang Z.B., Ruan J.B., Pan Y., Magupalli V.G., Wu H., Lieberman J. (2016) Inflammasome-activated gasdermin D causes pyroptosis by forming membrane pores. *Nature.* **535**(7610), 153–158.
7. Miao E.A., Leaf I.A., Treuting P.M., Mao D.P., Dors M., Sarkar A., Warren S.E., Wewers M.D., Aderem A.

<sup>1</sup> Статья представлена авторами на английском языке.

- (2010) Caspase-1-induced pyroptosis is an innate immune effector mechanism against intracellular bacteria. *Nat. Immunol.* **11**(12), 1136–1142.
8. Lin X.B., Wang H., An X.F., Zhang J., Kuang J., Hou J., Yan M. (2021) Baecklein E suppressed NLRP3 inflammasome activation through inhibiting both the priming and assembly procedure: implications for gout therapy. *Phytomedicine.* **84**, 153521.
  9. Zou Y., Luo X., Feng Y., Fang S., Tian J., Yu B., Li J. (2021) Luteolin prevents THP-1 macrophage pyroptosis by suppressing ROS production via Nrf2 activation. *Chem. Biol. Interact.* **345**, 109573.
  10. Lin M.B., Sun W., Gong W., Zhou Z., Ding Y., Hou Q. (2015) Methylophiopogonanone A protects against cerebral ischemia/reperfusion injury and attenuates blood-brain barrier disruption *in vitro*. *PLoS One.* **10**(4), e124558.
  11. He F., Xu B.L., Chen C., Jia H.J., Wu J.X., Wang X.C., Sheng J.L., Huang L., Cheng J. (2016) Methylophiopogonanone A suppresses ischemia/reperfusion-induced myocardial apoptosis in mice via activating PI3K/Akt/eNOS signaling pathway. *Acta Pharmacol. Sin.* **37**(6), 763–771.
  12. Li Z., Wu Y.Y., Yu B.X. (2020) Methylophiopogonanone A, an *Ophiopogon* homoisoflavonoid, alleviates high-fat diet-induced hyperlipidemia: assessment of its potential mechanism. *Braz. J. Med. Biol. Res.* **53**(3), e9201.
  13. Qiu Z., He Y.H., Ming H., Lei S., Leng Y., Xia Z.Y. (2019) Lipopolysaccharide (LPS) aggravates high glucose- and hypoxia/reoxygenation-induced injury through activating ROS-dependent NLRP3 inflammasome-mediated pyroptosis in H9C2 cardiomyocytes. *J. Diabetes Res.* **2019**, 8151836.
  14. Wree A., Eguchi A., McGeough M.D., Pena C.A., Johnson C.D., Canbay A., Hoffman H.M., Feldstein A.E. (2014) NLRP3 inflammasome activation results in hepatocyte pyroptosis, liver inflammation, and fibrosis in mice. *Hepatology.* **59**(3), 898–910.
  15. Zheng M., Kanneganti T.D. (2020) The regulation of the ZBP1-NLRP3 inflammasome and its implications in pyroptosis, apoptosis, and necroptosis (PANoptosis). *Immunol. Rev.* **297**(1), 26–38.
  16. Xie Q., Shen W.W., Zhong J., Huang C., Zhang L., Li J. (2014) Lipopolysaccharide/adenosine triphosphate induces IL-1 $\beta$  and IL-18 secretion through the NLRP3 inflammasome in RAW264.7 murine macrophage cells. *Int. J. Mol. Med.* **34**(1), 341–349.
  17. Zhang W.B., Tao A., Lan T., Cepinskas G., Kao R., Martin C.M., Rui T. (2017) Carbon monoxide releasing molecule-3 improves myocardial function in mice with sepsis by inhibiting NLRP3 inflammasome activation in cardiac fibroblasts. *Basic Res. Cardiol.* **112**(2), 16.
  18. Fernandes-Alnemri T., Wu J., Yu J.W., Datta P., Miller B., Jankowski W., Rosenberg S., Zhang J., Alnemri E.S. (2007) The pyroptosome: a supramolecular assembly of ASC dimers mediating inflammatory cell death via caspase-1 activation. *Cell Death Differ.* **14**(9), 1590–1604.
  19. Liu X., Zhang X., Ding Y., Zhou W., Tao L., Lu P., Wang Y., Hu R. (2017) Nuclear factor E2-related factor-2 negatively regulates NLRP3 inflammasome activity by inhibiting reactive oxygen species-induced NLRP3 priming. *Antioxid. Redox Signal.* **26**(1), 28–43.
  20. Lei F.Y., Weckerle C., Heinrich M. (2021) Liriopogons (Genera *Ophiopogon* and *Liriope*, Asparagaceae): a critical review of the phytochemical and pharmacological research. *Front. Pharmacol.* **12**, 769929.
  21. Lin Y., Zhu D., Qi J., Qin M., Yu B. (2010) Characterization of homoisoflavonoids in different cultivation regions of *Ophiopogon japonicus* and related antioxidant activity. *J. Pharm. Biomed. Anal.* **52**(5), 757–762.
  22. Han X.J., Xu T.S., Fang Q.J., Zhang H., Yue L., Hu G., Sun L. (2021) Quercetin hinders microglial activation to alleviate neurotoxicity via the interplay between NLRP3 inflammasome and mitophagy. *Redox Biol.* **44**, 102010.
  23. Burdette B.E., Esparza A.N., Zhu H., Wang S. (2021) Gasdermin D in pyroptosis. *Acta Pharm. Sin. B.* **11**(9), 2768–2782.
  24. Orning P., Lien E., Fitzgerald K.A. (2019) Gasdermins and their role in immunity and inflammation. *J. Exp. Med.* **216**(11), 2453–2465.

## Methylophiopogonanone A Inhibits LPS/ATP-Induced Macrophage Pyroptosis via ROS/NLRP3 Pathway

H. B. Zeng<sup>1</sup>, L. H. Zhang<sup>1</sup>, D. P. Yuan<sup>1</sup>, W. Wang<sup>2</sup>, X. M. Su<sup>1</sup>, W. X. Weng<sup>1</sup>, R. Miao<sup>1</sup>, J. Y. Xu<sup>1</sup>, J. Long<sup>1</sup>, \*, and Y. H. Song<sup>3</sup>, \*\*

<sup>1</sup> Jiangsu Key Laboratory for Pharmacology and Safety Evaluation of Chinese Materia Medica, School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

<sup>2</sup> Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

<sup>3</sup> Department of Cardiology, Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

\*e-mail: longjun@njucm.edu.cn

\*\*e-mail: sfy003@njucm.edu.cn

As a byproduct of mitochondrial respiration or metabolism, reactive oxygen species (ROS) can act as a signaling molecule to activate NLR family pyrin domain containing 3 (NLRP3) inflammasome, thereby triggering immune response. NLRP3 inflammasome acts as a sensor of various danger signals and is central to

the control of pyroptosis occurrence. Macrophage pyroptosis is closely related to atherosclerosis, arthritis, pulmonary fibrosis and other inflammatory diseases. Methylophiopogonanone A (MO-A) is a main homoisoflavonoid in Chinese herb *Ophiopogonis Radix*, which has antioxidant effect. However, it is not clear whether MO-A can alleviate macrophage pyroptosis by inhibiting oxidative stress. Here we have shown that MO-A increases the activities of superoxide dismutase (SOD) and catalase (CAT), inhibits the production of ROS, reduces the activation of NLRP3 inflammasome and the release of lactate dehydrogenase (LDH), and inhibits pyroptosis in macrophages induced by lipopolysaccharides (LPS) and adenosine triphosphate (ATP). These effects can be reversed by the ROS promoter hydrogen peroxide ( $H_2O_2$ ). Therefore, MO-A can inhibit macrophage pyroptosis through the ROS/NLRP3 pathway and may be considered as a candidate drug for the treatment of inflammatory diseases.

**Keywords:** NLRP3 inflammasome, pyroptosis, macrophages, ROS, methylophiopogonanone A

УДК 577.218

## ИНСУЛЯТОРНЫЙ БЕЛОК CP190 РЕГУЛИРУЕТ ЭКСПРЕССИЮ ГЕНОВ ДИФФЕРЕНЦИРОВКИ СПЕРМАТОЦИТОВ В КЛЕТКАХ МУЖСКОГО ЗАРОДЫШЕВОГО ПУТИ *Drosophila melanogaster*<sup>1</sup>

© 2023 г. С. Е. Романов<sup>a, b</sup>, В. В. Шлома<sup>a</sup>, Д. Е. Коряков<sup>a</sup>, С. Н. Белякин<sup>a</sup>, П. П. Лактионов<sup>a, b, \*</sup><sup>a</sup>Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, 630090 Россия<sup>b</sup>Новосибирский национальный исследовательский государственный университет, Новосибирск, 630090 Россия

\*e-mail: laktionov@mcb.nsc.ru

Поступила в редакцию 23.06.2022 г.

После доработки 04.08.2022 г.

Принята к публикации 15.08.2022 г.

Белок CP190 — один из ключевых компонентов инсуляторных комплексов дрозофилы, изучение которого важно для понимания механизмов регуляции генов в процессе дифференцировки клеток. Однако мутанты по гену *Cp190* погибают, не доживая до взрослого состояния, что существенно усложняет изучение функций этого гена в клетках взрослого организма. Чтобы преодолеть эту проблему и исследовать регуляторные эффекты CP190 в развитии тканей имаго, мы разработали систему условного спасения мутантов *Cp190*. Спасаящая конструкция, содержащая кодирующую последовательность гена *Cp190*, направленно удаляется при помощи Cre/loxP-опосредованной рекомбинации только в половых клетках семенников, что позволяет изучать эффекты мутации в клетках мужского зародышевого пути. С использованием высокопроизводительного транскриптомного анализа установлено, что инактивации CP190 влияет на экспрессию генов в клетках зародышевого пути. Обнаружено, что мутация *Cp190* противоположным образом влияет на тканеспецифичные гены, экспрессия которых подавляется CP190, и гены домашнего хозяйства, активатором которых выступает CP190. Мутация *Cp190* также приводила к активации экспрессии субпопуляции генов дифференцировки сперматоцитов, которые регулируются транскрипционным комплексом tMAC. Полученные результаты указывают на то, что основной функцией CP190 в процессе сперматогенеза является координация взаимодействий между генами дифференцировки и специфичными для них транскрипционными активаторами.

**Ключевые слова:** CP190, инсуляторные белки, *Drosophila melanogaster*, РНК-seq**DOI:** 10.31857/S0026898423010147, **EDN:** AWRYNA

### ВВЕДЕНИЕ

Специализация клеток многоклеточного организма сопровождается активацией “генов развития” под воздействием энхансеров [1], а ключевую роль в регуляции промотор–энхансерных взаимодействий играют инсуляторные белки [2]. Они образуют физические контакты между специальными участками ДНК-инсуляторами, и тем самым обеспечивают пространственные взаимодействия промоторов и *цис*-регуляторных элементов. Сходным образом инсуляторные белки блокируют неспецифичные промотор–энхансерные взаимодействия и устанавливают барьеры между соседними доменами на хромосомах.

У *Drosophila melanogaster* обнаружена почти дюжина специализированных белков, которые в

различных комбинациях связываются с инсуляторами [3–6]. Один из ключевых — белок CP190, взаимодействует с большинством известных инсуляторов. Белок CP190 участвует в формировании физических ДНК-контактов, регуляции трехмерной архитектуры ядра, а также в блокировке промотор–энхансерных взаимодействий [7–9]. В ряде случаев CP190 выступает в роли транскрипционного фактора и привлекает к генам-мишеням регуляторные белки и эпигенетические модификаторы [6, 10–12].

Потеря функции CP190 сопровождается нарушениями транскрипции, которые, в частности, затрагивают гомеозисные гены, кодирующие ключевые регуляторы развития [8, 13, 14]. Нуль-мутации гена *Cp190* являются рецессивными летальными, приводящими к гибели *D. melanogaster* на стадии куколки [15, 16], что затрудняет исследование регуляторных эффектов CP190 в клетках

<sup>1</sup> Дополнительная информация для этой статьи доступна по DOI 10.31857/S0026898423010147.

имаго. По этой причине влияние этого белка на экспрессию генов и морфологию ядра изучали либо на культурах клеток, либо на ранних стадиях развития [8, 13], что накладывает ряд ограничений при интерпретации результатов.

В нашей работе в качестве модельной системы клеточной дифференцировки мы использовали сперматогенез *D. melanogaster*. В ходе дифференцировки клеток мужского зародышевого пути под воздействием специализированных транскрипционных факторов tMAC и tTAF активируются более 1000 специфических генов [17, 18]. При этом лишь небольшая доля геномных сайтов связывания tMAC и tTAF находится в непосредственной близости от генов-мишеней [18]. Это может предполагать регуляцию генов посредством сети пространственных взаимодействий между регуляторными элементами генома при участии инсуляторных белков. Чтобы изучить влияние CP190 на регуляцию генов в процессе клеточной дифференцировки, разработана генетическая система, которая позволяет избирательно исследовать проявления летальных мутаций *Cp190* в клетках зародышевого пути взрослых самцов. С использованием этого подхода мы оценили влияние белка CP190 на экспрессию генов в семенниках и проанализировали его влияние на регуляцию генов дифференцировки сперматоцитов.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Молекулярное клонирование и трансформация дрозофилы.** Спасаящая конструкция (GenBank ON783212), содержащая кодирующую последовательность гена *Cp190* под контролем промотора гена *Ubi63E*, создана на основе вектора pUAST-attB, как описано в разделе “Результаты исследования”. С использованием этой конструкции проведен phiC31-опосредованный трансгенез линии  $y^1w^{67};P\{y\{+t.7\} = CaryP\}attP40;M\{vas-int.B\}ZH-102D$ , несущей сайт интеграции *attP40* во второй хромосоме, в результате чего получили линию  $y^1w^{67};CP190-mCD8-GFP$  [19].

**Линии мух.** Все линии *D. melanogaster* содержали при температуре 25°C на стандартном корме. В работе использовали следующие линии:  $y^1w^{67};CP190-mCD8-GFP$ , несущую спасающую конструкцию;  $y^1w^{67};nanos-Cre$ , содержащую трансген *nanos-Cre*, интегрированный в сайт *attP40* [18, 20];  $y^1w^{67};+/+;Cp190^2,e^1/TM6$  [15] и  $y^1w^{67};+/+;Cp190^3,e^1/TM6$  [16], содержащие мутантные аллели *Cp190*. Стандартными генетическими скрещиваниями получены линии  $y^1w^{67};CDSCP190;Cp190^2,e^1/TM6$  и  $y^1w^{67};nanos-Cre;Cp190^3,e^1/TM6$ . Для проведения Cre/loxP-опосредованной рекомбинации скрещивали самок, несущих спасающую конструкцию, с самцами, несущими конструкцию *nanos-Cre*. При проведении

контрольных экспериментов в рамках транскрипционного анализа использовали линию  $y^1w^{67};\{Ubi>stop>mCD8-GFP\}$ , несущую трансген  $\{Ubi>stop>mCD8-GFP\}$  в сайте *attP40* [20].

**Выделение РНК и обратная транскрипция.** Для анализа экспрессии в целых семенниках диссектировали 25 пар семенников и выделяли РНК с использованием реагента TRIZOL (“Invitrogen”, США, 15596018). кДНК нарабатывали, используя 1 мкг суммарной РНК, обратную транскрипцию проводили с помощью Superscript II Reverse Transcriptase (“Invitrogen”, 18080093) с использованием dT-олигонуклеотидов в качестве затравки.

**Количественная ПЦР.** Количественную ПЦР проводили с использованием реакционной смеси БиоМастер HS-qPCR SYBR Blue (“Биолабмикс”, Россия, МНС030). Последовательности использованных праймеров приведены в табл. S1 (см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Romanov\\_rus.pdf](http://www.molecbio.ru/downloads/2023/1/supp_Romanov_rus.pdf)). Эффективность удаления спасающей кассеты оценивали при помощи метода  $\Delta\Delta Ct$ , экспрессию *CP190* измеряли методом стандартной кривой. Для нормирования использовали ген *Actin42A*. Стандартную кривую строили, используя серийные разведения образцов геномной ДНК личиночных имагинальных дисков самок  $y^1w^{67};CP190-mCD8-GFP$ .

**Микроскопия.** Для анализа флуоресценции GFP *in vivo* семенники переносили на предметное стекло в каплю PBS, накрывали покровным стеклом и сразу анализировали на флуоресцентном микроскопе. Семенники для иммуноокрашивания фиксировали стандартным методом (2%-ный раствор формальдегида в PBST) [21]. В работе использовали первичные антитела крысы к CP190 (любезно предоставлены А.К. Головинным) [22] и вторичные антитела (anti-Rat Cross-Adsorbed A568, “Thermo Fisher Scientific”, США, A-11077). Препараты анализировали на конфокальном микроскопе Carl Zeiss LSM 710 (Германия).

**Анализ транскриптома методом РНК-seq.** В эксперименте использовали семенники самцов генотипов  $y^1w^{67};CP190-mCD8-GFP/nanos-Cre$  и  $y^1w^{67};CP190-mCD8-GFP/nanos-Cre;Cp190^2/Cp190^3$ . Контролем служили семенники самцов  $y^1w^{67};Ubi>stop>mCD8-GFP/nanos-Cre$  с репортерной конструкцией *Ubi>stop>mCD8-GFP*, интегрированной в сайт *attP40* во второй хромосоме [20]. Суммарную РНК выделяли из 25 пар семенников с использованием реагента TRIZOL. Пробоподготовка библиотек проводили с использованием набора TruSeq RNA Sample Preparation v2 Kit (“Illumina”, США, RS-121-2002). Секвенирование библиотек проводили на платформе Illumina MiSeq в режиме парных прочтений 2 × 75 п.н. Получено от 3.7 до 6.5 млн парных прочтений на образцах.

Геномное выравнивание осуществляли с помощью инструмента HiSat2 [23] (сборка генома dm6, Release 6 plus ISO1 MT, Aug. 2014) со стандартными параметрами. Количество прочтений для каждого гена суммировано при помощи FeatureCounts [24] с использованием геномной аннотации UCSC refGene (<https://hgdownload.soe.ucsc.edu/goldenPath/dm6/bigZips/genes/dm6.refGene.gtf.gz>). Использовали от двух до трех биологических повторов эксперимента с каждым генотипом; профили экспрессии генов в образцах разных генотипов показали высокую корреляцию (рис. S1, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supr\\_Romanov\\_rus.pdf](http://www.molecbio.ru/downloads/2023/1/supr_Romanov_rus.pdf)). Дифференциальную экспрессию генов анализировали при помощи алгоритма DESeq2 [25] с параметрами “ashr” [26] и “ihv” [27]. Данные по экспрессии генов в целых семенниках мутантов *cap*, *comr*, *mir40*, *bam* получены ранее и доступны в базе данных Gene Expression Omnibus (GSE97182) [18]. Данные по экспрессии генов в тканях дрозофилы получены из базы modENCODE [28] и кластеризованы при помощи программы Cluster 3.0 [29]. Обогащение тканеспецифичными генами анализировали при помощи пакета TissueEnrich [30]. Данные секвенирования транскриптома единичных клеток семенников получены из базы Fly Cell Atlas и проанализированы с использованием программного пакета Seurat [31, 32]. Результаты секвенирования доступны в базе данных Gene Expression Omnibus (PRJNA847720).

**Связывание Cap, Comr и Mir40 с CP190-зависимыми генами.** Данные по связыванию белков Cap, Comr и Mir40 с хромосомами клеток мужского зародышевого пути получены ранее при помощи метода DamID-seq (GSE97182) [18]. Сайты связывания белков картировали с использованием набора программ DamID-Seq [33]. Все гены в геноме были разделены на заданное число фрагментов. Обогащение связывания белков в области генов исследуемой выборки оценивали, сопоставляя частоту встречаемости центров GATC-фрагментов, связанных белком (согласно данным DamID-seq), в каждом фрагменте гена, со средним значением по всем генам *D. melanogaster*. Статистическую значимость обогащения оценивали с помощью точного теста Фишера.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

### *Система условного спасения обеспечивает жизнеспособность мутантов Cp190*

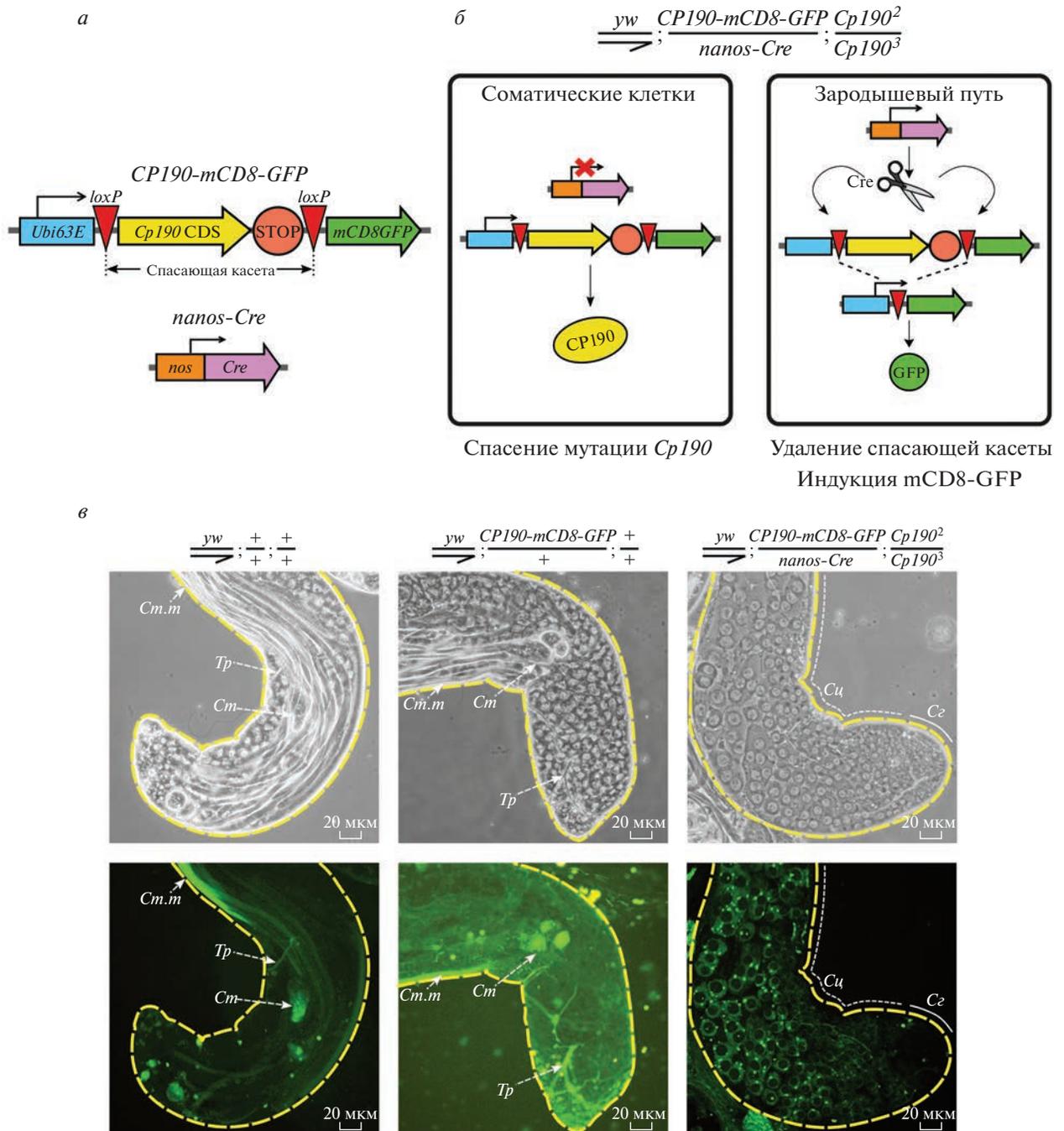
Мутанты по гену, кодирующему инсуляторный белок CP190, не доживают до взрослой стадии [15, 16], поэтому исследование функции этого гена в тканях имаго *D. melanogaster* сопряжено с экспериментальными трудностями. Чтобы преодолеть эту проблему и установить регуляторную функцию CP190 в клетках зародышевого пути

взрослых самцов, мы разработали генетическую систему для тканеспецифичного условного спасения мутации *Cp190* (рис. 1а).

Система состоит из двух генетических конструкций. Первая — спасающая конструкция *CP190-mCD8-GFP* — содержит кодирующую последовательность гена *Cp190* под контролем конститутивного промотора гена *Ubi63E*, который успешно использовали ранее для спасения летальной мутации *Cp190* [34, 35]. Вслед за кодирующей последовательностью *Cp190* помещен гибридный терминатор транскрипции HIS3-SV40 и кодирующая последовательность гена химерного репортерного белка mCD8-GFP [36, 37]. Кодирующая последовательность *Cp190* и терминатор транскрипции фланкированы loxP-сайтами узнавания рекомбиназы Cre и образуют спасающую генную кассету (рис. 1а). Вторая конструкция содержит ген рекомбиназы Cre под управлением промотора гена *nanos*, что обеспечивает наработку белка исключительно в клетках зародышевого пути (рис. 1а) [18, 20]. При наличии обеих конструкций в организме мутантов *Cp190* спасающая кассета будет специфически удалена из клеток зародышевого пути рекомбиназой Cre. В соматических клетках спасающая кассета не вырезается, обеспечивая жизнеспособность организма. При удалении кассеты в клетках зародышевого пути кодирующая последовательность гена химерного репортерного белка mCD8-GFP сближается с промотором гена *Ubi63E*, что позволяет получить клетки с дефицитом *Cp190*, помеченные GFP (рис. 1б).

С использованием системы phiC31-опосредованного трансгенеза генетическая конструкция *CP190-mCD8-GFP* была встроена в сайт attP40 во второй хромосоме [19]. Трансгенные мухи как с одной, так и с двумя копиями конструкции жизнеспособны и фертильны (данные не показаны). При этом одной копии оказалось достаточно для спасения летального фенотипа компаунд-гетерозигот *Cp190<sup>2</sup>/Cp190<sup>3</sup>*, у которых не образуется функциональный белок CP190 [15, 16]. Одновременно с этим экспрессия одной или двух копий спасающей конструкции не вызывала аномалий развития как в комбинации с *Cp190<sup>2</sup>/Cp190<sup>3</sup>*, так и на фоне дикого типа. Это свидетельствует об умеренном уровне эктопической экспрессии CP190, поскольку сильное повышение экспрессии этого гена вызывает серьезные дефекты развития [34].

В серии генетических скрещиваний получены самцы, несущие конструкции *CP190-mCD8-GFP* и *nanos-Cre* на фоне мутации *Cp190* (*y<sup>1w67</sup>; CP190-mCD8-GFP/nanos-Cre; Cp190<sup>2</sup>/Cp190<sup>3</sup>*). В клетках зародышевого пути таких самцов выявлена активация репортерного белка mCD8-GFP, начиная со стадии ранних сперматоцитов, что свидетельствовало об удалении спасающей генной кассеты



**Рис. 1.** Генетическая система для условного спасения мутантов с потерей функции *Cp190*. **а** – Схема спасающей конструкции *Cp190-mCD8-GFP* и конструкции *nanos-Cre* для экспрессии рекомбиназы Cre в клетках зародышевого пути дрозофилы. **б** – Принципиальная схема условного спасения мутантов *Cp190<sup>2</sup>/Cp190<sup>3</sup>* (см. пояснения в тексте). **в** – Снимки семенников самцов, не несущих спасающую конструкцию, с одной дозой спасающей конструкции и с условным спасением мутации *Cp190*. Режим фазового контраста (верхний ряд) и детекция флуоресценции GFP (нижний ряд). Желтым пунктиром обозначены границы семенников. Условные обозначения: *Cm.m* – хвосты сперматид; *Cm* – головы сперматид; *Tr* – дыхательные трубочки; *C2* – область семенника, содержащая сперматогонии (непрерывная линия); *C1* – область семенника, содержащая сперматциты (пунктирная линия).

(рис. 1в). Такие самцы обладали нормальной жизнеспособностью и не имели видимых нарушений сперматогенеза. GFP-негативные сперматозоиды и GFP-позитивные соматические клетки оболочки визуально не обнаружены, что указывает на высокую эффективность и специфичность удаления спасающей конструкции.

В контрольных образцах семенников  $y^1w^{67}$ , а также семенников самцов, несущих спасающую конструкцию в отсутствие Cre-рекомбиназы, наблюдалась только фоновая автофлуоресценция, характерная для дыхательных трубочек и фрагментов сперматид (рис. 1в). Это подтверждает надежное блокирование экспрессии репортерного белка во всех типах клеток семенника.

Дополнительная проверка эффективности удаления спасающей кассеты из генома клеток зародышевого пути при помощи количественной ПЦР показала, что спасающая конструкция сохраняется в 49.6% клеток целого семенника (рис. 2а). Это значение соответствует доле соматических клеток в семеннике (50.4%), оцененной исходя из данных секвенирования транскриптома единичных клеток семенника [31]. Этот результат свидетельствует о том, что спасающая кассета удаляется в подавляющем большинстве половых клеток.

*Система условного спасения значительно снижает уровень эктопической экспрессии Cp190 в клетках мужского зародышевого пути имаго*

Мы проанализировали уровни экспрессии мРНК эндогенного и трансгенного гена *Cp190* методом количественной ПЦР с праймерами на кодирующую область гена *Cp190* и 3'-нетранслируемую область эндогенного *Cp190* (рис. 2б). Первая пара праймеров позволяет измерить уровень экспрессии и эндогенного, и трансгенного *Cp190*, а вторая пара – только эндогенной мРНК *Cp190*. Установлено, что удаление спасающей кассеты под воздействием *nanos-Cre* приводило к снижению экспрессии кодирующей последовательности *Cp190* в целых семенниках в 3.8 раза (рис. 2в). Вместе с тем, уровень экспрессии кодирующей последовательности *Cp190* оставался в 2.4 раза выше по сравнению с контролем дикого типа, что, по всей видимости, можно объяснить повышенным (в сравнении с эндогенным) уровнем экспрессии трансгенного *Cp190* в соматических клетках, а также сохранением спасающей конструкции в единичных сперматозоидах (рис. 2г).

Получение мутантных по *Cp190* клеток зародышевого пути подтверждено иммуноокрашиванием (рис. 2з). В семенниках  $y^1w^{67};CP190-mCD8-GFP/nanos-Cre;Cp190^2/Cp190^3$  белок CP190 обнаружен в ядрах клеток на апикальном конце семенника, где располагаются недифференцированные сперматогонии, а также в ядрах соматических кле-

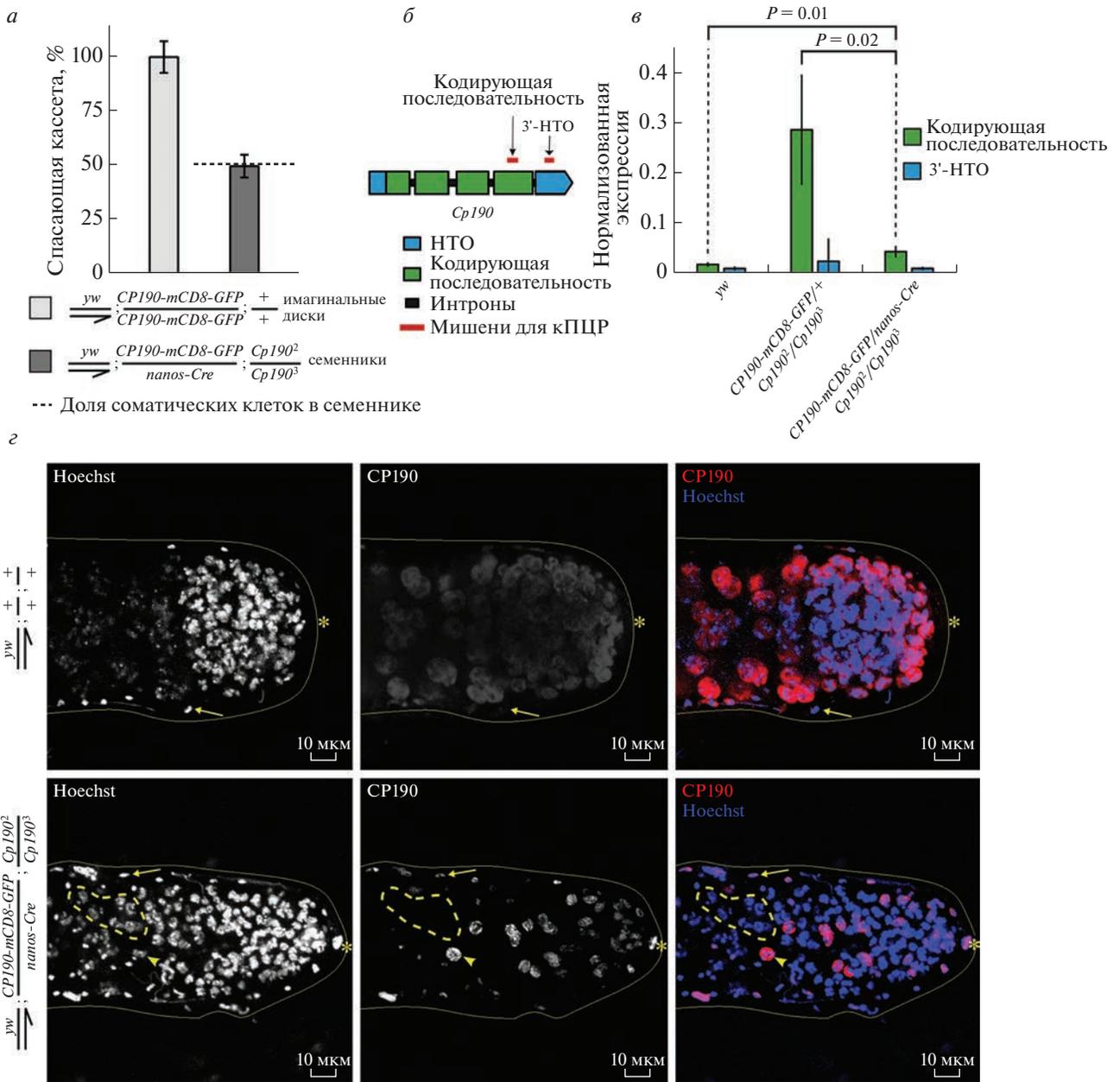
ток оболочки. Вместе с тем, экспрессия CP190 наблюдалась лишь в единичных ядрах сперматозоидов. В семенниках дикого типа стабильный уровень экспрессии CP190 наблюдался как в соматических, так и в половых клетках. Таким образом, разработанная система обеспечивает спасение мутантного фенотипа *Cp190* в соматических клетках, а также достаточно эффективное удаление спасающей кассеты в клетках мужского зародышевого пути, что на фоне мутации позволяет изолированно воспроизвести мутантный фенотип *Cp190* в половых клетках взрослых самцов.

*Мутация Cp190 влияет на экспрессию генов дифференцировки в клетках мужского зародышевого пути*

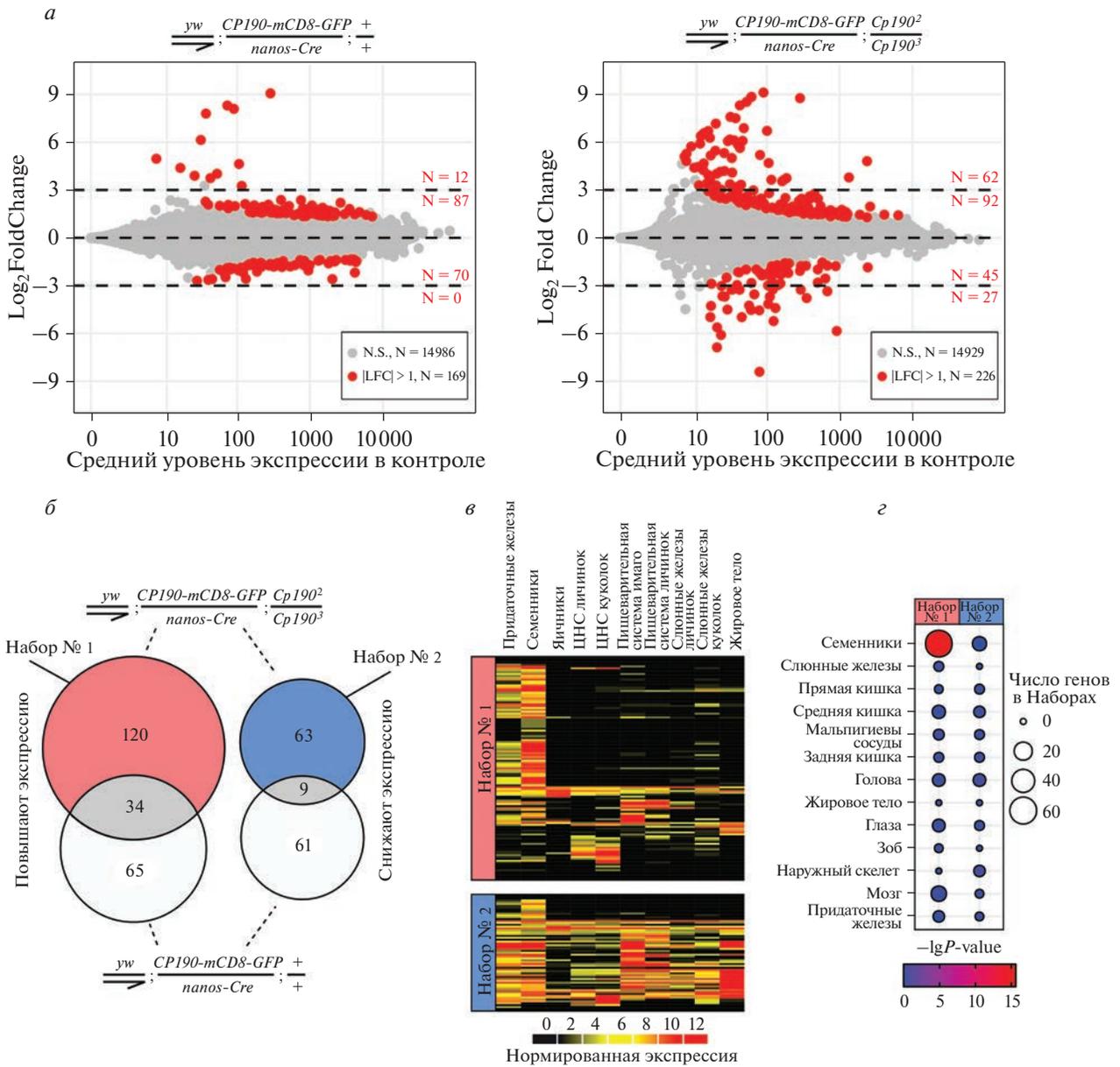
Влияние CP190 на экспрессию генов в семенниках имаго оценивали с помощью высокопроизводительного секвенирования транскриптома (РНК-seq) и последующего анализа дифференциальной экспрессии (табл. S2, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Romanov\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Romanov_rus.zip)). В экспериментах использовали трехнедельных самцов, так как в этом возрасте семенники  $y^1w^{67};CP190-mCD8-GFP/nanos-Cre$  и  $y^1w^{67};CP190-mCD8-GFP/nanos-Cre;Cp190^2/Cp190^3$  имеют эквивалентный уровень экспрессии *Cp190* (рис. S2, см. Дополнительные материалы на сайте [http://www.molecbio.ru/downloads/2023/1/supp\\_Romanov\\_rus.zip](http://www.molecbio.ru/downloads/2023/1/supp_Romanov_rus.zip)). Важно отметить, что активность спасающей кассеты контролируется промотором гена *Ubi63E*, из-за чего уровень транскрипции *Cp190* превышает эндогенный (рис. 2в), что может влиять на экспрессию генов. Чтобы аккуратно учесть этот эффект в нашем эксперименте, мы оценили влияние повышенной экспрессии *Cp190* в соматических клетках на транскриптом целых семенников.

С этой целью мы измерили экспрессию генов в семенниках самцов  $y^1w^{67};CP190-mCD8-GFP/nanos-Cre;+/+$ , в которых экспрессируется эндогенный функциональный ген *Cp190*, а спасающая кассета активна только в соматических клетках. Для сравнения использовали экспрессию генов в семенниках самцов контрольной линии. Присутствие спасающей конструкции привело к умеренным изменениям транскрипции: экспрессия 12 генов изменялась более чем в 8 раз (рис. 3а). Анализ семенников с условным спасением мутации *Cp190* в соматических клетках ( $y^1w^{67};CP190-mCD8-GFP/nanos-cre;Cp190^2/Cp190^3$ ) выявил 89 генов, чья экспрессия изменялась более чем в 8 раз в сравнении с контролем (рис. 3а).

Кроме того, из 226 генов, дифференциально экспрессирующихся на фоне условного спасения мутации *Cp190* (экспрессия 154 генов повыси-



**Рис. 2.** Cre-опосредованная рекомбинация приводит к удалению спасающей кассеты и утрате CP190 в клетках мужского зародышевого пути. **а** – Количество копий спасающей кассеты в семенниках с условным спасением сравнивали с количеством в имажинальных дисках личинок, несущих двойную дозу конструкции *CP190-mCD8-GFP*. Число копий спасающей кассеты нормировали на дозу конструкции. Пунктирной линией обозначена доля соматических клеток в семенниках дрозофилы по данным секвенирования транскриптома единичных клеток [31]. **б** – Расположение мишеней для ОТ-кПЦР в гене *Cp190*. Соответствующие праймеры использовали для измерения экспрессии эндогенного и трансгенного *Cp190*. **в** – Нормированная экспрессия *Cp190* в семенниках трехдневных имаго с конструкциями *CP190-mCD8-GFP* и *nanos-Cre*. Экспрессию измеряли при помощи праймеров к локусам, обозначенным на диаграмме **б** – Экспрессию нормировали на активность гена домашнего хозяйства *Actin42A*. Величина *P*-value получена с помощью теста Стьюдента с двусторонним распределением. **г** – Конфокальные изображения семенников с иммуноокрашиванием антителами анти-CP190. В верхнем ряду – семенник дикого типа (*yw*). В нижнем ряду – семенник с условным спасением. ДНК окрашивали красителем Hoechst. Непрерывной желтой линией подчеркнуты границы семенников. Звездочкой отмечен апикальный конец семенника. Длинной стрелкой показаны соматические клетки оболочки. Пунктирной линией показано скопление CP190-негативных сперматоцитов. Короткой стрелкой показан CP190-позитивный сперматоцит. Масштаб 10 мкм.



**Рис. 3.** Истощение CP190 в клетках мужского зародышевого пути сопровождается активацией семенник-специфичных генов и инактивацией генов домашнего хозяйства. *а* – Дифференциальная экспрессия генов в семенниках с повышенной экспрессией CP190 в соматических клетках (диаграмма слева) или в семенниках с условным спасением (диаграмма справа) по результатам РНК-seq. По вертикальной оси отложен логарифмированный диапазон изменения экспрессии ( $\text{Log}_2\text{FoldChange}$ , LFC) трансгенных семенников по сравнению с контролем. По горизонтальной оси отложена нормированная средняя экспрессия в семенниках контрольной линии. Гены со статистически значимым изменением экспрессии ( $|\text{LFC}| > 1$ ,  $P < 0.05$ ) помечены красным цветом. Значения LFC генов с незначимым изменением экспрессии (серые точки, N.S.) были для наглядности усечены при помощи эмпирического байесовского подхода. Пунктирными линиями указаны значения  $\text{LFC} \pm 3$  и 0. Красным показано число красных точек со значением LFC в интервале между ближайшими пунктирными линиями. *б* – Пересечение между множествами генов, экспрессия которых значимо повысилась и понизилась в семенниках с условным спасением или в семенниках с повышенной экспрессией CP190 в соматических клетках (красные точки на диаграмме *а*). Гены, экспрессия которых изменилась только на фоне условного спасения, обозначены как Набор № 1 и Набор № 2. *в* – Нормированная экспрессия генов из Наборов № 1 и № 2 в разных тканях дрозофилы по данным modENCODE [28]. *г* – Анализ тканеспецифичной экспрессии CP190-зависимых генов в клетках мужского зародышевого пути.

лась, 72 — понизилась), транскрипция 43 изменялась и на фоне повышенной экспрессии *Cr190* (у 34 генов повысилась, у 9 снизилась) (рис. 3б). Эти 43 гена были исключены из дальнейшего анализа, направленного на выделение специфического эффекта мутации в гене *Cr190* на экспрессию генов в клетках зародышевого пути.

После этой поправки обнаружили 120 генов, чья экспрессия усиливалась на фоне мутации *Cr190* (Набор № 1), и 63 гена, уровень экспрессии которых снижался (Набор № 2) (рис. 3б). Чтобы охарактеризовать эти выборки, мы проанализировали профили экспрессии входящих в них генов в разных тканях с использованием транскриптомных данных базы modENCODE [28]. Оказалось, что в Наборе №1 можно выделить три основных кластера, представленных генами, активными в мужской репродуктивной, нервной и пищеварительной системах (рис. 3в). Гены же Набора № 2 экспрессируются повсеместно и лишь 14 из них можно причислить к семенник-специфичным. Анализ паттернов экспрессии генов в различных тканях дрозофилы с использованием программы TissueEnrich [30] подтвердил преобладание семенник-специфичных генов в Наборе № 1, но не в Наборе № 2 (рис. 3г). Таким образом, белок *Cr190* может участвовать как в поддержании экспрессии повсеместно активных генов в семенниках, так и в репрессии (или модуляции) активности тканеспецифичных генов, в том числе и генов, участвующих в сперматогенезе.

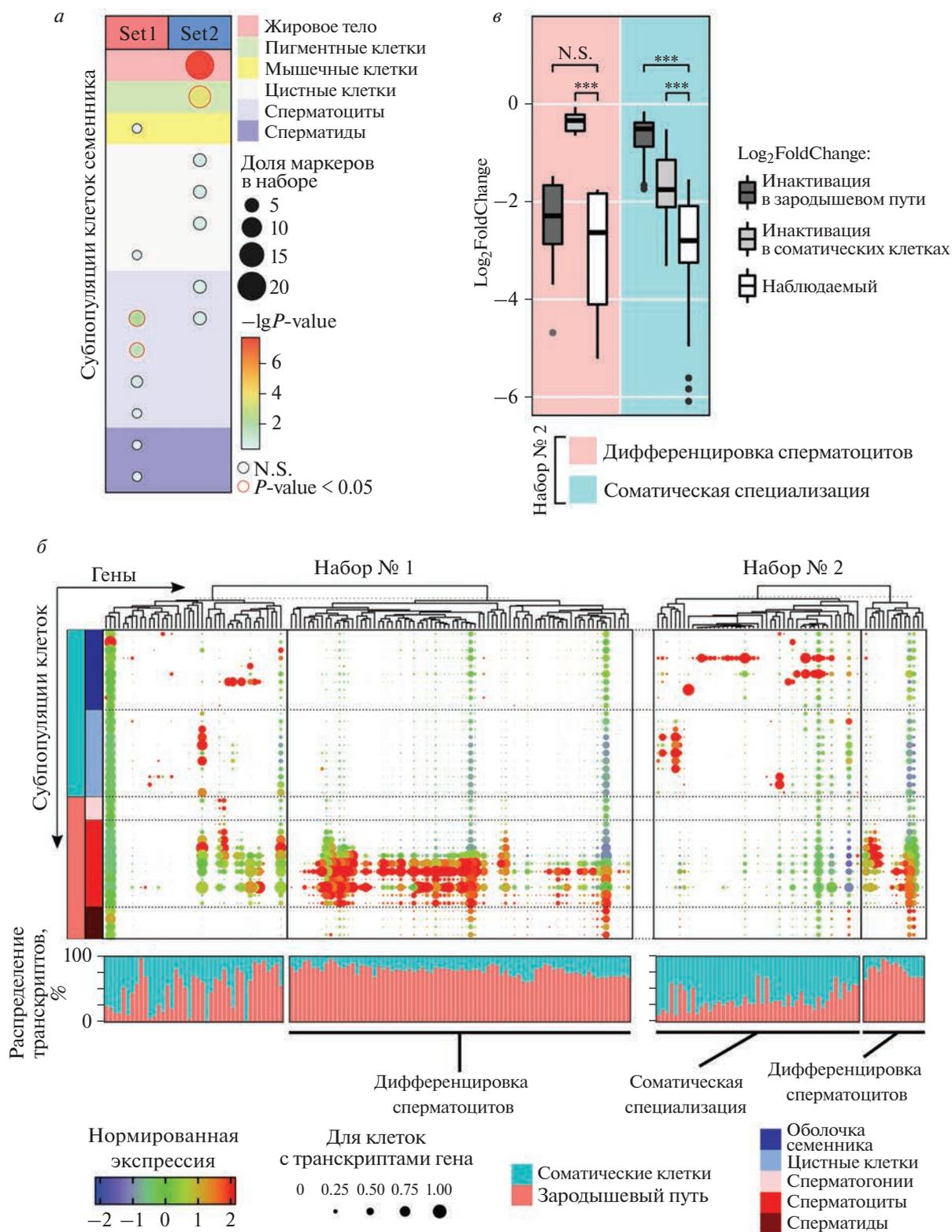
Вместе с тем, необходимо учитывать, что клеточный состав семенника достаточно разнообразен, он представлен несколькими типами соматических клеток, а также клетками зародышевого пути, находящимися на разных стадиях дифференцировки. Различные клеточные типы имеют характерные отличия в профиле экспрессии, что подтверждается данными секвенирования тран-

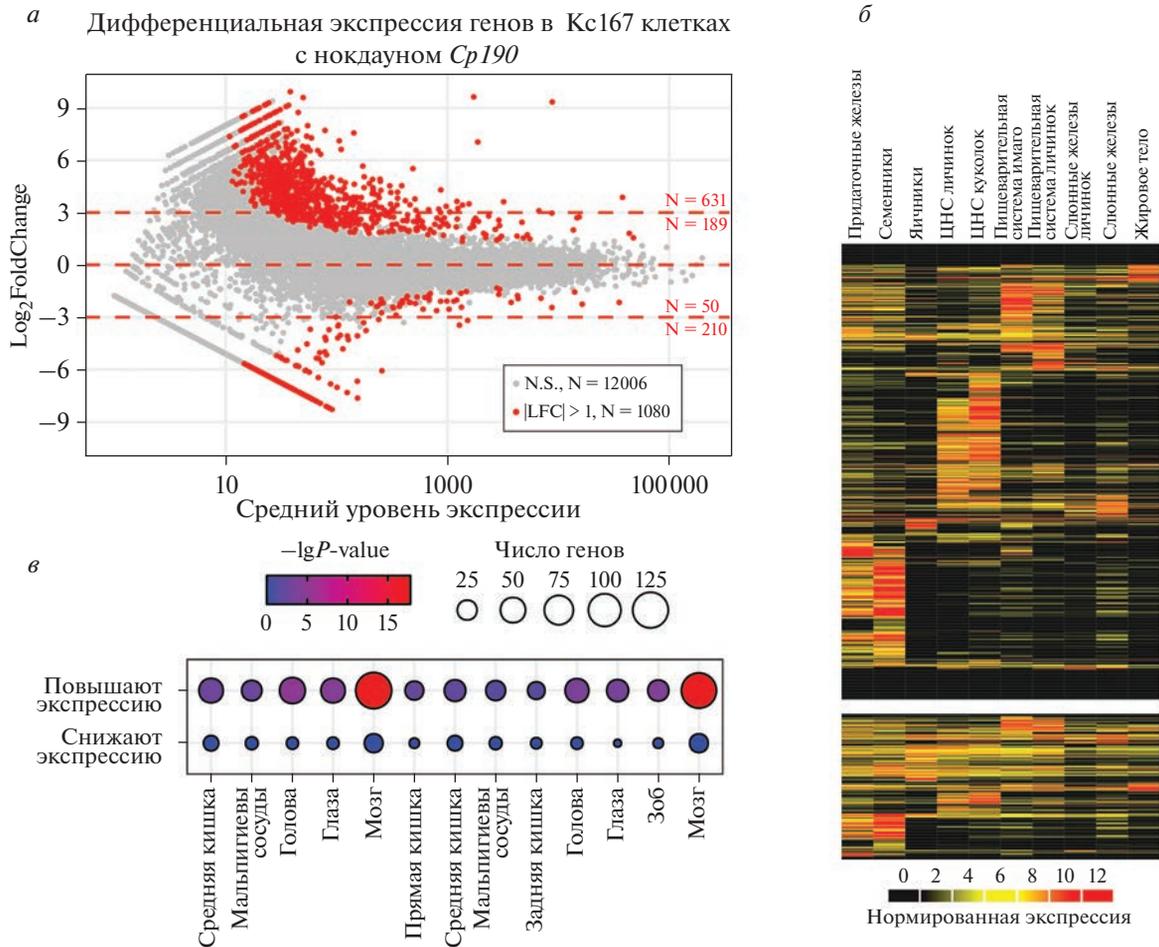
скриптома единичных клеток семенников. Охарактеризованные профили экспрессии отдельных клеточных популяций позволяют выделить специфичные гены-маркеры, а также установить, в каких клетках активен тот или иной ген в норме.

Чтобы определить, в каких конкретно клетках семенника в норме активны *Cr190*-зависимые гены, мы использовали гены-маркеры различных клеточных популяций семенников дрозофилы из базы данных DRscDB [31, 38]. В Наборе № 1 обнаружены лишь единичные гены-маркеры соматических клеток, тогда как маркеры сперматоцитов в сумме составляли 10% (рис. 4а). В Наборе № 2 маркерами жировых и пигментных клеток были 22% генов, маркеры же клеток зародышевого пути были единичными (рис. 4а). Кластеризация выборки генов Набора № 1 по уровню экспрессии в разных клеточных популяциях семенника с использованием базы данных транскриптома единичных клеток Fly Cell Atlas [31] позволила установить, что транскрипция 66% генов характерна для сперматоцитов (рис. 4б). Не менее 60% экспрессии таких генов обеспечивается клетками зародышевого пути, что позволяет отнести большинство генов Набора № 1 к генам дифференцировки сперматоцитов (рис. 4б). Напротив, лишь пятую часть генов Набора № 2 можно идентифицировать как гены дифференцировки сперматоцитов, а их общее количество в 5.7 раза меньше, чем в Наборе № 1. Все вместе это позволяет предположить, что *Cr190* оказывает преимущественно репрессорное воздействие по отношению к генам дифференцировки сперматоцитов.

Основная часть генов из Набора № 2 в норме активна преимущественно в соматических клетках семенников (рис. 4б). Вместе с тем, оставался открытым вопрос, изменяется ли уровень экспрессии этих генов в клетках зародышевого пути на фоне истощения *Cr190*. С использованием

**Рис. 4.** Утрата *Cr190* в клетках мужского зародышевого пути приводит к нарушению активности генов дифференцировки сперматоцитов. *a* — Обогащение генов-маркеров клеточных типов в Наборе № 1 и № 2 по данным DRscDB [38]. Клеточные типы аннотированы в эксперименте по секвенированию транскриптома единичных клеток [31], в некоторых случаях они включают несколько субпопуляций с собственным набором генов-маркеров. Принадлежность субпопуляций клеточным типам обозначена цветными прямоугольниками. Размер кругов характеризует представленность маркеров во всем наборе. Статистическая значимость обогащения вычислена в точном тесте Фишера. Маркеры со значимым обогащением обведены красным. *b* — Экспрессия генов из Наборов № 1 и № 2 в клеточных популяциях семенника по данным Fly Cell Atlas [31]. Принадлежность популяций клеточным типам и анатомическим структурам в семеннике закодирована цветными прямоугольниками. Цвет точек характеризует уровень экспрессии гена в сравнении со средней экспрессией в клетках семенника. Размер точек описывает долю клеток в популяции, в которых обнаружены транскрипты данного гена. На столбчатой диаграмме показан вклад клеток зародышевого пути и соматических клеток в общую транскрипцию каждого гена в целом семеннике. При помощи иерархической кластеризации по уровням экспрессии в популяциях клеток гены в каждом наборе разделены на два кластера. В Наборе № 1 отчетливо выделяются гены с высокой активностью в сперматоцитах (гены дифференцировки сперматоцитов). В Наборе № 2 кластеры можно охарактеризовать как гены соматической специализации и гены дифференцировки сперматоцитов. *c* — Сравнение оценок изменения экспрессии в двух кластерах из Набора 2. Оценки диапазона изменений экспрессии ( $\text{Log}_2\text{FoldChange}$ ) получены отдельно для генов дифференцировки сперматоцитов (красный цвет) и генов соматической специализации (синий цвет) из предположения, что в семенниках с условным спасением инактивация произошла только в клетках зародышевого пути (темно-серые прямоугольники) или только в соматических клетках (светло-серые прямоугольники). Наблюдаемые значения  $\text{Log}_2\text{FoldChange}$  показаны белыми прямоугольниками. Условные обозначения: \*\*\* $P < 5 \times 10^{-8}$  в тесте Вилкоксона, N.S. —  $P > 0.05$ .



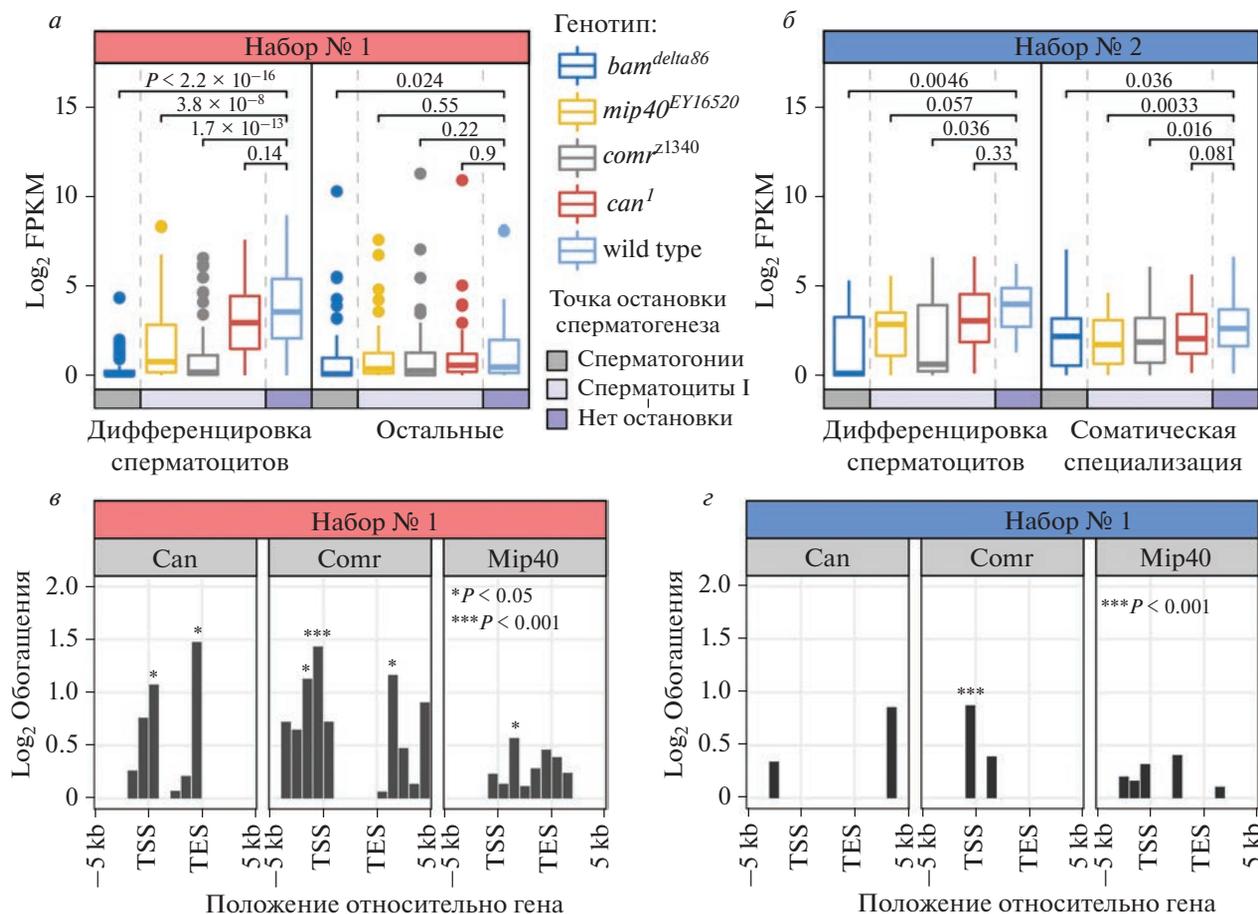


**Рис. 5.** CP190 подавляет активность тканеспецифичных генов и способствует активации генов домашнего хозяйства в клетках Кс167. *а* – Эффект инактивации *Cp190* в клетках Кс167 согласно данным эксперимента [13]. Дифференциальную экспрессию анализировали таким же способом, как на рис. 3*а*. *б* – Экспрессия CP190-зависимых генов (красные точки на диаграмме *а*) в тканях дрозофилы согласно данным modENCODE [28]. *в* – Анализ тканеспецифичной экспрессии CP190-зависимых генов клеток Кс167.

данных о транскриптомах единичных клеток [31], позволяющих оценить вклад отдельной клеточной популяции семенников в уровень экспрессии генов, мы смоделировали два профиля экспрессии, соответствующих образцам целых семенников, в одном профиле исследуемые гены были инактивированы в клетках зародышевого пути, а в другом – в соматических клетках. На следующем этапе мы сопоставили реальные изменения экспрессии генов Набора № 2 с моделями инактивации генов в отдельных клеточных популяциях. Инактивация генов, активных преимущественно в соматических клетках, только в соматических или только в половых клетках не позволяет смоделировать такое же сильное падение транскрипции, какое мы наблюдали в семенниках с условным спасением ( $P < 5 \times 10^{-8}$  в тесте Вилкоксона) (рис. 4*в*). Напротив, оценка генов дифференцировки сперматоцитов, полученная на основе гипотезы об инактивации в зародышевом пути, не

отличается от реальных значений ( $P = 0.23$  в тесте Вилкоксона) (рис. 4*в*). Такое наблюдение может указывать на то, что падение транскрипции генов Набора № 2, характерных для соматических клеток, происходит как в половых, так и в соматических клетках.

Интересно отметить, что схожие регуляторные эффекты CP190 обнаружены и в культуре соматических клеток дрозофилы. С использованием полученных ранее данных транскриптомного анализа клеток Кс167 на фоне инактивации *Cp190* посредством РНК-интерференции мы обнаружили повышение экспрессии 820 генов и снижение только 260 (рис. 5*а*) [13]. Активирующиеся гены также представлены преимущественно генами, специфичными для семенников, нервной ткани и пищеварительной системы (рис. 5*б*, *в*). Гены же, экспрессия которых снижалась на фоне истощения *Cp190*, имели неспецифичные профили экспрессии (рис. 5*б*, *в*).



**Рис. 6.** Делеция CP190 в клетках мужского зародышевого пути приводит к нарушению активности tMAC-зависимых генов дифференцировки сперматоцитов. *а* – Экспрессия генов из двух кластеров Набора № 1 в семенниках дикого типа и мутантов с нарушением дифференцировки клеток мужского зародышевого пути. Цветными прямоугольниками показано на каких стадиях сперматогенеза останавливается дифференцировка клеток в семенниках каждого генотипа. Гены *mip40* и *comr* кодируют компоненты семенник-специфичного транскрипционного активатора tMAC. Ген *can* кодирует один из белков tTAF. Статистическая значимость отличий экспрессии генов в семенниках мутантов и дикого типа установлена в тесте Вилкоксона. *б* – Обогащение Набора № 1 генами, вокруг которых в хромосомах клеток зародышевого пути самцов в разных участках связываются транскрипционные факторы Can, Comr и Mir40 [18]. Статистическая значимость обогащения измерена с помощью точного теста Фишера. Обозначения: TSS – сайт начала транскрипции; TES – сайт завершения транскрипции. *в* – То же, что в *а*, для генов из двух кластеров Набора № 2. *г* – То же, что в *б*, для генов из Набора № 2.

### Мутация *Sr190* нарушает экспрессию генов-мишеней tMAC

Активация генов дифференцировки происходит скоординировано в фазе роста сперматоцитов первого порядка. Описано несколько мутаций, которые нарушают этот процесс [17]. Так, мутация гена *bam* вызывает остановку сперматогенеза на стадии сперматогониев, в результате чего семенники обогащаются слабо дифференцированными клетками зародышевого пути [39]. Гены *mip40*, *comr* и *can* являются генами задержки мейоза, они кодируют транскрипционные факторы, а их мутации приводят к остановке сперматогенеза на стадии первичных сперматоцитов [17].

Нарушения в работе Mir40 и Comr, которые входят в состав семенник-специфичного ком-

плекса задержки мейоза tMAC, вызывают драматичные изменения экспрессии генов в семенниках [17, 18]. Нарушения же в работе белков из группы семенник-специфичных факторов транскрипции tTAF, включая Can, вызывают менее выраженные изменения в транскриптите семенников. Вместе tTAF и tMAC обеспечивают активацию программы дифференцировки сперматоцитов, причем только часть генов регулируется этими факторами напрямую [18].

Учитывая связь CP190 с генами дифференцировки, мы предположили функциональное взаимодействие между CP190, tTAF и tMAC. Ранее мы получили профили экспрессии генов в семенниках мутантов *bam*, *mip40*, *comr* и *can*, а также картировали сайты связывания Mir40, Comr и Can с хромосомами в клетках зародышевого пути [18].

Мы проанализировали экспрессию генов, которые активируются в семенниках с условным спасением (Набор № 1), в семенниках дикого типа и мутантов *bam*, *mip40*, *comr* и *can* (рис. 6а).

Активность генов дифференцировки сперматоцитов из Набора № 1 (рис. 4б) значительно снижена у мутантов *bam*, *mip40* и *comr*, но не у мутантов *can* (рис. 6а). Остальные гены из Набора № 1 почти не реагируют на мутации *mip40*, *comr* и *can*, только мутация *bam* приводит к незначительному снижению их экспрессии, что можно объяснить присутствием среди них генов с высокой активностью в сперматоцитах (рис. 4б). Таким образом, CP190 ограничивает активность tMAC-зависимых генов дифференцировки сперматоцитов.

Профили распределения *Comr*, *Mip40* и *Can* на хромосомах показывают, что эти белки чаще ожидаемого связываются с генами из Набора № 1, причем наиболее интенсивно связывается транскрипционный фактор *Comr* (рис. 6б). Таким образом, мутация *Sp190* затрагивает активность генов, непосредственно регулируемых комплексом tMAC.

Для экспрессии генов Набора № 2 характерны менее выраженные различия между мутантами *bam*, *mip40*, *comr* и *can*, и диким типом (рис. 6в), хотя и у мутантов с нарушениями в tMAC, и у мутантов *bam* снижается экспрессия генов дифференцировки сперматоцитов и генов, характерных для соматических клеток. Гены из Набора № 2 имеют некоторую связь с транскрипционным фактором *Comr* (рис. 6г), что указывает на возможность прямой активации их экспрессии комплексом tMAC, и белок CP190 участвует в этом процессе.

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В этой работе мы подробно изучили влияние белка CP190 на активность генов в клетках зародышевого пути *D. melanogaster*. Как оказалось, основное влияние мутация *Sp190* оказывает на субпопуляцию генов дифференцировки сперматоцитов, которые регулируются транскрипционным комплексом tMAC. Интересно, что в отсутствие белка CP190 активность этих генов усиливается в 2 раза и более, что позволяет сделать вывод о модуляции их активности этим белком в норме, ограничивающим уровень их экспрессии.

В целом, на культурах клеток дрозофилы показаны умеренные изменения транскриптома на фоне истощения инсуляторного белка CP190 [8, 13, 40]. Нами также выявлены лишь 183 гена, экспрессия которых зависит от CP190 в сперматоцитах. Экспрессия 120 из этих генов повышалась на фоне мутации *Sp190*, а 63 – снижалась, что говорит в пользу преобладающей репрессорной функции CP190. На культурах клеток показано

также, что CP190, как и другие инсуляторные белки, в большей степени склонен к подавлению транскрипции, несмотря на одновременную активаторную функцию [13, 41].

Мы обнаружили, что CP190 противоположным образом влияет на активность тканеспецифичных генов и генов домашнего хозяйства, что указывает на возможность существования нескольких механизмов, с помощью которых CP190 контролирует активность этих генов. Можно предположить, что CP190 способен выступать в качестве транскрипционного фактора для генов, активация которых зависит от CP190. Это косвенно подтверждается тем, что для активации CP190-связанных генов требуется кооперативное действие CP190 и фактора M1BP – регулятора генов домашнего хозяйства [13]. Причем промоторы генов, экспрессия которых снижается в клетках с нокдауном, обогащены CP190, в то время как репрессорный эффект CP190, по всей вероятности, реализуется не напрямую [13].

Важно отметить, что активация генетической программы дифференцировки клеток мужского зародышевого пути сопровождается значительной перестройкой архитектуры хроматина [42]. Так, на стадии сперматоцитов уменьшается количество дистальных контактов, а промоторы активных генов дифференцировки оказываются в изолированном окружении [42]. Показано участие CP190, наряду с белком Chro, в формировании и поддержании топологически ассоциированных доменов (ТАД) [41]. Причем обнаружено, что нарушение экспрессии генов в результате удаления CP190 коррелирует с изменением границ окружающего их ТАД [41]. Более того, ТАДы, содержащие гены домашнего хозяйства, имеют тенденцию к более высокой стабильности [41]. Это позволяет предположить, что наблюдаемое дифференциальное влияние CP190 на экспрессию тканеспецифичных генов и генов домашнего хозяйства может быть обусловлено различиями в механизмах регуляции на уровне структуры хроматина, в результате чего гены домашнего хозяйства активируются под действием CP190, тогда как тканеспецифичные гены подавляются.

Учитывая отсутствие влияния на дифференцировку клеток мужского зародышевого пути, нормальную фертильность мутантных самцов, а также умеренный эффект на изменение паттерна экспрессии генов, инсуляторный белок CP190 вряд ли можно рассматривать в качестве ключевого регуляторного фактора сперматогенеза. Для сравнения, мутация *comr* (tMAC) затрагивает экспрессию более 2500 генов и вызывает задержку мейоза [43]. Однако с учетом ключевой роли CP190 в регуляции архитектуры ядра можно предположить, что недостаточность CP190 приводит к сбоям в коммуникации специфичных

факторов транскрипции и их регуляторных элементов, а также к более общим нарушениям структуры хроматина, что вызывает нарушения в механизмах тонкой регуляции экспрессии.

В ходе исследования мы обнаружили, что нарушение экспрессии CP190 приводит к инактивации tMAC-зависимых генов дифференцировки сперматоцитов. Примечательно, что в соматических клетках дрозофилы функционирует гомологичный tMAC репрессорный комплекс dREAM — партнер CP190, основная функция которого состоит в регуляции генов дифференцировки и клеточного цикла [44–47]. Взаимодействие CP190 и dREAM в культурах клеток играет важную роль в управлении парами дивергентных генов — близлежащих генов с противоположно направленными промоторами [46]. Привлечение dREAM к дивергентным генам дрозофилы происходит за счет CP190 и позволяет добиваться снижения экспрессии только одного гена в паре. Наличие у dREAM и tMAC общих субъединиц (например, Mir40 и Caf1), а также связь между CP190 и tMAC-зависимыми генами позволяет предположить, что CP190 участвует в привлечении tMAC к хромосомам. Дальнейшие исследования позволят объяснить, почему основной эффект истощения CP190 в клетках мужского зародышевого пути заключается в дерегуляции генов дифференцировки сперматоцитов.

Авторы работы выражают благодарность А.К. Головнину (ИБГ РАН) за предоставленные антитела, А.В. Пиндюрину (ИМКБ СР РАН) за помощь в клонировании, а также ЦКП “Молекулярная и клеточная биология” за предоставленное оборудование.

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований в рамках научных проектов № 19-34-90108, 17-00-00181, 20-34-70141 и 19-04-00872, а также при поддержке Программы фундаментальных научных исследований FWGZ-2021-0017 (122011900429-5).

Условия содержания *Drosophila melanogaster* соответствовали нормативам, указанным в Приказе Минздрава России № 267 от 19.06.2003 “Об утверждении правил лабораторной практики в Российской Федерации”.

Авторы заявляют об отсутствии конфликта интересов.

## СПИСОК ЛИТЕРАТУРЫ

- Long H.K., Prescott S.L., Wysocka J. (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*. **167**, 1170–1187.
- Kyrchanova O., Georgiev P. (2014) Chromatin insulators and long-distance interactions in *Drosophila*. *FEBS Lett*. **588**, 8–14.
- Yang J., Corces V.G. (2011) Chromatin insulators: a role in nuclear organization and gene expression. In: *Advances in Cancer Res.* Ed. Woude G.W. **110**. London: Acad. Press, 43–76.
- Nègre N., Brown C.D., Shah P.K., Kheradpour P., Morrison C.A., Henikoff J.G., Feng X., Ahmad K., Russell S., White R.A.H., Stein L., Henikoff S., Kellis M., White K.P. (2010) A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet*. **6**, e1000814.
- Maksimenko O., Bartkuhn M., Stakhov V., Herold M., Zolotarev N., Jox T., Buxa M.K., Kirsch R., Bonchuk A., Fedotova A., Kyrchanova O., Renkawitz R., Georgiev P. (2015) Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Res*. **25**, 89–99.
- Zolotarev N., Maksimenko O., Kyrchanova O., Sokolinskaya E., Osadchiy I., Girardot C., Bonchuk A., Ciglar L., Furlong E.E.M., Georgiev P. (2017) Opbp is a new architectural/insulator protein required for ribosomal gene expression. *Nucl. Acids Res*. **45**, 12285–12300.
- Cuartero S., Fresán U., Reina O., Planet E., Espinàs M.L. (2014) Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *EMBO J*. **33**, 637–647.
- Schwartz Y.B., Linder-Basso D., Kharchenko P.V., Tolstorukov M.Y., Kim M., Li H.-B., Gorchakov A.A., Minoda A., Shanower G., Alekseyenko A.A., Riddle N.C., Jung Y.L., Gu T., Plachetka A., Elgin S.C.R., Kuroda M.I., Park P.J., Savitsky M., Karpen G.H., Pirrotta V. (2012) Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Res*. **22**, 2188–2198.
- Kaushal A., Dorier J., Wang B., Mohana G., Taschner M., Cousin P., Waridel P., Iseli C., Semenova A., Restrepo S., Guex N., Aiden E. L., Gambetta M.C. 2022. Essential role of Cp190 in physical and regulatory boundary formation. *Sci. Adv*. **8**, eabl8834.
- Sabirov M., Kyrchanova O., Pokholkova G.V., Bonchuk A., Klimenko N., Belova E., Zhimulev I.F., Maksimenko O., Georgiev P. (2021) Mechanism and functional role of the interaction between CP190 and the architectural protein Pita in *Drosophila melanogaster*. *Epigenet. Chromatin*. **14**, 16.
- Bohla D., Herold M., Panzer I., Buxa M.K., Ali T., Demmers J., Krüger M., Scharfe M., Jarek M., Bartkuhn M., Renkawitz R. (2014) A functional insulator screen identifies NURF and dREAM components to be required for enhancer-blocking. *PLoS One*. **9**, e107765.
- Ali T., Krüger M., Bhujji S., Jarek M., Bartkuhn M., Renkawitz R. (2017) Chromatin binding of Gcn5 in *Drosophila* is largely mediated by CP190. *Nucl. Acids Res*. **45**, 2384–2395.
- Bag I., Chen S., Rosin L.F., Chen Y., Liu C.-Y., Yu G.-Y., Lei E.P. (2021) M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. *Nat. Commun*. **12**, 4170.
- Savitsky M., Kim M., Kravchuk O., Schwartz Y.B. (2016) Distinct roles of chromatin insulator proteins in control of the *Drosophila* bithorax complex. *Genetics*. **202**, 601–617.

15. Butcher R.D.J., Chodagam S., Basto R., Wakefield J.G., Henderson D.S., Raff J.W., Whitfield W.G.F. (2004) The *Drosophila* centrosome-associated protein CP190 is essential for viability but not for cell division. *J. Cell Sci.* **117**, 1191–1199.
16. Oliver D., Sheehan B., South H., Akbari O., Pai C.Y. (2010) The chromosomal association/dissociation of the chromatin insulator protein Cp190 of *Drosophila melanogaster* is mediated by the BTB/POZ domain and two acidic regions. *BMC Cell Biol.* **11**, 101.
17. White-Cooper H. (2010) Molecular mechanisms of gene regulation during *Drosophila* spermatogenesis. *Reproduction.* **139**, 11–21.
18. Laktionov P.P., Maksimov D.A., Romanov S.E., Antoshina P.A., Posukh O.V., White-Cooper H., Koryakov D.E., Belyakin S.N. (2018) Genome-wide analysis of gene regulation mechanisms during *Drosophila* spermatogenesis. *Epigenetics Chromatin.* **11**, 14.
19. Markstein M., Pitsouli C., Villalta C., Celniker S.E., Perrimon N. (2008) Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat. Genet.* **40**, 476–483.
20. Лактионов П.П., Максимов Д.А., Андреева Е.Н., Шлома В.В., Белякин С.Н. (2013) Генетическая система для мечения соматических и герминальных клеточных линий в гонадах *Drosophila melanogaster*. *Цитология.* **55**, 185–189.
21. Solovei I., Cremer M. (2010) 3D-FISH on cultured cells combined with immunostaining. In: *Methods Mol. Biol.* **659**. Eds Bridger J.M., Volpi E.M. New York: Humana Press, 117–126.
22. Golovnin A., Volkov I., Georgiev P. (2012) SUMO conjugation is required for the assembly of *Drosophila* Su(Hw) and Mod(mdg4) into insulator bodies that facilitate insulator complex formation. *J. Cell Sci.* **125**, 2064–2074.
23. Kim D., Paggi J.M., Park C., Bennett C., Salzberg S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915.
24. Liao Y., Smyth G.K., Shi W. (2014) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* **30**, 923–930.
25. Love M.I., Huber W., Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
26. Stephens M. (2016) False discovery rates: a new deal. *Biostatistics.* **18**, 275–294.
27. Ignatiadis N., Klaus B., Zaugg J.B., Huber W. (2016) Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Meth.* **13**, 577–580.
28. Brown J.B., Boley N., Eisman R., May G.E., Stoiber M.H., Duff M.O., Booth B.W., Wen J., Park S., Suzuki A.M., Wan K.H., Yu C., Zhang D., Carlson J.W., Cherbas L., Eads B.D., Miller D., Mockaitis K., Roberts J., Davis C.A., Frise E., Hammonds A.S., Olson S., Shenker S., Sturgill D., Samsonova A.A., Weizmann R., Robinson G., Hernandez J., Andrews J., Bickel P.J., Carninci P., Cherbas P., Gingers T.R., Hoskins R.A., Kaufman T.C., Lai E.C., Oliver B., Perrimon N., Graveley B.R., Celniker S.E. (2014) Diversity and dynamics of the *Drosophila* transcriptome. *Nature.* **512**, 393–399.
29. de Hoon M.J.L., Imoto S., Nolan J., Miyano S. (2004) Open source clustering software. *Bioinformatics.* **20**, 1453–1454.
30. Jain A., Tuteja G. (2019) TissueEnrich: tissue-specific gene enrichment analysis. *Bioinformatics.* **35**, 1966–1967.
31. Li H., Janssens J., De Waegeneer M., Kolluru S.S., Davie K., Gardeux V., Saelens W., David F.P.A., Brbić M., Spanier K., Leskovec J., McLaughlin C.N., Xie Q., Jones R.C., Brueckner K., Shim J., Tattikota S.G., Schnorrer F., Rust K., Nystul T.G., Carvalho-Santos Z., Ribeiro C., Pal S., Mahadevaraju S., Przytycka T.M., Allen A.M., Goodwin S.F., Berry C.W., Fuller M.T., White-Cooper H., Matunis E.L., DiNardo S., Galenza A., O'Brien L.E., Dow J.A.T.; FCA Consortium, Jasper H., Oliver B., Perrimon N., Deplancke B., Quake S.R., Luo L., Aerts S., Agarwal D., Ahmed-Braimah Y., Arbeitman M., Ariss M.M., Augsburg J., Ayush K., Baker C.C., Banisch T., Birker K., Bodmer R., Bolival B., Brantley S.E., Brill J.A., Brown N.C., Buehner N.A., Cai X.T., Cardoso-Figueiredo R., Casares F., Chang A., Clandinin T.R., Crasta S., Desplan C., Detweiler A.M., Dhakan D.B., Donà E., Engert S., Floc'hlay S., George N., González-Segarra A.J., Groves A.K., Gumbin S., Guo Y., Harris D.E., Heifetz Y., Holtz S.L., Horns F., Hudry B., Hung R.J., Jan Y.N., Jaszczak J.S., Jeffers G.S.X.E., Karkanas J., Karr T.L., Katheder N.S., Kezou J., Kim A.A., Kim S.K., Kockel L., Konstantinides N., Kornberg T.B., Krause H.M., Labott A.T., Laturney M., Lehmann R., Leinwand S., Li J., Li J.S.S., Li K., Li K., Li L., Li T., Litovchenko M., Liu H.H., Liu Y., Lu T.C., Manning J., Mase A., Matera-Vatnick M., Matias N.R., McDonough-Goldstein C.E., McGeever A., McLachlan A.D., Moreno-Roman P., Neff N., Neville M., Ngo S., Nielsen T., O'Brien C.E., Osumi-Sutherland D., Özel M.N., Papatheodorou I., Petkovic M., Pilgrim C., Pisco A.O., Reisenman C., Sanders E.N., Dos Santos G., Scott K., Sherlekar A., Shiu P., Sims D., Sit R.V., Slaidina M., Smith H.E., Sterne G., Su Y.H., Sutton D., Tamayo M., Tan M., Tastekin I., Treiber C., Vacek D., Vogler G., Waddell S., Wang W., Wilson R.I., Wolfner M.F., Wong Y.E., Xie A., Xu J., Yamamoto S., Yan J., Yao Z., Yoda K., Zhu R., Zinzen R.P. (2022) Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science.* **375**, eabk2432.
32. Hao Y., Hao S., Andersen-Nissen E., Mauck W.M., Zheng S., Butler A., Lee M.J., Wilk A.J., Darby C., Zager M., Hoffman P., Stoeckius M., Papalexi E., Mimitou E.P., Jain J., Srivastava A., Stuart T., Fleming L.M., Yeung B., Rogers A.J., McElrath J.M., Blish C.A., Gottardo R., Smibert P., Satija R. (2021) Integrated analysis of multimodal single-cell data. *Cell.* **184**, 3573–3587. e29.
33. Maksimov D.A., Laktionov P.P., Belyakin S.N. (2016) Data analysis algorithm for DamID-seq profiling of chromatin proteins in *Drosophila melanogaster*. *Chromosome Res.* **24**(4), 481–494.
34. Akbari O.S., Oliver D., Eyer K., Pai C.-Y.Y. (2009) An entry/gateway cloning system for general expression of genes with molecular tags in *Drosophila melanogaster*. *BMC Cell Biol.* **10**, 8.

35. Evans C.J., Olson J.M., Ngo K.T., Kim E., Lee N.E., Kuoy E., Patananan A.N., Sitz D., Tran P., Do M.-T., Yackle K., Cespedes A., Hartenstein V., Call G.B., Banerjee U. (2009) G-TRACE: rapid Gal4-based cell lineage analysis in *Drosophila*. *Nat. Methods*. **6**, 603–605.
36. Pindyurin A.V., Pagie L., Kozhevnikova E.N., van Arensbergen J., van Steensel B. (2016) Inducible DamID systems for genomic mapping of chromatin proteins in *Drosophila*. *Nucl. Acids Res.* **44**, 5646–5657.
37. Lee T., Luo L. (2001) Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development. *Trends Neurosci.* **24**, 251–254.
38. Hu Y., Tattikota S.G., Liu Y., Comjean A., Gao Y., Forman C., Kim G., Rodiger J., Papatheodorou I., Santos G.D., Mohr S.E., Perrimon N. (2021) DRscDB: A single-cell RNA-seq resource for data mining and data comparison across species. *Comput. Struct. Biotechnol. J.* **19**, 2018–2026.
39. Gönczy P., Matunis E., DiNardo S. (1997) bag-of-marbles and benign gonial cell neoplasm act in the germline to restrict proliferation during *Drosophila* spermatogenesis. *Development*. (Cambridge, England). **124**, 4361–4371.
40. Bartkuhn M., Straub T., Herold M., Herrmann M., Rathke C., Saumweber H., Gilfillan G.D., Becker P.B., Renkawitz R. (2009) Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J.* **28**, 877–888.
41. Chathoth K.T., Mikheeva L.A., Crevel G., Wolfe J.C., Hunter I., Beckett-Doyle S., Cotterill S., Dai H., Harrison A., Zabet N.R. (2022) The role of insulators and transcription in 3D chromatin organization of flies. *Genome Res.* **32**, 682–698.
42. Ilyin A.A., Kononkova A.D., Golova A.V., Shloma V.V., Olenkina O.M., Nenashcheva V.V., Abramov Y.A., Kotov A.A., Maksimov D.A., Laktionov P.P., Pindyurin A.V., Galitsyna D.A., Ulianov S.V., Khrameeva E.E., Gelfand M.S., Belyakin S.N., Razin S.V., Shevelyov Y.Y. (2022) Comparison of genome architecture at two stages of male germline cell differentiation in *Drosophila*. *Nucl. Acids Res.* **50**, 3203–3225.
43. Лактионов П.П., White-Соорер Н., Максимов Д.А., Белянкин С.Н. (2014) Транскрипционный фактор Comg играет роль прямого активатора в генетической программе сперматогенеза у *D. melanogaster*. *Молекуляр. биология*. **48**, 153–165. <https://doi.org/10.1134/S0026893314010087>
44. Lee H., Ohno K., Voskoboynik Y., Ragusano L., Martinez A., Dimova D.K. (2010) *Drosophila* RB proteins repress differentiation-specific genes via two different mechanisms. *Mol. Cell. Biol.* **30**, 2563–2577.
45. DeBruhl H., Wen H., Lipsick J.S. (2013) The complex containing *Drosophila* Myb and RB/E2F2 regulates cytokinesis in a histone H2Av-dependent manner. *Mol. Cell. Biol.* **33**, 1809–1818.
46. Korenjak M., Kwon E., Morris R.T., Anderssen E., Amzallag A., Ramaswamy S., Dyson N.J. (2014) dREAM co-operates with insulator-binding proteins and regulates expression at divergently paired genes. *Nucl. Acids Res.* **42**, 8939–8953.
47. Beall E.L., Lewis P.W., Bell M., Rocha M., Jones D.L., Botchan M.R. (2007) Discovery of tMAC: a *Drosophila* testis-specific meiotic arrest complex paralogous to Myb-Muv B. *Genes Dev.* **21**, 904–919.

## Insulator Protein CP190 Regulates Expression of Spermatocyte Differentiation Genes in *Drosophila melanogaster* Male Germline

S. E. Romanov<sup>1, 2</sup>, V. V. Shloma<sup>1</sup>, D. E. Koryakov<sup>1</sup>, S. N. Belyakin<sup>1</sup>, and P. P. Laktionov<sup>1, 2, \*</sup>

<sup>1</sup>*Institute of Molecular and Cellular Biology, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

<sup>2</sup>*Novosibirsk State University, Novosibirsk, 630090 Russia*

\*e-mail: laktionov@mcb.nsc.ru

CP190 protein is one of the key components of *Drosophila* insulator complexes, and its study is important for understanding the mechanisms of gene regulation during cell differentiation. However, *Cp190* mutants die before reaching adulthood, which significantly complicates the study of its functions in imago. To overcome this problem and to investigate the regulatory effects of CP190 in adult tissues development, we have designed a conditional rescue system for *Cp190* mutants. Using Cre/loxP-mediated recombination, the rescue construct containing *Cp190* coding sequence is effectively eliminated specifically in spermatocyte, allowing us to study the effect of the mutation in male germ cells. Using high-throughput transcriptome analysis we determined the function of CP190 on gene expression in germline cells. *Cp190* mutation was found to have opposite effects on tissue-specific genes, which expression is repressed by CP190, and housekeeping genes, that require CP190 for activation. Mutation of *Cp190* also promoted expression of a set of spermatocyte differentiation genes that are regulated by tMAC transcriptional complex. Our results indicate that the main function of CP190 in the process of spermatogenesis is the coordination of interactions between differentiation genes and their specific transcriptional activators.

**Keywords:** CP190, insulator proteins, *Drosophila melanogaster*, RNA-seq

## СТРУКТУРНО-ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ БИОПОЛИМЕРОВ И ИХ КОМПЛЕКСОВ

УДК 57.013:577.359;57.088

### РАЗДЕЛЕНИЕ ФАЗ В ПРЕПАРАТЕ ОЧИЩЕННОГО БЕЛКА ЧЕЛОВЕКА LSM4<sup>1</sup>

© 2023 г. Н. Li<sup>a</sup>, Y. Ju<sup>b</sup>, W. W. Liu<sup>a</sup>, Y. Y. Ma<sup>c</sup>, H. Ye<sup>c</sup>, \*, N. Li<sup>c</sup>, \*\*

<sup>a</sup>Health Care Office, Service Bureau of The General Administration of Affairs, The Central Military Commission, Beijing, 100034 China

<sup>b</sup>Department of Obstetrics and Gynecology, The Seventh Medical Center of Chinese People's Liberation Army General Hospital, Beijing, 100071 China

<sup>c</sup>Department of Minimally Invasive Gynecologic Center, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing Maternal and Child Health Care Hospital, Beijing, 100006 China

\*e-mail: yehong@cmmu.edu.cn

\*\*e-mail: linaog@126.com

Поступила в редакцию 17.02.2022 г.

После доработки 04.07.2022 г.

Принята к публикации 20.07.2022 г.

Жидкофазное разделение белков происходит в ряде биологических процессов, таких как регуляция транскрипции, процессинг и созревание РНК. Sm-подобный белок 4 (LSM4) участвует во множестве процессов, включая сплайсинг пре-мРНК и сборку Р-телец. Перед проведением исследований по участию LSM4 в разделении двух жидких фаз в ходе процессинга или созревания РНК необходимо сначала обнаружить эту способность белка LSM4 в системе *in vitro*. Плазмиду, экспрессирующую белок mCherry-LSM4, получали на основе вектора рЕТ30а и использовали для выделения белка mCherry-LSM4 из прокариотических клеток (*Escherichia coli* штамм BL21). Белок mCherry-LSM4 очищали с использованием аффинной хроматографии на Ni-NTA-агарозе, а затем жидкостной хроматографии быстрого разрешения (FPLC). Широкопольную флуоресцентную микроскопию Delta-Vision использовали для наблюдения за динамическим разделением фаз жидкость–жидкость в препаратах белка LSM4 *in vitro*. В ходе анализа структуры белка LSM4 с помощью базы данных Predictor of Natural Disordered Regions установлено, что его С-конец содержит домен низкой сложности. Из *E. coli* получен очищенный препарат полноразмерного белка LSM4 человека. Показано, что LSM4 человека обеспечивает зависимое от концентрации разделение фаз жидкость–жидкость *in vitro* в буфере с краудинг-агентами. Соли в высокой концентрации и 1,6-гександиол блокируют LSM4-индуцированное разделение двух жидких фаз. Кроме того, *in vitro* наблюдается слияние капель белка LSM4. Полученные результаты свидетельствуют о том, что полноразмерный белок LSM4 человека обладает способностью образовывать жидкие включения и вызывать разделение фаз жидкость–жидкость *in vitro*.

**Ключевые слова:** разделение фаз жидкость–жидкость, Sm-подобный белок 4, жидкостная экспресс-хроматография белков, очистка белков, биофизический процесс

**DOI:** 10.31857/S0026898423010068, **EDN:** AWQDCU

#### СПИСОК ЛИТЕРАТУРЫ

1. Banani S.F., Lee H.O., Hyman A.A., Rosen M.K. (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell. Biol.* **18**(5), 285–298.
2. Mitrea D.M., Cika J.A., Stanley C.B., Nourse A., Onuchic P.L., Banerjee P.R., Phillips A.H., Park C.G., Deniz A.A., Kriwacki R.W. (2018) Self-interaction of
3. McSwiggen D.T., Hansen A.S., Teves S.S., Marie-Nelly H., Hao Y., Heckert A.B., Umemoto K.K., Dugast-Darzacq C., Tjian R., Darzacq X. (2019) Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *Elife.* **8**, e47098.

<sup>1</sup> Статья представлена авторами на английском языке.

4. Hyman A.A., Weber C.A., Julicher F. (2014) Liquid-liquid phase separation in biology. *Annu. Rev. Cell. Dev. Biol.* **30**, 39–58.
5. Aguzzi A., Altmeyer M. (2016) Phase separation: linking cellular compartmentalization to disease. *Trends Cell Biol.* **26**(7), 547–558.
6. Bergeron-Sandoval L.P., Safaee N., Michnick S.W. (2016) Mechanisms and consequences of macromolecular phase separation. *Cell.* **165**(5), 1067–1079.
7. Toretzky J.A., Wright P.E. (2014) Assemblages: functional units formed by cellular phase separation. *J. Cell Biol.* **206**(5), 579–588.
8. Lyon A.S., Peeples W.B., Rosen M.K. (2021) A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**(3), 215–235.
9. Sanulli S., Trnka M.J., Dharmarajan V., Tibble R.W., Pascal B.D., Burlingame A.L., Griffin P.R., Gross J.D., Narlikar G.J. (2019) HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature.* **575**(7782), 390–394.
10. Zhao Y.G., Zhang H. (2020) Phase separation in membrane biology: the interplay between membrane-bound organelles and membraneless condensates. *Dev. Cell.* **55**(1), 30–44.
11. Jiang H., Wang S., Huang Y., He X., Cui H., Zhu X., Zheng Y. (2015) Phase transition of spindle-associated protein regulate spindle apparatus assembly. *Cell.* **163**(1), 108–122.
12. Huang Y., Li T., Ems-McClung S.C., Walczak C.E., Prigent C., Zhu X., Zhang X., Zheng Y. (2018) Aurora A activation in mitosis promoted by BuGZ. *J. Cell. Biol.* **217**(1), 107–116.
13. Lee K.H., Zhang P., Kim H.J., Mitrea D.M., Sarkar M., Freibaum B.D., Cika J., Coughlin M., Messing J., Mollieux A., Maxwell B.A., Kim N.C., Temirov J., Moore J., Kolaitis R.M., Shaw T.I., Bai B., Peng J., Kriwacki R.W., Taylor J.P. (2016) C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. *Cell.* **167**(3), 774–788. e717.
14. Altmeyer M., Neelsen K.J., Teloni F., Pozdnyakova I., Pellegrino S., Grøfte M., Rask M.D., Streicher W., Jungmichel S., Nielsen M.L., Lukas J. (2015) Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun.* **6**, 8088.
15. Boeynaems S., Alberti S., Fawzi N.L., Mittag T., Polyimenidou M., Rousseau F., Schymkowitz J., Shorter J., Wolozin B., Van Den Bosch L., Tompa P., Fuxreiter M. (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol.* **28**(6), 420–435.
16. Shin Y., Brangwynne C.P. (2017) Liquid phase condensation in cell physiology and disease. *Science.* **357**, 6357.
17. Patel A., Lee H.O., Jawerth L., Maharana S., Jahnke M., Hein M.Y., Stoyanov S., Mahamid J., Saha S., Franzmann T.M., Pozniakowski A., Poser I., Maghelli N., Royer L.A., Weigert M., Myers E.W., Grill S., Drechsel D., Hyman A.A., Alberti S. (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell.* **162**(5), 1066–1077.
18. Pannone B.K., Kim S.D., Noe D.A., Wolin S.L. (2001) Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein. *Genetics.* **158**(1), 187–196.
19. Arribas-Layton M., Dennis J., Bennett E.J., Damgaard C.K., Lykke-Andersen J. (2016) The C-terminal RGG domain of human Lsm4 promotes processing body formation stimulated by arginine dimethylation. *Mol. Cell. Biol.* **36**(17), 2226–2235.
20. Reijns M.A., Alexander R.D., Spiller M.P., Beggs J.D. (2008) A role for Q/N-rich aggregation-prone regions in P-body localization. *J. Cell Sci.* **121**(Pt 15), 2463–2472.
21. Roth A.J., Shuman S., Schwer B. (2018) Defining essential elements and genetic interactions of the yeast Lsm2-8 ring and demonstration that essentiality of Lsm2-8 is bypassed via overexpression of U6 snRNA or the U6 snRNP subunit Prp24. *RNA.* **24**(6), 853–864.
22. Lyons S.M., Ricciardi A.S., Guo A.Y., Kambach C., Marzluff W.F. (2014) The C-terminal extension of Lsm4 interacts directly with the 3' end of the histone mRNP and is required for efficient histone mRNA degradation. *RNA.* **20**(1), 88–102.
23. Decker C.J., Teixeira D., Parker R. (2007) Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in *Saccharomyces cerevisiae*. *J. Cell Biol.* **179**(3), 437–449.
24. Mennie A.K., Moser B.A., Nakamura T.M. (2018) LARP7-like protein Pof8 regulates telomerase assembly and poly(A)<sup>+</sup> TERRA expression in fission yeast. *Nat. Commun.* **9**(1), 1–12.
25. Adamson B.S., Smogorzewska A., Sigoillot F.D., King R.W., Elledge S.J. (2012) A genome-wide study of homologous recombination in mammalian cells identifies RBMX, a novel component of the DNA damage response. *Nat. Cell Biol.* **14**(3), 318–328.
26. Alabrudzinska M., Skoneczny M., Skoneczna A. (2011) Diploid-specific genome stability genes of *S. cerevisiae*: genomic screen reveals haploidization as an escape from persisting DNA rearrangement stress. *PLoS One.* **6**(6), e21124.
27. Lin Y., Protter D.S., Rosen M.K., Parker R. (2015) Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell.* **60**(2), 208–219.
28. Linding R., Jensen L.J., Diella F., Bork P., Gibson T.J., Russell R.B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure.* **11**(11), 1453–1459.
29. Schuster B.S., Reed E.H., Parthasarathy R., Jahnke C.N., Caldwell R.M., Bermudez J.G., Ramage H., Good M.C., Hammer D.A. (2018) Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **9**(1), 2985.
30. Uversky V.N. (2017) Intrinsically disordered proteins in overcrowded milieu: membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.* **44**, 18–30.
31. Sasahara K., McPhie P., Minton A.P. (2003) Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J. Mol. Biol.* **326**(4), 1227–1237.
32. Alberti S., Saha S., Woodruff J.B., Franzmann T.M., Wang J., Hyman A.A. (2018) A user's guide for phase

- separation assays with purified proteins. *J. Mol. Biol.* **430**(23), 4806–4820.
33. Alberti S., Gladfelter A., Mittag T. (2019) Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell.* **176**(3), 419–434.
34. Verdone L., Galardi S., Page D., Beggs J.D. (2004) Lsm proteins promote regeneration of pre-mRNA splicing activity. *Curr. Biol.* **14**(16), 1487–1491.
35. Rao B.S., Parker R. (2017) Numerous interactions act redundantly to assemble a tunable size of P bodies in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA.* **114**(45), E9569–E9578.
36. Tang W., Kannan R., Blanchette M., Baumann P. (2012) Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature.* **484**(7393), 260–264.

## Phase Separation of Purified Human LSM4 Protein

H. Li<sup>1</sup>, Y. Ju<sup>2</sup>, W. W. Liu<sup>1</sup>, Y. Y. Ma<sup>3</sup>, H. Ye<sup>3, \*</sup>, and N. Li<sup>3, \*\*</sup>

<sup>1</sup>Health Care Office, Service Bureau of The General Administration of Affairs, The Central Military Commission, Beijing, 100034 China

<sup>2</sup>Department of Obstetrics and Gynecology, The Seventh Medical Center of Chinese People's Liberation Army General Hospital, Beijing, 100071 China

<sup>3</sup>Department of Minimally Invasive Gynecologic Center, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing Maternal and Child Health Care Hospital, Beijing, 100006 China

\*e-mail: yehong@ccmu.edu.cn

\*\*e-mail: linaog@126.com

Liquid–liquid phase separation of proteins occur in a number of biological processes, such as regulation of transcription, processing, and RNA maturation. Sm-like protein 4 (LSM4) is involved in multiple processes, including pre-mRNA splicing and P-bodies assembly. Before investigating the involvement of LSM4 in the separation of the two liquid phases during RNA processing or maturation, the separation of the liquid phases in an *in vitro* preparation of LSM4 protein should be first be detected. The mCherry-LSM4 plasmid was derived from pET30a and used to isolate mCherry-LSM4 protein from prokaryotic cells (*Escherichia coli* strain BL21). The mCherry-LSM4 protein was purified using Ni-NTA resin. The protein was further purified by fast protein liquid chromatography. Delta-Vision wide-field fluorescence microscopy was used to observe the dynamic liquid–liquid phase separation of the LSM4 protein *in vitro*. Analysis of the LSM4 protein structure using the Predictor of Natural Disordered Regions database revealed that its C-terminus contains a low complexity domain. A purified preparation of full-length human LSM4 protein was obtained from *E. coli*. Human LSM4 was shown to provide concentration-dependent separation of liquid–liquid phases *in vitro* in buffer with crowding reagents. Salts in high concentration and 1,6-hexanediol block the LSM4-induced separation of the two liquid phases. In addition, *in vitro* fusion of LSM4 protein droplets is observed. These results indicate that the full-length human LSM4 protein has the ability to form liquid inclusions and induce liquid–liquid phase separation *in vitro*.

**Keywords:** liquid–liquid phase separation, Sm-like protein 4, fast protein liquid chromatography, protein purification, biophysical process

УДК 575.322

## МЕТОД КОМПЛЕКСНОГО ФОРМИРОВАНИЯ ПРЕДИКТОРОВ ДЛЯ ПРИМЕНЕНИЯ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧЕ ПРЕДСКАЗАНИЯ СТРУКТУРЫ И ФУНКЦИЙ БЕЛКА

© 2023 г. Ю. В. Мильчевский<sup>a, \*</sup>, В. Ю. Мильчевская<sup>a, b</sup>, Ю. В. Кравацкий<sup>a, c</sup><sup>a</sup>Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук,  
Москва, 119991 Россия<sup>b</sup>Institute of Medical Statistics and Bioinformatics, Faculty of Medicine, University of Cologne,  
Cologne, 50931 Germany<sup>c</sup>Центр высокоточного редактирования и генетических технологий для биомедицины,  
Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991 Россия

\*e-mail: milch@eimb.ru

Поступила в редакцию 10.06.2022 г.

После доработки 31.07.2022 г.

Принята к публикации 01.09.2022 г.

Повышение точности предсказания структуры и функций белков в последнее время связано в основном с применением и совершенствованием методов машинного обучения. Кодирование информации, содержащейся в последовательности аминокислот, – первый этап предсказания структуры, и поэтому оно играет фундаментальную роль в конечном успехе этих методов. Мы предлагаем единую методику генерации предикторов сложного вида, позволяющую формализовать предположения о факторах, которые влияют на структуру и функцию белка. Кроме того, в рамках этой задачи предложен подход к созданию и использованию баз данных структурных свойств, предоставляющих новые возможности для статистического описания и анализа структурных свойств. Предложенные методы позволяют создавать и тестировать наборы предикторов (описывающих факторы, которые влияют на структуру и функцию белка) как для конкретных задач, так и универсальных. Статистические методы построения моделей, которые мы используем, позволяют отбирать статистически значимые предикторы и улучшать таким образом предсказательные модели. На классическом примере предсказания вторичной структуры белка мы показали эффективность данного подхода, получив точность предсказания для трех классов DSSP: Q3 = 81.3%. Предложенный метод реализован в виде мультиплатформенной программы на языке C++ для командной строки. Исходный код и использованные в этой работе данные расположены по ссылке <https://github.com/Milchevskiy/protein-encoding-projects>

**Ключевые слова:** белок, вторичная структура, функция, предикторы, предсказание, пошаговый регрессионный анализ, пошаговый дискриминантный анализ

**DOI:** 10.31857/S0026898423010093, **EDN:** AWNZLZ

### ВВЕДЕНИЕ

Согласно термодинамической гипотезе Anfinsen [1], информация о структуре и функциях белка содержится в его аминокислотной последовательности. Так, из этой последовательности может быть извлечена информация о физико-химических свойствах белка, контактные потенциалы аминокислот в цепи, эволюционные данные и другая информация. На основе этих данных строят предсказания структуры и свойств белков.

В последние годы наблюдали значительный прогресс в методах предсказания структуры и функций белка на основе первичной структуры. Существенные продвижения произошли в таких

задачах, как предсказание вторичной [2] и локальной [3] структуры белка, белковых контактов [4], белоксвязывающих участков [5] и др. Прогресс в этих задачах достигнут в основном за счет использования методов машинного обучения, особенно глубокого обучения [6]. Фактически, в текущий момент высокая точность предсказания структуры и функций белка достижима только методами машинного обучения. Подготовка данных для обучения предсказательных моделей – наиболее трудоемкая задача, по-видимому, как в методическом, так и в алгоритмическом смысле. Суть проблемы в том, что генерация входных данных для машинного обучения в задачах предсказания структур и функций белков – нестандарт-

ная процедура, которая обычно реализуется в контексте конкретной задачи. Кроме того, огромный массив данных о свойствах аминокислот содержится в разрозненных источниках.

Чтобы получить адекватные результаты, основанные на машинном обучении, обычно необходимы три составляющих: репрезентативные обучающие выборки, мощные алгоритмы машинного обучения и эффективное представление объектов через набор предикторов. Последние достижения в предсказании структуры белков связаны с увеличением объема доступных данных о структуре и функции белков, а также с революционным прогрессом в развитии алгоритмов глубокого машинного обучения [2]. В то же время методы представления белковых данных (*amino acid encoding method*, кодирование последовательности аминокислот) не привлекали пристального внимания исследователей.

Экстенсивные методы совершенствования предсказаний, связанные с ростом объемов доступных баз данных, а также с улучшением методов собственно машинного обучения, практически исчерпали себя. Мы полагаем, что повышение точности предсказаний лежит в области развития подходов к кодированию биологических данных в предикторы для последующего использования в предсказательных моделях. Создаваемый набор предикторов должен давать возможность формулировать и тестировать наши предположения о природе связи между последовательностью и структурой через физико-химические свойства, определяемые этой последовательностью.

## МАТЕРИАЛЫ И МЕТОДЫ

В последние десятилетия предложены различные алгоритмы кодирования аминокислотных последовательностей в предикторы [7, 8]. Наиболее широко для формирования предикторов используют следующие кодировки: *one-hot encoding*, *position specific scoring matrix (PSSM)*, а также физико-химические кодировки свойств аминокислот в белковой цепи. В дополнение к этим предложены такие варианты, как кодирование, оцененное по энергиям контакта между остатками [9] и полученное в результате выравнивания структуры белка [10] и из контекста последовательности [11]. Подробный обзор и классификация современных методов формирования предикторов для белковых последовательностей представлены в обзоре X. Jing и др. [12].

Сформулированные ранее подходы разнородны. Единый подход, включающий в себя формирование предикторов на основе всей доступной информации по аминокислотной последовательности (физической, химической, биологической),

отсутствует. Например, в базе *AAindex* [13] собрана информация о более чем 550 физико-химических свойствах аминокислот. Подавляющее большинство этих свойств не использовали при предсказаниях методами машинного обучения.

Мы предлагаем единый алгоритм генерации предикторов сложного вида, позволяющий формализовать предположения о факторах, влияющих на структуру и функцию белков. Этот алгоритм включает, в том числе, генерацию регулярных выражений для создания предикторов, создание и использование баз данных структурных свойств и формирование предикторов непосредственно из физико-химических свойств.

Кроме того, метод позволяет создавать комбинированные предикторы, составленные из всех выше перечисленных типов. Ниже мы подробно описываем типы создаваемых предикторов, методы их создания и анализа в процессе построения предварительных моделей с помощью пошагового регрессионного анализа и пошагового дискриминантного анализа.

**Предикторы, основанные на базе свойств аминокислот AAindex.** Как уже упоминали, в базе данных *AAindex* [13] систематизировано более 550 физико-химических свойств аминокислот. Кроме них эта база данных также содержит различные матрицы аминокислотных мутаций (*amino acid mutation matrices*) и потенциалы попарных контактов (*pair-wise contact potentials*). По предлагаемому нами методике принцип формирования предикторов для всех этих данных одинаков.

Простейшим способом генерации предикторов для характеристики некоторого свойства белковой цепи (например, гидрофобности) является сопоставление последовательности набору соответствующих каждой аминокислоте значений гидрофобности. Однако наивно предполагать, что свойства аминокислотной последовательности описываются линейно гидрофобностью отдельных аминокислот.

Проиллюстрируем генерацию набора предикторов по предсказанию локальной структуры белка [14]. Рассмотрим фрагмент последовательности из  $n$  остатков. Каждая аминокислота  $\{a_i, i = 1, \dots, n\}$  имеет определенное значение гидрофобности  $\{H_i, i = 1, \dots, n\}$ . Далее к этому массиву значений свойств применимо функциональное преобразование:

$$F(H_1, H_2, \dots, H_n) = \sqrt{\left(\sum_{k=1}^n H_k \cos\left(k \frac{2\pi}{T}\right)\right)^2 + \left(\sum_{k=1}^n H_k \sin\left(k \frac{2\pi}{T}\right)\right)^2}, \quad (1)$$

где  $T$  – период,  $k$  – номер аминокислоты внутри фрагмента,  $H_k$  – гидрофобность  $k$ -й аминокислоты. В данном случае функциональная трансформация

**Таблица 1.** Пример распределения аминокислот по группам внутри тетрапептида

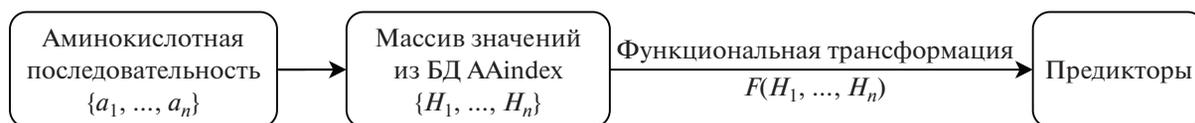
Сдвиг	Распределение аминокислот по группам эквивалентности
-2	[ARNDCQENILKMFPSTWYV] [G] [P] [O]
-1	[HRK] [DENQ] [C] [STPAG] [MILV] [FYW] [O]
0	[A] [R] [N] [D] [C] [Q] [E] [H] [I] [L] [K] [M] [F] [P] [S] [T] [W] [Y] [V] [G] [P]
1	[ARNDCQENILKMFPSTWYV] [G] [P] [O]

ция отражает периодическое изменение гидрофобности с периодом  $T$  во фрагменте из  $n$  остатков.

Если рассматриваемый участок находится в периодической (например, спиральной) конформации, то величина, рассчитанная по формуле (1), будет максимальна для периода  $T$  этой спирали. Остается “угадать” этот период, а также минимальное число остатков  $n$ , на котором сформированный предиктор будет значимым, т.е. отличаться от случайного шума по заданным статистическим критериям. Способы задания статистических критериев подробно описаны ниже в подразделе “пошаговый регрессионный анализ”. Далее, в базе AAindex представлено более 30 свойств, называемых шкалами гидрофобности. Какую из них

выбрать? Расчетные методы, которые мы использовали для отладки модели, позволяют количественно сравнивать значимость множества предикторов. В рассматриваемом случае мы собрали в набор предикторы, построенные по формуле (1), для всех шкал гидрофобности и при различных диапазонах параметров  $T$  и  $n$ . Применение пошагового регрессионного анализа позволило выявить наиболее значимые предикторы и отбросить несущественные, что представляет основной этап усовершенствования модели.

В общем виде схема генерации предикторов, основанных на физико-химических свойствах, предполагает этап функциональной трансформации:



Алгоритм предполагает также возможность создания исследователями собственных функций функциональной трансформации, отражающих их гипотезы о возможных факторах, влияющих на структуру и функции белков.

**Регулярные выражения как предикторы.** Одна из широко используемых кодировок – так называемая “binary encoding” [15]. В ней последовательность кодируется числами 0 и 1. В простейшем случае так называемого “one-hot encoding” каждую позицию в аминокислотной последовательности описывает бинарный вектор (0 и 1) размерности 20. Такой вектор содержит только одну единицу в позиции (поэтому он назван one-hot), соответствующей рассматриваемой аминокислоте, тогда как все остальные заполнены нулями. Из-за большой размерности и разреженности представления в “one-hot encoding” используют также вырожденные представления. Вырождение состоит в том, что 20 стандартных аминокислот распределяются по группам. Например, M. Dayhoff [16] опираясь на статистику точечных мутаций, распределил стандартные аминокислоты по шести группам: [H, R, K], [D, E, N, Q], [C], [S, T, P, A, G], [M, I, L, V], [F, Y, W].

Предлагаемый нами метод обладает возможностью генерировать предикторы для заранее выбранных распределений аминокислот по группам для более общей ситуации: когда рассматривают не отдельные аминокислоты, а некоторый участок последовательности (так называемое word, или “слово”) заданной длины, при этом для каждого положения этого участка определяют свои правила разделения на группы. Эта задача представляет задание так называемых регулярных выражений. В качестве примера в табл. 1 приведен вариант распределения по группам для тетрапептидов. Такой способ уменьшения размерности задачи называют “заданием редуцированных алфавитов” [14].

В положении левее на 2 остатка от текущего (“-2” – сдвиг) положения все аминокислоты, кроме глицина и пролина, считают неразличимыми. В положении “-1” предполагают эквивалентность H, K, R, затем D, E, N, Q и т.д. В текущем положении “0” все аминокислоты различаются. В табл. 1 приведена также виртуальная аминокислота “O”, которая нужна для представления фрагментов, находящихся на концах последовательности. Например, слово “OOMA” означает виртуальный тетрапептид, содержащий две ами-

нокислоты, М и А, на N-конце белка. Чтобы описать положение последовательности по правилам, приведенным в табл. 1, необходимо  $4 \times 6 \times 20 \times 4 = 1920$  предикторов. Это все еще разреженный массив, однако он существенно отличается по размеру от one-hot массива, который состоял бы из всех возможных тетрапептидов и имел размер  $20^4 = 160000$ . Более сжатый характер массива и его существенно меньший размер позволяют использовать статистические методы для связи последовательности со структурой и функциями белка. В частности, в следующем разделе описано использование предикторов, построенных на статистических характеристиках встречаемости структурных элементов.

**Предикторы, построенные на статистических характеристиках встречаемости структурных элементов.** Создание базы данных структурных элементов. В качестве примера рассмотрим применение статистики встречаемости структурных элементов – 16 пентапептидов из работы А. de Brevern и др. [17]. Этот набор структурных фрагментов (Protein Blocks – PB) широко используют для более детального описания конформации белковой цепи (по сравнению с общепринятой разметкой вторичной структуры DSSP [18]). Расстояние между фрагментами структуры главной цепи в терминах RMSD описывает формула:

$$\text{RMSD}(V_1, V_2) = \min \left( \sqrt{\frac{\sum_{i=1}^{3M} (x_{i,1} - x_{i,2})^2 + (y_{i,1} - y_{i,2})^2 + (z_{i,1} - z_{i,2})^2}{3M}} \right), \quad (2)$$

где  $V_1$  и  $V_2$  – фрагменты структуры главной цепи белка, а  $x_{i,1} = x_i(V_1), y_{i,1} = y_i(V_1), z_{i,1} = z_i(V_1), i = 1, \dots, 3M$  – декартовы координаты атомов главной цепи N, C $_{\alpha}$ , C, а  $M$  – число остатков во фрагменте.

Рассмотрим один из этих структурных элементов:  $PB_j, j \in \{1, \dots, 16\}$  – и произвольный фрагмент последовательности длиной 5 остатков.  $N_{\text{occ}}(seq)$  – частота встречаемости последовательности  $seq$

среди последовательностей с известной структурой (например, обучающая выборка). Пусть  $\bar{\mu}_j(seq)$  – среднее расстояние (в терминах RMSD) между атомами главной цепи пентапептидов, образующими последовательность  $seq$ , и атомами главной цепи  $j$ -го блока  $PB_j$ . Далее,  $\bar{\mu}_j$  – среднее расстояние между  $PB_j$  и всеми пентапептидами обучающей выборки. Соответственно  $s_j^2(seq)$  и  $s_j^2$  – дисперсии расстояния для пентапептидов, имеющих последовательность  $seq$  и для всех пентапептидов выборки  $t_j = \frac{\bar{\mu}_j - \bar{\mu}_j(seq)}{s_j(seq)}$ , где  $N$  – размер

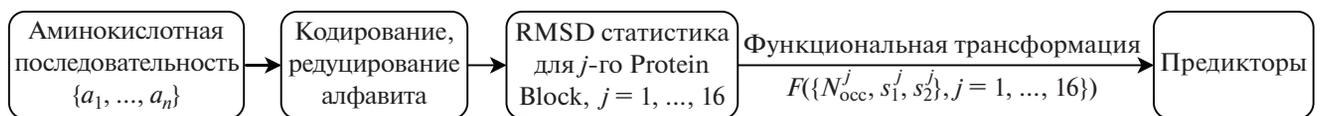
обучающей выборки, а  $s_j(seq) = \frac{\sigma^2(seq)}{N_{\text{occ}}(seq)} + \frac{s_j^2}{N}$ .

Поскольку  $N \gg N_{\text{occ}} > 1$ ,  $s_j^2(seq) = \frac{\sigma^2(seq)}{N_{\text{occ}}(seq)}$ . Следовательно,

$$t_j \approx \frac{\bar{\mu}_j - \bar{\mu}_j(seq)}{\sigma_j(seq)} \sqrt{N_{\text{occ}}}, \quad (3)$$

где  $\sigma_j^2(seq)$  – оценка дисперсии расстояния. Формула (3) определяет статистику Стьюдента для задачи сравнения выборочных средних. По значению этой величины мы оцениваем вероятность того, что среднее расстояние до  $j$ -го блока PB для пентапептидов всей обучающей выборки больше, чем для выборки пентапептидов с последовательностью  $seq$ . Из формулы (3) следует, что число вхождений  $N_{\text{occ}}(seq)$  фрагмента  $seq$  оказывает существенное влияние на результат, поскольку малые значения  $N_{\text{occ}}(seq)$  могут приводить к искаженной оценке  $\sigma_j^2(seq)$ . Именно с этим обстоятельством связано использование редуцированных алфавитов, описанных выше.

В общем виде схему, описывающую формирование предикторов, построенных на статистических характеристиках встречаемости структурных элементов, можно представить следующим образом:



Чтобы применить формулу (3), необходимо составить базу данных встречаемости белковых структур для получения величин  $\bar{\mu}_j, \bar{\mu}_j(seq), N_{\text{occ}}(seq), \sigma(seq)$ .

Функциональная трансформация, задаваемая формулой (3), не единственное преобразование, которое можно использовать для генерации пре-

дикторов, связанных со встречаемостью структурных элементов, но для всех таких предикторов нужно хранить именно указанный выше набор величин. Приведенные выше параметры сохраняют в базах данных свойств структурных элементов. При построении предсказательной модели можно использовать множество таких баз данных, раз-

личающихся редуцированными алфавитами, длиной фрагмента последовательности и метрикой, задающей расстояние между фрагментами. Например, вместо метрики RMSD (root mean square deviation) можно использовать RMSDA (root mean square deviation of angular values) [17], либо другую метрику.

**Комбинированные предикторы.** В некоторых случаях может возникнуть необходимость использования предикторов, учитывающих одновременно несколько физико-химических или функциональных свойств белковой последовательности.

Комбинированные предикторы представляют дополнительные возможности для создания предикторов сложного вида. Они могут быть составлены по заранее заданному сценарию из нескольких описанных выше предикторов и условий. Чтобы реализовать эту возможность, каждому предиктору присваивается свое имя, по которому к нему обращаются для задания комбинированного предиктора. Простейший случай – перемножение двух или нескольких предикторов. Для логических переменных это означает одновременное выполнение условий, определяемых для каждого из предикторов. В качестве примера рассмотрим три булевых предиктора:

1. Присутствие двух алифатических аминокислот на расстоянии от  $-6$  до  $-2$  от текущего положения.
2. Присутствие двух алифатических аминокислот на расстоянии от  $+3$  до  $+6$  от текущего положения.
3. Присутствие как пролина, так и глицина на расстоянии от  $-2$  до  $+3$ .

Логическим произведением этих трех предикторов будет комбинированный предиктор, который устойчив к единичным вставкам и делециям. Комбинированные предикторы, составленные подобным образом, мы использовали для описания и последующего поиска  $\beta$ -шпилек.

В других случаях можно задавать различные логические условия, например результирующий предиктор, составленный из двух предикторов, который равен первому предиктору, если второй лежит в заданных пределах, в противном случае равен нулю.

Также возможно задание вложенных процедур генерации предикторов, в которых используют комбинированные предикторы, составленные на предыдущих шагах.

**Расчетные методы для построения предсказательной модели.** Чтобы оценить качество сформированной системы предикторов и для ее усовершенствования, предложено использование оригинальных модификаций регрессионного и дискриминантного анализов.

**Пошаговый регрессионный анализ.** Линейный регрессионный анализ [19] для задач предсказания структуры и функций белка нуждается в модификации, которая позволяет добавлять (для улучшения качества предсказания) или удалять (для упрощения регрессионной функции) предикторы. Пошаговая регрессия позволяет частично автоматизировать процедуру получения набора регрессионных функций без существенных дополнительных вычислительных затрат. Предикторы, включенные в итоговую регрессионную функцию, должны удовлетворять определенным критериям. Мы использовали подход, основанный на статистике Фишера (F-статистики) [20], для проверки значимости регрессионного коэффициента предиктора при принятии решения о включении или исключении его из регрессионной модели. Состав предикторов, включенных в итоговую регрессионную модель, зависит от пороговых значений F-статистики исключения и F-статистики включения. Следовательно, пороговые значения F-статистики влияют на состав и, таким образом, на качество предсказательной модели. Схема используемого в нашей методике пошагового регрессионного анализа изображена на рис. 1а.

Реализованный в алгоритме (и программном пакете) пошаговый регрессионный анализ обладает следующими возможностями:

1. *Определение оптимального порогового значения F-статистики.* Пороговые значения F-статистики и оптимальный набор предикторов выбирают с помощью метода перекрестной проверки (*k-fold cross validation*). Иначе говоря, для задачи, содержащей  $N$  объектов,  $N$  раз создают регрессионную модель, на каждом шаге из нее исключают один из объектов и предсказывают значение зависимой переменной для этого объекта. Итог этой процедуры – массивы предсказанных и экспериментальных значений зависимых переменных. Коэффициент корреляции Пирсона между этими массивами выбран мерой качества модели. Повторяя эту процедуру с различными пороговыми значениями F-статистики, мы выбирали ту модель, для которой коэффициент корреляции максимален. Анализ состава предикторов в окончательной модели и сравнения их статистической значимости дает ценную информацию о факторах, определяющих локальную структуру.

2. *Возможность одновременно рассчитывать не одну зависимую переменную.* В задаче предсказания локальной структуры начальный набор предикторов (до отбора удовлетворяющих пороговым значениям F-статистики) одинаков для каждой из зависимых переменных (например, для шаблонов локальной структуры, таких как  $\alpha$ -спираль,  $\beta$ -слой и т.д.). Подобная ситуация может встречаться во многих задачах. Самым затратным в смысле числа вычислительных опера-



**Рис. 1.** Схемы формирования статистической модели для предсказания: *а* – локальной структуры (LS) по аминокислотной последовательности белка; *б* – вторичной структуры белка с использованием пошагового дискриминационного анализа.

ций (требующих на 2 порядка больше операций в случае реальных задач, чем все остальные) является этап вычисления матрицы перекрестных произведений, пропорциональный  $M \times (M + 6)N$  [20], где  $M$  – число предикторов в начальном наборе,  $N$  – число объектов. Мы модифицировали стандартный метод, вычисляя расширенную матрицу перекрестных произведений размером  $(M + P)(M + P)$  один раз. Впоследствии эту матрицу используют для получения регрессионных функций для всех  $P$  зависимых переменных. Эта модификация обеспечивает радикальное повышение производительности.

**Пошаговый дискриминантный анализ.** Применение пошагового дискриминантного анализа [20] имеет много общего с применением пошагового регрессионного. В этом случае с помощью  $F$ -статистики также определяют набор статистически значимых предикторов, но не для предсказания зависимой переменной, а для класси-

фикации наблюдаемых объектов по группам. Алгоритмы формирования значимых предикторов в стандартном пошаговом методе дискриминантного анализа и пошаговом регрессионном анализе практически идентичны. Так же, как и в регрессионном анализе, происходит определение оптимального набора предикторов с помощью метода перекрестной проверки ( $k$ -fold cross validation). В качестве меры качества предсказательной модели выбрана доля правильно классифицированных (предсказанных) объектов. Схема используемого в нашей методике пошагового дискриминационного анализа изображена на рис. 1б.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Описанные в предыдущих разделах подходы реализованы в виде пакета программ для генерации и тестирования наборов предикторов описанными выше статистическими методами.

Иллюстрация применения нашего пакета программ к реальным задачам представляет обращение к классической задаче предсказания вторичной структуры белка. При этом использовали все указанные выше типы предикторов. В частности, мы выполняли создание и отладку предсказательной модели с применением вычислительных процедур, которые обеспечивают оценку вклада отдельных предикторов. Отладка модели – итеративная процедура: на каждой итерации строят предсказательную модель, затем редактируют набор предикторов, исходя из статистической значимости каждого предиктора и качества полученной модели. При создании модели использовали последовательно пошаговые регрессионный и дискриминантный анализы. Технически современные компьютеры позволяют задавать большое число предикторов для обучения модели (до 10000 и более). Процедура выявления значимых предикторов позволяет шаг за шагом выявлять имеющиеся закономерности, модифицировать и комбинировать предикторы, улучшая при этом модель.

Создание модели предсказания состоит из двух этапов. На первом этапе использована регрессионная модель для предсказания классификации по обобщенным координатам для 16 базовых структур (protein blocks) [14] (рис. 1а), затем эти обобщенные координаты использовали как предикторы (вместе с дополнительными предикторами) для построения модели с помощью пошагового дискриминантного анализа (рис. 1б).

Проиллюстрируем первый этап создания предсказательной модели на примере двух наиболее важных наборов предикторов. Первый из них основан исключительно на физико-химических свойствах аминокислот из базы данных AAindex, использующихся для функционального преобразования по формуле (1). В процессе отладки таких предикторов опробованы не только все 30 шкал гидрофобности из этой базы, но и все остальные свойства из этой базы. Оказалось, что наиболее значимы предикторы, составленные по шкале WERD780101, которая формально не относится к гидрофобности, а характеризует склонность аминокислот быть внутри глобулы белка [21]. При этом среди всех опробованных значений периода выявлен период, равный 3.6 остаткам, что соответствует  $\alpha$ -спиральной конформации. Оптимальная длина фрагмента, на которой этот предиктор наиболее статистически весом, составляет 8 остатков. На рис. 2 видно, что гидрофильные аминокислоты ориентированы к водной поверхности белка через 3–4 остатка.

Второй тип предикторов менее очевиден, однако для предсказания многих из 16 структурных классов де Бреверна (de Brevern) [17] предикторы этого типа обладают наибольшей статистической

значимостью. На примере этого типа предикторов мы иллюстрируем поиск адекватного функционального преобразования, отражающего скрытые закономерности. При этом мы исходили из предположения, что если для всех фрагментов (в данном случае пентапептидов), имеющих последовательность *seq*, характерна конформация, близкая к одной из 16 шаблонных структур де Бреверна, то среднее расстояние (RMSD) между этими фрагментами и этой шаблонной структурой будет меньше, чем среднее расстояние между этой же шаблонной структурой и фрагментами всей выборки. Это расстояние может быть записано так:

$$F(k, seq) = \frac{\sum_i^N R_k}{N} - \frac{\sum_i^{N_{occ}(seq)} R_k(seq)}{N_{occ}(seq)}, \quad (4)$$

где *k* – номер шаблонной структуры (protein block), *N* – общее число фрагментов обучающей выборки, *N<sub>occ</sub>(seq)* – число фрагментов с последовательностью *seq*, *R<sub>i</sub>(k)* – расстояние от *i*-го фрагмента до *k*-го protein block, *R<sub>i</sub>(k, seq)* – расстояние от *i*-го фрагмента с последовательностью *seq* до *k*-го protein block. Максимум выражения (4) соответствует максимальной близости фрагментов с последовательностью *seq* к *k*-й шаблонной структуре (protein block) де Бреверна.

Но как можно сравнивать случаи, когда *N<sub>occ</sub>(seq) = 1* и когда *N<sub>occ</sub>(seq) = 100*? Как добиться того, чтобы вклад этого предиктора в регрессионную функцию был линейным? Выполнили тестирование различных функциональных преобразований и в итоге предикторы, задаваемые формулой (5), стали наиболее статистически значимыми:

$$F(k, seq) = \ln[1 + \ln(1 + N_{occ}(seq))] \times \left( \frac{\sum_i^N R_i(k)}{N} - \frac{\sum_i^{N_{occ}(seq)} R_i(k, seq)}{N_{occ}(seq)} \right). \quad (5)$$

Алфавит не обязательно имеет длину 5 остатков, совпадающую с длиной шаблонных структур де Бреверна. В данной задаче в процессе отладки модели мы подбирали редуцированные алфавиты длиной до 15 остатков, используя различные способы распределения аминокислот по эквивалентным группам. В частности, самый значимый предиктор для предсказания конформации  $\beta$ -слоя (protein block “D” де Бреверна) получен по формуле (5) для случая редуцированного алфавита для фрагмента длиной 7 остатков (табл. 2).

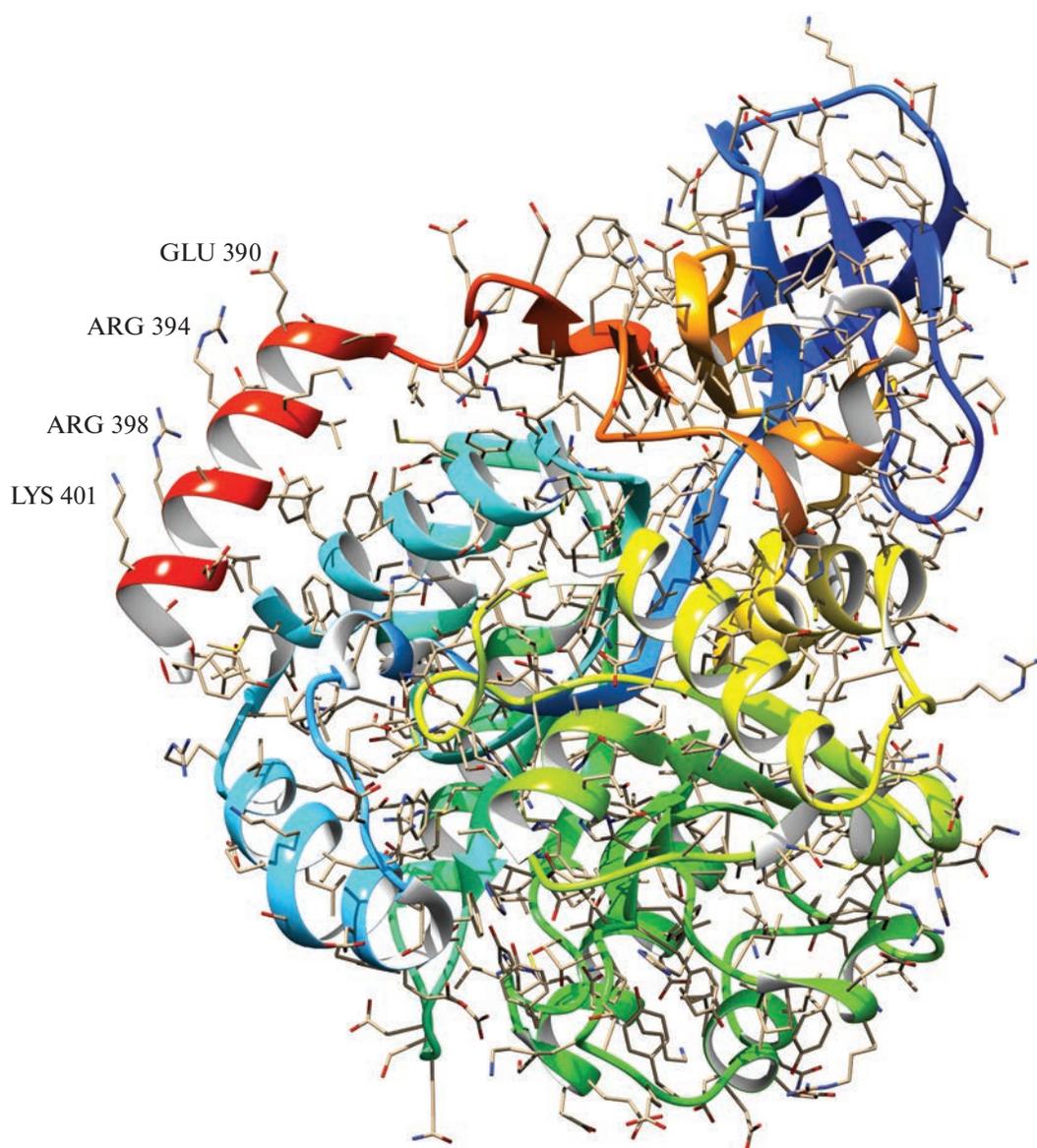


Рис. 2. Формирование предиктора, связанного с периодическими вариациями гидрофобности, вдоль цепи в белке 1P1M.

#### *Обучающая выборка*

Мы использовали стандартный метод создания обучающих выборок PISCES30 [22]. Отбор белковых цепей из PDB выполнен по следующим параметрам: разрешение не хуже 2.5 Å, R-фактор не менее 1.0 и гомология последовательностей не более 30%. В итоге получена обучающая выборка, состоящая из 17148 негомологичных белковых цепей, структуры которых определены с помощью рентгеноструктурного анализа.

#### *Предсказание вторичной структуры*

Обучающая выборка для оценки эффективности модели предсказания разделена на две части в соотношении 4 : 1. Большую часть использовали

для обучения, меньшую — для оценки качества предсказания.

Предсказание выполняли для трех DSSP классов ( $\alpha$ -спираль “Н”,  $\beta$ -слой “Е”, неупорядоченная структура “—”). Качество предсказательной модели можно оценить при помощи матрицы ошибок (confusion matrix), диагональные элементы которой содержат число правильно предсказанных конформаций соответствующего класса, а элементы каждой строки характеризуют распределение предсказаний по классам. В табл. 3 приведены результаты, полученные для контрольной выборки. В частности, класс Н ( $\alpha$ -спираль) встретился 307026 раз (281 863 + 5 + 25 158). Из них этот класс верно предсказали 281 863 раз, 5 раз

**Таблица 2.** Распределения аминокислот по группам для фрагмента длиной 7 остатков

Сдвиг	Распределение аминокислот по группам эквивалентности
–3	[A] [R] [N] [D] [C] [Q] [E] [H] [I] [L] [K] [M] [F] [P] [S] [T] [W] [Y] [V] [G] [P]
–2	[A] [R] [N] [D] [C] [Q] [E] [H] [I] [L] [K] [M] [F] [P] [S] [T] [W] [Y] [V] [G] [P]
–1	[ALMC] [VIFYWKRHDENQST] [GP] [O] [X]
0	[ALMC] [VIFYWKRHDENQST] [GP] [X]
1	[ALMC] [VIFYWKRHDENQST] [GP] [O] [X]
2	[A] [R] [N] [D] [C] [Q] [E] [H] [I] [L] [K] [M] [F] [P] [S] [T] [W] [Y] [V] [G] [P]
3	[A] [R] [N] [D] [C] [Q] [E] [H] [I] [L] [K] [M] [F] [P] [S] [T] [W] [Y] [V] [G] [P]

**Таблица 3.** Матрица ошибок для трех классов DSSP

DSSP-класс	Содержание ошибок					
	число			%		
	Н	Е	–	Н	Е	–
Н	281863	5	25158	91.80	0.001	8.19
Е	212	139392	42698	0.12	76.46	23.42
–	26264	44671	239587	8.46	14.39	77.16

ошибочно как  $\beta$ -слой и 25 158 раз неверно как неупорядоченную структуру.

Чтобы дополнительно проконтролировать качество предсказательной модели, использовали классическую выборку пептидов CB513 [23], которую широко применяли другие авторы для анализа эффективности методов предсказания и сравнения этих методов между собой. Выборка представляет собой набор 513 белковых цепей (и частей белковой цепи), подобранных специально для перекрестной проверки методов предсказания вторичной структуры так, чтобы минимизировать влияние внутренней гомологии. Также мы выполнили сравнение точности предсказаний нашего метода с другими современными методами предсказания структур, такими как GApred [24], SPIDER2 [25], Jpred4 [26], FSVM [27], SSpro5 [28]. Результаты предсказания, выполненные предлагаемым нами подходом, и их сравнение с другими методами (данные других программ получены из работы [24]) приведены в табл. 4.

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученное качество предсказания для трех DSSP-классов составляет  $Q3 = 82.62\%$  для контрольной подвыборки из нашей обучающей выборки и  $81.44\%$  для выборки CB513 [23]. Этот результат говорит о высоком качестве предсказания. Лучшие, так называемые state-of-the-art, методы предсказания обеспечивают качество  $Q3$  в диапазоне  $80–85\%$  [29–32], хотя в этой работе предсказание выполняли без использования методов машинного обучения и данных о гомологичных белках (фактически, результат можно назвать

**Таблица 4.** Сравнение точности предсказаний трех DSSP-классов, выполненное для выборки CB513 различными методами предсказания

Метод	DSSP-класс			
	Q3	Н	Е	–
GApred	85.3	83.3	85.5	84.8
SPIDER2	73.5	72.4	79.3	78.7
Jpred4	65.6	60.2	63.4	67.1
FSVM	83.0	82.0	80.0	79.0
SSpro5	82.0	78.3	83.0	80.5
Наш метод	81.4	91.2	76.3	76.2

предварительным для последующего применения методов машинного обучения).

Очевидно (табл. 3 и 4), что наилучшая точность предсказаний ( $91\%$ ) получена для DSSP-класса Н ( $\alpha$ -спираль). Это связано с тем, что класс Н наиболее структурно однороден и почти полностью совпадает с предсказанием класса М из структурной классификации де Бреверна [17]. При этом ошибочные ложноотрицательные предсказания ( $8\%$ ) относятся к классу неупорядоченных структур (обозначаемых “–” в разметке 3 классов DSSP). Внутри этого класса, согласно структурной классификации DSSP, имеется класс Т (turn), который структурно близок  $\alpha$ -спирали. Этим фактором объясняют ложноположительные предсказания ( $8\%$ ) для класса Н со стороны класса “–”. В нашей модели предсказания  $\alpha$ -спирального класса и класса Е ( $\beta$ -слой) практически не пересекаются. Это вполне ожидаемо, поскольку эти классы максимально структурно различаются (в смысле RMSD между пентапептидами), а пред-

сказания DSSP-классов основаны на предсказании распределения по 16 структурным блокам де Бреверна. Отметим, что расстояние между структурными блоками M, соответствующими идеальной  $\alpha$ -спирали, и D, близкими к антипараллельному  $\beta$ -слою, составляет 3.12 Å и близко к максимальному среди попарных расстояний для 16 структурных классов [14].

Точность предсказания DSSP класса E ( $\beta$ -слой) (76%) заметно ниже класса H ( $\alpha$ -спираль) ввиду того, что среди класса неупорядоченных структур присутствуют конформации, близкие (в смысле RMSD) к предсказываемому классу E. Это приводит к значительным ложноположительным (23%) и ложноотрицательным (14%) ошибкам.

Структурную классификацию DSSP определяют преимущественно по водородным связям белковых цепей, поэтому такие DSSP-классы, как  $\beta$ -слои, имеют размытую структуру. Даже для случая 8 DSSP-классов эта проблема остается актуальной. Чтобы описать более детально локальную структуру, создали структурные классификации, основанные на других принципах. В частности, структурная классификация де Бреверна основана на фиксированных углах  $\Phi$  и  $\Psi$  для каждого из 16 структурных блоков. Точность предсказания, основанного на физико-химических свойствах аминокислот, в этом случае составила  $Q_{16} = 67.9\%$  [14], при этом для  $\alpha$ -спирально-структурного блока “M” точность предсказания превысила 97%. Поскольку для предсказания мы не использовали множественные выравнивания последовательностей, эта точность, по-видимому, определена тем, что удалось найти существенные физико-химические параметры, определяющие структуру этого структурного блока. Эти параметры заданы в виде предикторов, созданных и отобранных с помощью предлагаемых нами алгоритмов создания предикторов и отладки предсказывающей модели. Многие предикторы удалось интерпретировать (например, предикторы, образованные по формуле (1)).

Как видно из табл. 4, на выборке CB513 [23] наш метод в его текущей реализации существенно лучше предсказывает  $\alpha$ -спирали (превосходя на 8% даже “генетические” алгоритмы и на 9% и более другие state-of-the-art методы) и проявил себя несколько хуже в предсказании  $\beta$ -слоев (уступая 9% “генетическому” алгоритму и отставая от наиболее передовых state-of-the-art методов до 7%) и неупорядоченных структур (отставание от “генетического” алгоритма – 9%, от наиболее передовых state-of-the-art методов – 4%). Это может быть связано с тем, что наше предсказание  $\alpha$ -спирали проще связать именно со структурными параметрами, поскольку структура  $\alpha$ -спирали задана более формализовано, чем структура  $\beta$ -слоев и тем более неупорядоченных структур.

Кроме того,  $\beta$ -слои структурно близки к некоторым участкам, относимым к неупорядоченным (например, левая спираль типа полипролин II). Практически одинаковые пространственные структуры, близкие к конформации  $\beta$ -слоя, могут быть отнесены к разным DSSP-классам только по наличию/отсутствию характерных водородных связей, некоторые из которых вообще могут быть сформированы разными белковыми цепями. Информацию для отнесения таких спорных участков к одному из классов, по-видимому, можно получить исключительно из известной структуры гомологичных белков после их выравнивания по последовательности. В предлагаемом методе мы не использовали выравнивание последовательностей, поскольку занимались иллюстрацией возможностей метода генерации и отладка набора предикторов.

## ЗАКЛЮЧЕНИЕ

Предложена методика формирования и усовершенствования наборов предикторов для последующего использования в задачах предсказания структуры и функций белка методами машинного обучения. Представлены примеры формирования наборов предикторов разных типов. Приведены примеры, иллюстрирующие работу пакета программ. Описана процедура задания пользователем собственных предикторов по аналогии с уже присутствующими в библиотеке. Представленная процедура усовершенствования предсказательной модели с помощью регрессионного и дискриминантного анализов – мощный инструмент поиска и тестирования наличия скрытых закономерностей, заложенных в исходных данных. Полученные с помощью машинного обучения результаты зачастую весьма сложно интерпретировать, в то время как предложенный метод позволяет выявить физико-химические факторы, определяющие структуру белка.

Генерацию предикторов и улучшение модели с применением классических статистических методов мы рассматриваем как подготовительную процедуру для современных методов машинного обучения, хотя при удачном выборе и последующей оптимизации набора входных предикторов использование классических методов построения предсказательных моделей обеспечивает вполне конкурентоспособные результаты.

Применение предложенной методики создания предикторов для предсказания структуры и функций белка методами машинного обучения будет продолжено нами в последующих работах. Кроме применения методов машинного обучения мы планируем подключить информацию о структуре гомологичных белков (template-based method) и создать алгоритм переключения между template-based method и non-template-based method

в зависимости от предсказываемого белка. Показано, что подобный подход приводит к повышению точности на 3–6% [33]. Весьма перспективным представляется применение изложенных в работе методов для предсказания более детальных, чем вторичная структура, структурных классификаций, например по структурным блокам де Бреверна [17] или еще более детальных классификаций с различным числом элементов и различной длиной фрагментов, которые мы специально разрабатываем [14].

В текущем виде предложенный метод и программное обеспечение могут быть использованы в составе комбинированного подхода, сочетающего использование различных методов/программ предсказания для первичного предсказания координат  $\alpha$ -спиральных участков в предсказываемых пептидах, поскольку по точности предсказания таких участков предложенный нами метод и его программная реализация существенно превосходят все существующие аналоги.

Авторы благодарны Центру высокоточного редактирования и генетических технологий для биомедицины за вычислительные мощности, предоставленные для выполнения этой работы.

Работа выполнена при поддержке Российского научного фонда (22-24-01088).

Настоящая статья не содержит каких-либо исследований с участием людей или животных в качестве объектов исследований.

Авторы заявляют об отсутствии конфликта интересов.

#### ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Реализация предложенного метода в виде пакета кроссплатформенных программ на языке C++ доступна по адресу <https://github.com/Milchevskiy/protein-encoding-projects>.

#### СПИСОК ЛИТЕРАТУРЫ

1. Anfinsen C.B. (1973) Principles that govern the folding of protein chains. *Science*. **181**, 223–230.
2. Yang Y., Gao J., Wang J., Heffernan R., Hanson J., Palliwala K., Zhou Y. (2018) Sixty-five years of the long march in protein secondary structure prediction: the final stretch? *Brief. Bioinform.* **19**, 482–494.
3. Zimmermann O., Hansmann U.H. (2008) LOCUSTRA: accurate prediction of local protein structure using a two-layer support vector machine approach. *J. Chem. Inf. Model.* **48**, 1903–1908.
4. Wuyun Q., Zheng W., Peng Z., Yang J. (2018) A large-scale comparative assessment of methods for residue-residue contact prediction. *Brief. Bioinform.* **19**, 219–230.
5. Zhang J., Kurgan L. (2018) Review and comparative assessment of sequence-based predictors of protein-binding residues. *Brief Bioinform.* **19**, 821–837.
6. Min S., Lee B., Yoon S. (2017) Deep learning in bioinformatics. *Brief. Bioinform.* **18**, 851–869.
7. Hu H.J., Pan Y., Harrison R., Tai P.C. (2004) Improved protein secondary structure prediction using support vector machine with a new encoding scheme and an advanced tertiary classifier. *IEEE Trans Nanobiotechnology*. **3**, 265–271.
8. Yoo P.D., Sikder A.R., Zhou B.B., Zomaya A.Y. (2008) Improved general regression network for protein domain boundary prediction. *BMC Bioinformatics*. **9**(Suppl. 1), S12.
9. Miyazawa S., Jernigan R.L. (1999) Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. *Proteins*. **34**, 49–68.
10. Lin K., May A.C., Taylor W.R. (2002) Amino acid encoding schemes from protein structure alignments: multi-dimensional vectors to describe residue types. *J. Theor. Biol.* **216**, 361–365.
11. Asgari E., Mofrad M.R. (2015) Continuous distributed representation of biological sequences for deep proteomics and genomics. *PLoS One*. **10**, e0141287.
12. Jing X., Dong Q., Hong D., Lu R. (2020) Amino acid encoding methods for protein sequences: a comprehensive review and assessment. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **17**, 1918–1931.
13. Kawashima S., Pokarowski P., Pokarowska M., Kolinski A., Katayama T., Kanehisa M. (2008) AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res.* **36**, D202–205.
14. Milchevskaya V., Nikitin A.M., Lukshin S.A., Filatov I.V., Kravatsky Y.V., Tumanyan V.G., Esipova N.G., Milchevskiy Y.V. (2021) Structural coordinates: a novel approach to predict protein backbone conformation. *PLoS One*. **16**, e0239793.
15. Taha K., Yoo P.D. (2015) Predicting protein function from biomedical text. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2015**, 3275–3278.
16. Dayhoff M.O. (1972) *Atlas of protein sequence and structure*. Silver Spring, Md.: National Biomedical Research Foundation.
17. de Brevern A.G., Etchebest C., Hazout S. (2000) Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins*. **41**, 271–287.
18. Kabsch W., Sander C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. **22**, 2577–2637.
19. Hocking R.R. (1983) Developments in linear regression methodology: 1959–1982. *Technometrics*. **25**, 219–223.
20. Ralston A., Wilf H.S., Enslein K. (1960) *Mathematical methods for digital computers*. New York: Wiley.
21. Wertz D.H., Scheraga H.A. (1978) Influence of water on protein structure. An analysis of the preferences of amino acid residues for the inside or outside and for specific conformations in a protein molecule. *Macromolecules*. **11**, 9–15.
22. Wang G., Dunbrack R.L., Jr. (2005) PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res.* **33**, W94–98.

23. Cuff J.A., Barton G.J. (1999) Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins*. **34**, 508–519.
24. Rout S.B., Mishra S., Sahoo S.K. (2021) Q3 Accuracy and SOV measure analysis of application of GA in protein secondary structure prediction. *Revue d'Intelligence Artificielle*. **35**, 403–408.
25. Yang Y., Heffernan R., Paliwal K., Lyons J., Dehzangi A., Sharma A., Wang J., Sattar A., Zhou Y. (2017) SPIDER2: a package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. *Methods Mol. Biol.* **1484**, 55–63.
26. Drozdetskiy A., Cole C., Procter J., Barton G.J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–394.
27. Xie S., Li Z., Hu H. (2018) Protein secondary structure prediction based on the fuzzy support vector machine with the hyperplane optimization. *Gene*. **642**, 74–83.
28. Magnan C.N., Baldi P. (2014) SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics*. **30**, 2592–2597.
29. Ma Y., Liu Y., Cheng J. (2018) Protein secondary structure prediction based on data partition and semi-random subspace method. *Sci. Rep.* **8**, 9856.
30. Guo Z., Hou J., Cheng J. (2021) DNSS2: improved ab initio protein secondary structure prediction using advanced deep learning architectures. *Proteins*. **89**, 207–217.
31. Wang S., Peng J., Ma J., Xu J. (2016) Protein secondary structure prediction using deep convolutional neural fields. *Sci. Rep.* **6**, 18962.
32. Zhang B., Li J., Lu Q. (2018) Prediction of 8-state protein secondary structures by a novel deep learning architecture. *BMC Bioinformatics*. **19**, 293.
33. Krieger S., Kececioglu J. (2020) Boosting the accuracy of protein secondary structure prediction through nearest neighbor search and method hybridization. *Bioinformatics*. **36**, i317–i325.

## A Method to Generate Complex Predictive Features for ML-Based Prediction of the Local Protein Structure

Y. V. Milchevskiy<sup>1, \*</sup>, V. Y. Milchevskaya<sup>1, 2</sup>, and Y. V. Kravatsky<sup>1, 3</sup>

<sup>1</sup>Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia

<sup>2</sup>Institute of Medical Statistics and Bioinformatics, Faculty of Medicine, University of Cologne, Cologne, 50931 Germany

<sup>3</sup>Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia

\*e-mail: milch@imb.ru

Recently, the prediction of protein structure and function from its sequence underwent a rapid increase in performance. It is primarily due to the application of machine learning methods, many of which rely on the predictive features supplied to them. It is thus crucial to retrieve the information encoded in the amino acid sequence of a protein. Here, we propose a method to generate a set of complex yet interpretable predictors, which aids in revealing factors that influence protein conformation. The proposed method allows us to generate predictive features and test them for significance in two scenarios: for a general description of the protein structures and functions, as well as for highly specific predictive tasks. Having generated an exhaustive set of predictors, we narrow it down to a smaller curated set of informative features using feature selection methods, which increases the performance of subsequent predictive modelling. We illustrate the effectiveness of the proposed methodology by applying it in the context of local protein structure prediction, where the rate of correct prediction for DSSP Q3 (three-class classification) is 81.3%. The method is implemented in C++ for command line use and can be run on any operating system. The source code is released on GitHub: <https://github.com/Milchevskiy/protein-encoding-projects>.

**Keywords:** local structure prediction, protein secondary structure prediction, protein function, protein sequence encoding, protein conformation, stepwise regression, stepwise discriminant analysis

УДК 575.1/2:577.1

## ВЛИЯНИЕ НОКАУТА ОСНОВНЫХ ГЕНОВ ТРАНССУЛЬФУРАЦИИ НА ПАТТЕРН БЕЛКОВОГО СИНТЕЗА У *D. melanogaster*

© 2023 г. А. Ш. Заклута<sup>а</sup>, В. Ю. Шилова<sup>а</sup>, О. Г. Зацепина<sup>а</sup>, \*<sup>а</sup>Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991 Россия

\*e-mail: olzacepina@yandex.ru

Поступила в редакцию 01.07.2022 г.

После доработки 21.07.2022 г.

Принята к публикации 21.07.2022 г.

Ферменты пути транссульфурации и продукции сероводорода – цистатионин-β-синтаза (CBS), цистатионин-γ-лиаза (CSE) и 3-меркаптопируват-сульфотрансфераза (3-MST) – играют важную цитопротекторную роль в организме. Ранее с помощью технологии CRISPR/Cas9 мы получили линии *Drosophila melanogaster* с делециями генов *cbs*, *cse*, *mst*, а также с двойной делецией генов *cbs* и *cse*. Нами проанализировано влияние этих делеций на паттерн белкового синтеза в слюнных железах личинок третьего возраста и в яичниках половозрелых мух. В слюнных железах линий с делециями *cbs* и *cse* обнаружено снижение накопления белка FBP2, содержащего 20% остатков метионина. В яичниках выявлены изменения в уровне экспрессии и в точках изофокусирования белков, участвующих в защите клеток от окислительного стресса, гипоксии и в деградации белков. Показано, что степень окисленности белков в линиях мух с делециями ферментов транссульфурации сходна со степенью окисленности в контрольной линии. Обнаружено снижение общего количества протеасом и их каталитических субъединиц в линиях с делециями генов *cbs* и *cse*.

**Ключевые слова:** гены пути транссульфурации, CRISPR/Cas9-технология, протеасомы, дрозофила**DOI:** 10.31857/S0026898423010160, **EDN:** AWNKFK

### ВВЕДЕНИЕ

Сероводород (H<sub>2</sub>S), считавшийся ранее относительно токсичным газом, исследуется в настоящее время в качестве биологического медиатора, участвующего во многих физиологических процессах у всех высших организмов.

Сероводород имеет важное значение для устойчивости клеток к окислительному стрессу и влияет на продолжительность жизни [1–5]. H<sub>2</sub>S является эндогенным модулятором окислительного стресса, удаляющим активные радикалы кислорода (ROS) и азота [6], увеличивающим уровень внутриклеточного глутатиона (GSH) [7], повышающим уровень антиоксидантной защиты [6]. Ферменты цистатионин-β-синтаза (CBS), цистатионин-γ-лиаза (CSE) и 3-меркаптопируват-сульфотрансфераза (3-MST) играют центральную роль в синтезе H<sub>2</sub>S, превращая гомоцистеин в цистеин после распада метионина [8, 9] в результате реакции транссульфурации (TSP). Показано, что подавление экспрессии гена *cbs* приводит к на-

коплению гомоцистеина и развитию гомоцистинурии – мультисистемного расстройства, затрагивающего соединительную и мышечную ткань, центральную нервную систему и сердечно-сосудистую систему [10]. CBS также необходима для женской репродуктивной функции [11]. Противоречивые результаты получены при исследовании мышей со сниженной экспрессией гена *cse*. Одни авторы показали, что делеция гена *cse* или подавление его экспрессии улучшают выживаемость мышей при эндотоксическом шоке, снижая уровень воспалительной реакции [12–14]. Согласно другим данным, экспрессия *cse* играет важную роль в функционировании эндотелия и состоянии сосудов, способствуя ангиогенезу в ишемизированных участках и увеличивая проницаемость эндотелия [15, 16], а также может ингибировать прогрессию нейродегенеративных заболеваний [17].

Нами получены линии *Drosophila melanogaster* с делецией основных генов серного метаболизма [18]. Анализ транскриптома целых мух с делецией генов *cbs* и *cse*, а также двойных трансформантов выявил изменения в уровне экспрессии множества генов, вовлеченных в различные биологические процессы. Наиболее существенными были

Сокращения: CBS – цистатионин-β-синтаза, CSE – цистатионин-γ-лиаза, 3-MST – меркаптопируват-сульфотрансфераза, FBP2 – белок жирового тела 2, TSP – транссульфурация.

изменения уровня экспрессии генов, участвующих в процессах окислительного восстановления и репродукции [18].

Модельная система дрозофилы делает возможным изучение влияния делеций генов серного метаболизма на уровне целого организма. Интересно было понять, какие физиологические параметры подвергаются наибольшим изменениям при делеции этих генов. Выявлены значимые нарушения в функционировании репродуктивной системы самок. Отмечена задержка в развитии яичников и в откладке яиц у линий мух с делецией гена *cbs*, задержка в развитии на стадиях от откладки яиц до окукливания и сокращение длительности репродуктивного возраста у линий с двойной делецией генов *cbs* и *cse* [19].

В представленной работе методом двумерного электрофореза изучены паттерны белков из слюнных желез личинок третьего возраста, а также из яичников 5-дневных мух *D. melanogaster* с делецией генов *cbs*, *cse* и *mst*. В линиях с делецией генов серного метаболизма выявлены изменения в уровне экспрессии ряда белков. Исследована активность и общее количество протеасом в изучаемых линиях мух.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Содержание дрозофил.** В работе использовали лабораторные линии *D. melanogaster*, полученные из стокового центра Bloomington Drosophila Stock Center (линия 58492), а также линии с делецией генов серного метаболизма, полученные в нашей лаборатории – линия с делецией гена *cbs* (*CBS*<sup>-/-</sup>), линия с делецией гена *cse* (*CSE*<sup>-/-</sup>), линии с делециями обоих генов (*CBS*<sup>-/-</sup>; *CSE*<sup>-/-</sup> (1) и *CBS*<sup>-/-</sup>; *CSE*<sup>-/-</sup> (2)), линия с делецией гена *mst* (*MST*<sup>-/-</sup>). В эксперименте использовали девственных самок, которых содержали в течение 5 дней в пробирках со стандартным кормом при температуре +24°C, не более 20 самок в пробирке.

**Двумерный электрофорез белков (2D).** Для получения белковых проб для 2D-электрофореза из личинок третьего возраста извлекали слюнные железы (или яичники из 5-дневных самок) в фосфатно-солевом буфере (PBS). Слюнные железы метили в течение 1 ч в 200 мкл PBS, добавляя 1 мкл (1.85 MBq) *L*-<sup>35</sup>S-метионина (“Amersham Biosciences Corp.”, США). Одновременно готовили немеченые пробы. Далее железы лизировали в 75 мкл буфера О’Фаррелла с 1% PMSF и 1% коктейля ингибиторов протеаз (“Amresco”, США), растирали и центрифугировали (10 мин, 10000 g). Супернатант замораживали в жидком азоте и хранили при –70°C. 2D-электрофорез проводили согласно классическому методу О’Фаррелла [20] с некоторыми вариациями. На первое направление с изофокусом 3.5–9 pH наносили лизат из 10 пар

слюнных желез. После изофокусирования столбики геля извлекали из трубок, промывали в течение 40 мин в буфере Лэммли для подготовки к электрофорезу во втором направлении и замораживали при –80°C. Во втором направлении белки разделяли методом диск-электрофореза в 11%-ном полиакриламидном геле (ПААГ). После окрашивания гели высушивали для последующей радиоавтографии. Радиоавтографию проводили на Typhoon FLA 9500 (“GE Healthcare Life Sciences”, США). Гели окрашивали в 0.2%-ном растворе Coomassie Brilliant Blue G-250, вырезали белковые пятна, которые использовали для проведения масс-спектрометрического анализа.

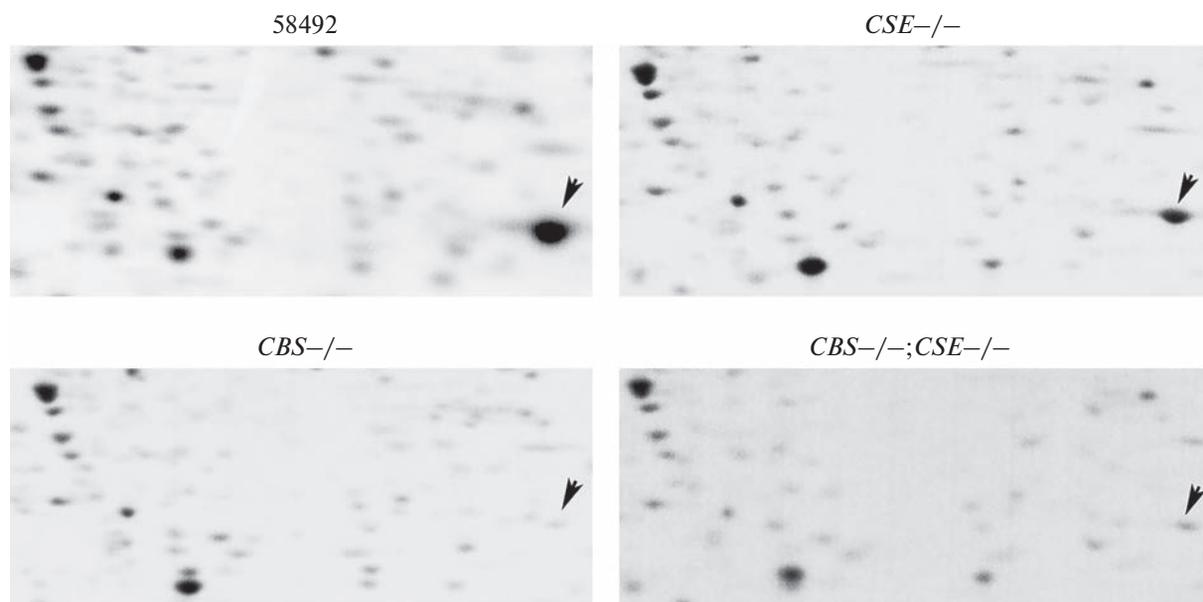
**Измерение активности протеасом.** Активность протеасом измеряли в соответствии с [21] с некоторыми изменениями. Пятидневных самок дрозофилы замораживали в жидком азоте, далее гомогенизировали в лизирующем буфере NP40 (50 mM Трис-НСl pH 7.4, 150 mM NaCl, 1% NP40). Концентрацию белка определяли по методу Бредфорда. В полученные 25 мкг лизата с концентрацией белка 1 мкг/мкл добавляли 50 мкМ Me4 VодipyFL-Ahx 3 Leu 3 VS (флуоресцентно меченный аналог субстрата, реагирующий с N-концевым треонином всех активных субъединиц протеасом).

Образцы инкубировали в течение 1 ч при +37°C. Белки разделяли с помощью гель-электрофореза в денатурирующих условиях (SDS-PAGE) в 12.5%-ном ПААГ. Сканирование геля проводили на системе ChemiDoc MP (возбуждение при 480 нм и эмиссия при 530 нм).

**Измерение общего количества протеасом.** Для определения количества протеасом мух замораживали в жидком азоте и лизировали в буфере Лэммли. Белки разделяли в 12.5%-ном ПААГ. После электрофореза белки переносили на нитроцеллюлозную мембрану (“Amersham Protran 0.45 μm”, Великобритания) методом полусухого блоттинга. Протеасомы выявляли с помощью первичных антител к субъединицам α1, 2, 3, 5, 6 & 7 20S протеасомы (MCP231, “Enzo Life Sciences”, Швейцария, разведение 1 : 1000) и вторичных антител козы к IgG мыши, конъюгированных с пероксидазой хрена (HRP) (разведение 1 : 5000). Иммуные комплексы выявляли с помощью хемилюминесцентного набора Thermo Scientific SuperSignal West Pico Plus (США) с последующей визуализацией на системе ChemiDoc MP.

**MALDI-TOF-масс-спектрометрия.** Масс-спектрометрию пептидов после трипсинолиза белков проводили, как описано ранее [22]. Белки идентифицировали, используя программное обеспечение Mascot (www.matrixscience.com) в базе NCBI.

**Обсчет данных и статистический анализ.** В каждом эксперименте анализировали по три биологических повтора. При измерении активности



**Рис. 1.** Включение  $S^{35}$ -метионина в белки слюнных желез *Drosophila melanogaster*. Белки разделяли методом двумерного электрофореза. Слюнные железы выделяли из линий: 58492 – контроль, линии с делециями *CSE*<sup>-/-</sup> и *CBS*<sup>-/-</sup>, и с двойной делецией *CBS*<sup>-/-</sup>;*CSE*<sup>-/-</sup>. Стрелка указывает положение белка FBP2.

протеасом интенсивность свечения обсчитывали по общему количеству белка, окрашенного Coomassie Brilliant Blue G-250. При измерении общего количества протеасом в качестве внутреннего контроля использовали актин. Результаты обрабатывали с помощью программы Fiji. В каждом эксперименте рассчитывали стандартные отклонения и значение *t*-коэффициента Стьюдента.

**Детекция окисленных белков.** Пятидневных мух замораживали в жидком азоте. Гомогенизировали в буфере RIPA, содержащем 2%  $\beta$ -меркаптоэтанола. Окисление белков оценивали с помощью набора OxyBlot Protein Oxidation Detection Kit (“Millipore”, США) в соответствии с инструкциями производителя. В этом методе карбонильные группы белков дериватируют до 2,4-динитрофенилгидразона (ДНФ-гидразон) путем реакции с 2,4-динитрофенилгидразином (ДНФГ). Далее ДНФ-дериватизированные белки выявляют методом Вестерн-блоттинга с помощью антител, специфичных к ДНФ.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

### *Изменения в уровне накопления белка жирового тела у линий мух с делецией гена *cbs**

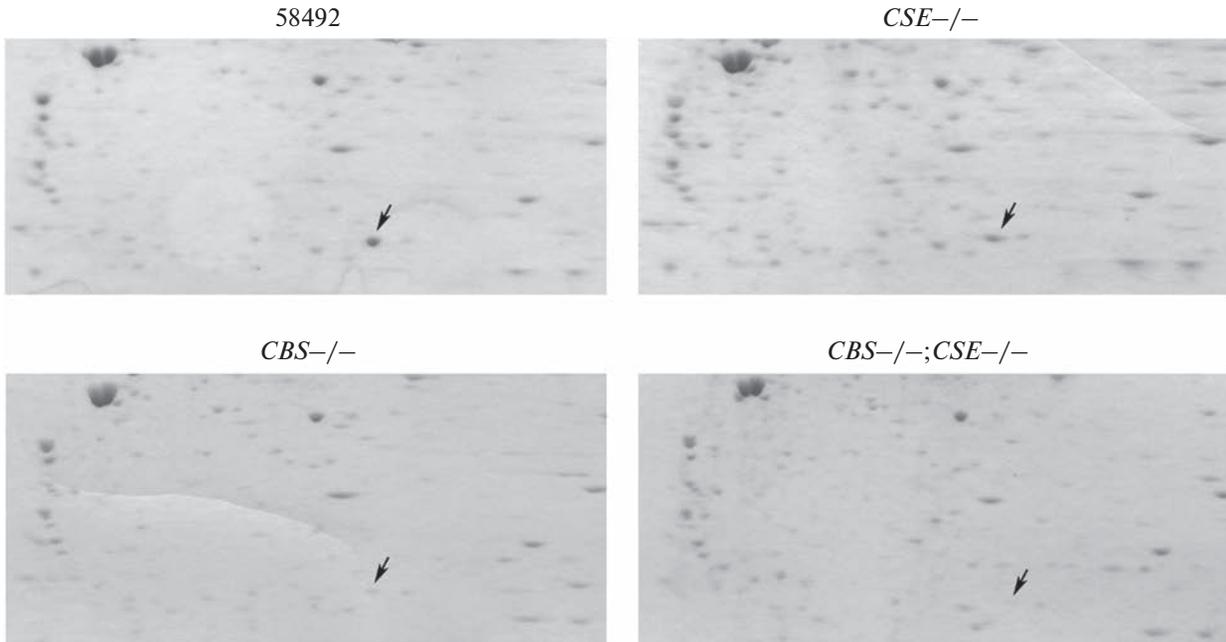
Ранее мы показали, что у линий *D. melanogaster* с делецией гена *cbs*, особенно у линий с двойной делецией генов *cbs* и *cse*, наблюдается задержка развития на стадиях от откладки яиц до окукливания [19]. С целью выявления возможных причин удлинения времени развития мы провели про-

теомный анализ белков, выделенных из слюнных желез личинок третьей стадии развития.

Анализ меченных метионином белков из слюнных желез изучаемых линий мух выявил множественные незначительные различия в уровне синтеза белков с низким уровнем экспрессии, детектируемые при включении метионина, но не видимые при окрашивании гелей Coomassie. Нами обнаружено значимое различие в уровне экспрессии белка жирового тела 2 (FBP2) (рис. 1). Ген *fbp2* кодирует белок из 256 аминокислотных остатков, который накапливается в жировом теле личинок третьей стадии и в куколках. Этот белок гомологичен алкогольдегидрогеназе, кодируемой геном *Adh*, но не обладает ферментативной активностью. Кодируемый *fbp2* полипептид содержит 20% остатков метионина, тогда как в зрелом полипептиде, кодируемом *Adh*, имеется лишь один остаток метионина [23]. Предполагается, что FBP2 является белком, служащим источником метионина и серы, необходимых на стадии метаморфоза дрозофилы.

На рис. 1 видно мощное включение метионина в белок FBP2 (отмечен стрелкой) в контрольной линии и снижение уровня включения метионина в этот белок в линии с делецией гена *cse*. В линиях с делецией гена *cbs* и с двойной делецией генов *cbs* и *cse* наблюдали наиболее значительное снижение синтеза этого белка.

Анализ гелей, окрашенных в растворе Coomassie, подтвердил различия в уровне накопления белка FBP2 в анализируемых линиях (рис. 2). Мы также проанализировали линию с делецией гена *mst*.



**Рис. 2.** Двумерный электрофорез белков слюнных желез. Гели окрашивали в 0.2%-ном растворе Coomassie Brilliant Blue G-250. Стрелкой отмечен белок FBP2.

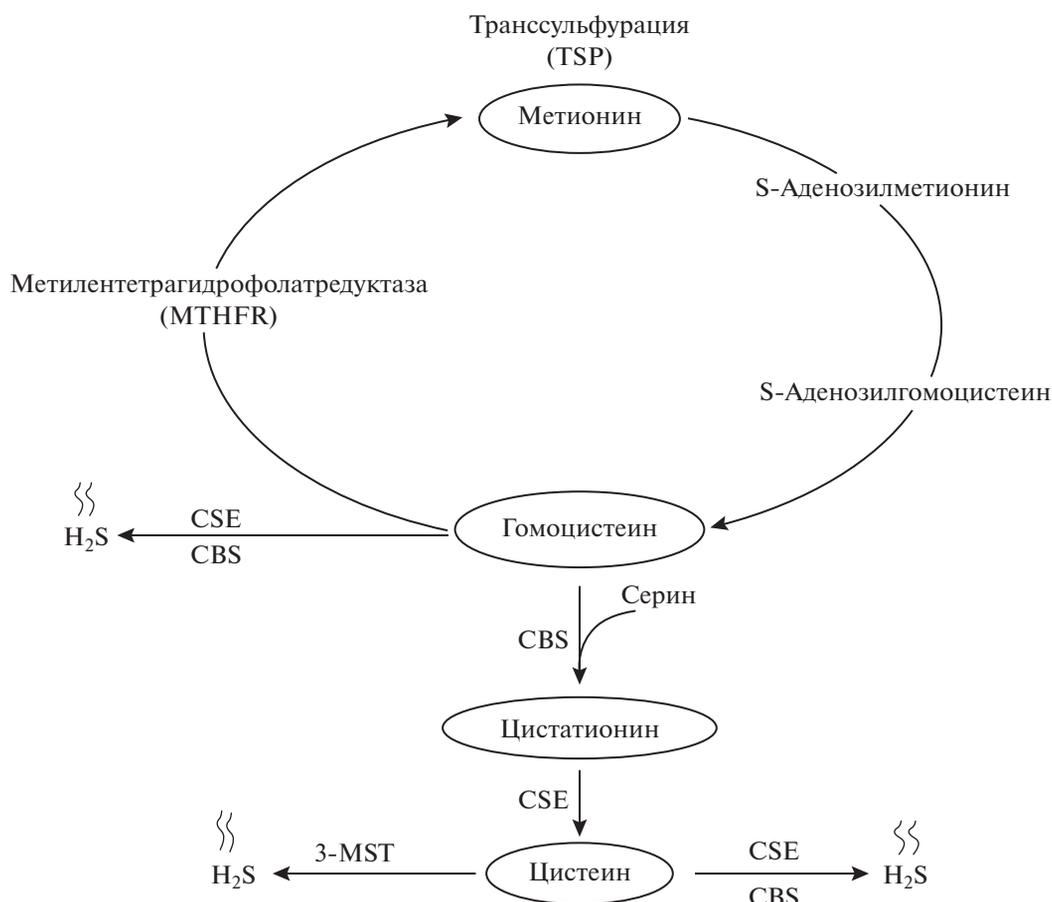
Делеция этого гена не влияет на путь транссульфурации и, соответственно, на синтез аминокислот. Уровень синтеза белка FBP2 в этой линии остается таким же, как в контрольной линии (не показано).

Проведенный нами ранее анализ уровня аминокислот [24] выявил двух–трехкратное увеличение количества свободного метионина в линиях мух с делецией гена *cbs* и с двойной делецией генов *cbs* и *cse*. Таким образом, наши результаты показывают существование обратной корреляции между уровнем свободного метионина и уровнем экспрессии белка FBP2 и предполагают, что накопление этого белка зависит от уровня накопления метионина. Можно предположить, что нарушение пути транссульфурации (рис. 3), приводящее к накоплению метионина и снижению уровня цистеина, может влиять на уровень синтеза запасующих белков, используемых в процессе метаморфоза.

#### Анализ протеома яичников

Ранее мы выявили изменения в развитии яичников и плодовитости у самок дрозофилы с делецией гена *cbs* и с двойной делецией генов *cbs* и *cse* [19]. В данной работе методом 2D-электрофореза изучали синтез белков в яичниках девственных пятидневных самок этих линий. Наиболее значимые различия наблюдали при сравнении линий с двойной делецией *CBS-/-; CSE-/-* с контрольной линией (рис. 4).

Интересно, что среди белков с измененным уровнем экспрессии или с измененной точкой изофокусирования идентифицированы два белка, относящихся к семейству тиоредоксинов (рис. 4, белки #1 (Jafrac1) и #7 (Thioredoxin-like); табл. 1). Тиоредоксины – высококонсервативные белки, содержащие в активном центре каталитически активный цистеиновый S–S-мостик. Тиоредоксины обладают свойствами редокс-активного дисульфида белковой природы и служат переносчиками электронов, а также играют важную роль в защите клеток от окислительного стресса путем детоксикации пероксидов. Экспрессия гена *jafrac1* в нейронах снижает уровень ROS, восстанавливает функцию митохондрий, ослабляет активацию JNK, вызванную паракватом [25]. Из рис. 4 видно, что в линии *CBS-/-; CSE-/-* повышен уровень экспрессии белков #1 и #7, а белок Jafrac1 сдвинут в кислую сторону, что свидетельствует о модификации этого важного белка. Транскриптомный анализ, проведенный нами ранее, выявил увеличение уровня экспрессии *jafrac1* у самцов с делетированными генами *cbs* и *cse* [18]. Также мы наблюдали повышение уровня экспрессии мРНК [18] и белка Lethal(2)-37Cс. Уровень экспрессии белка Lethal(2)-37Cс увеличивается при хронической гипоксии и имеет адаптивный эффект [26]. Наш транскриптомный анализ выявил увеличение экспрессии всех генов, кодирующих белки хориона. Мы также обнаружили увеличение уровня экспрессии белка хориона Cp7Fc (Chorion protein c at 7F). Двумерный электрофорез (рис. 4) выявил смещение в



**Рис. 3.** Путь транссульфурации и трансметилирования. Гомоцистеин синтезируется из метионина, получаемого с пищей. Гомоцистеин может быть реметилирован в метионин под действием 5,10-метилентетрагидрофолатредуктазы (MTHFR). CBS конденсирует гомоцистеин с серином с образованием цистатионина, который служит субстратом для CSE при синтезе цистеина. CSE может генерировать H<sub>2</sub>S либо из цистеина, либо из гомоцистеина. CBS при генерировании H<sub>2</sub>S использует комбинацию цистеина и гомоцистеина. Третий фермент, генерирующий H<sub>2</sub>S, 3-MST, в сочетании с цистеинамино-трансферазой (CAT) использует цистеин для образования H<sub>2</sub>S.

кислую область точки изофокусирования одной из основных субъединиц протеасомы (Proteasome beta 4 subunit) и изменение уровня экспрессии металлопротеазного компонента 26S протеасомы CSN5. Эти изменения могут влиять на активность протеасом.

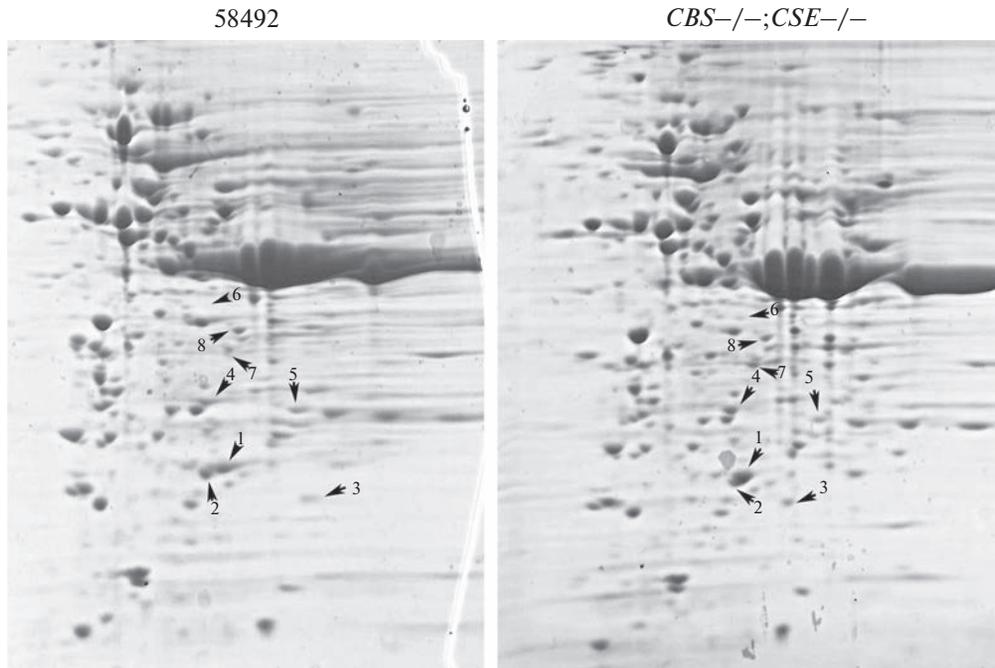
#### *Активность протеасом и уровень их экспрессии в изучаемых линиях дрозофилы*

На основании результатов 2D-анализа представлялось интересным проанализировать количество и активность протеасом в изучаемых линиях дрозофилы. Известно, что основной функцией протеасом является деградация различных белков, в том числе окисленных и поврежденных, до коротких пептидов [27, 28]. Ранее была выявлена активация протеасом в условиях окислительного стресса и в ходе восстановления после теплового шока [29, 30].

Мы предположили возможность изменения активности протеасом в линиях мух с делетированными генами серного метаболизма. Для проверки этой гипотезы измерили общую активность протеасом в белковом лизате, полученном из пятидневных самок (рис. 5). Видно, что во всех линиях с делетированными генами *cbs* и *cse*, а также у двойных трансформантов наблюдается уменьшение количества каталитически-активных субъединиц протеасом. В линии с делецией гена *mst* это снижение минимально.

Для того, чтобы выяснить, не связано ли снижение активности протеасом с уменьшением их общего количества, был проведен Вестерн-блот-анализ содержания структурных субъединиц протеасом в лизатах тканей (рис. 6).

Видно, что в линиях с делецией генов серного метаболизма наблюдается снижение содержания субъединиц  $\alpha 1, 2, 3, 5, 6, 7$  протеасом. Таким образом наблюдается корреляция между снижением



**Рис. 4.** Двумерный электрофорез белков, выделенных из яичников 5-дневных девственных самок контрольной линии 58492 и линии с двойной делецией *CBS*<sup>-/-</sup>;*CSE*<sup>-/-</sup>. Окрашивание Coomassie Brilliant Blue G-250. Цифрами отмечены белки, идентифицированные методом пептидного фингерпринтинга (табл. 1).



**Рис. 5.** *а* – Выявление каталитически активных субъединиц протеасом, специфически связывающихся с флуоресцентным зондом Me4BodipyFL–Ahx3Leu3VS. 1, 2 – Контрольная линия 58492; 3, 4 – *CSE*<sup>-/-</sup>; 5, 6 – *CBS*<sup>-/-</sup>; 7, 8 – *CBS*<sup>-/-</sup>, *CSE*<sup>-/-</sup>(1); 9, 10 – *CBS*<sup>-/-</sup>, *CSE*<sup>-/-</sup>(2); 11 – *MST*<sup>-/-</sup>; К(–) – отрицательный контроль. *б* – Изменение количества каталитически-активных субъединиц, нормализация по актину, окрашенному Coomassie G-250. 1 – 58492; 2 – *CSE*<sup>-/-</sup>; 3, 4 – *CBS*<sup>-/-</sup>; 5 – *CBS*<sup>-/-</sup>, *CSE*<sup>-/-</sup>(1); 6 – *CBS*<sup>-/-</sup>; *CSE*<sup>-/-</sup>(2); 7 – *MST*<sup>-/-</sup>. Объединены данные трех независимых экспериментов. Статистическая значимость изменения по сравнению с контрольной линией. \**p* < 0.05, \*\**p* < 0.01.

ем количества каталитически-активных и структурных субъединиц протеасом. Полученные данные оказались довольно неожиданными для нас. Ранее транскриптомный анализ библиотек РНК, полученных из самок исследуемых нами линий, не выявил значимых различий в уровне экспрессии генов, кодирующих структурные белки протеасом и убиквитинлигаз [18]. При этом мы выявили активацию генов, участвующих в окисли-

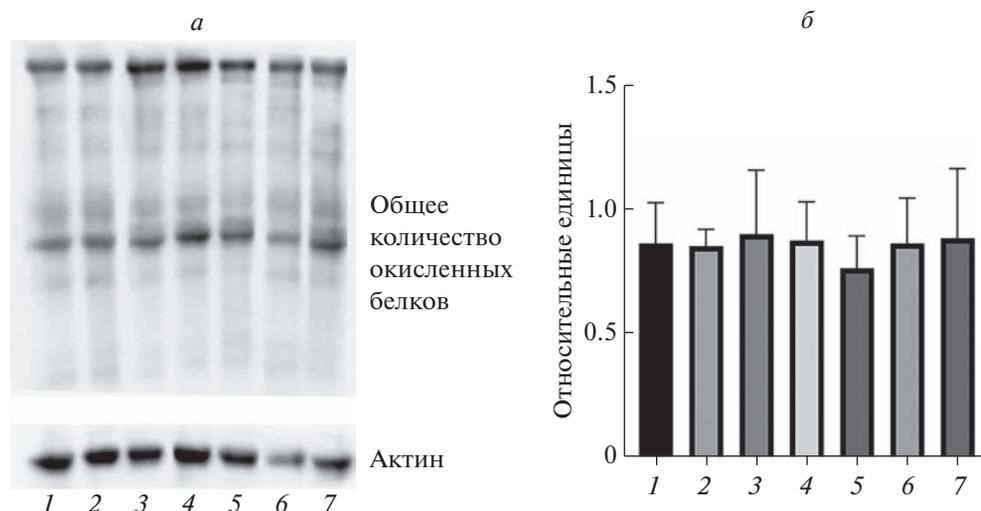
тельно-восстановительных процессах в линиях с делециями генов *cbs*, *cse*, что свидетельствует об окислительном стрессе. Известно, что накопление окисленных белков приводит к активации протеасом и деградации окисленных белков [29–31]. Исходя из этих соображений, мы проверили, имеются ли различия в количестве окисленных белков в линиях с делециями генов серного метаболизма. Оксид-Блот-анализ экстрактов, полученных из пя-



**Рис. 6.** *a* – Вестерн-блот-анализ белков со специфическими антителами к субъединицам  $\alpha$ 1, 2, 3, 5, 6, 7 протеасомы 20S (см. Экспериментальную часть). 1 – Контрольная линия 58492; 2 – *CSE*<sup>-/-</sup>; 3, 4 – *CBS*<sup>-/-</sup>; 5 – *CBS*<sup>-/-</sup>; *CSE*<sup>-/-(1)</sup>; 6 – *CBS*<sup>-/-</sup>; *CSE*<sup>-/-(2)</sup>; 7, 8 – *MST*<sup>-/-</sup>; K(+) – положительный контроль. *b* – Изменение общего количества субъединиц протеасом, нормализация по актину. 1 – 58492; 2 – *CSE*<sup>-/-</sup>; 3, 4 – *CBS*<sup>-/-</sup>; 5 – *CBS*<sup>-/-</sup>; *CSE*<sup>-/-(1)</sup>; 6 – *CBS*<sup>-/-</sup>; *CSE*<sup>-/-(2)</sup>; 7 – *MST*<sup>-/-</sup>. Общее количество протеасом вычисляли по данным трех независимых экспериментов. Статистическая значимость изменения по сравнению с контрольной линией. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.005$ .

**Таблица 1.** Основные белки, характер экспрессии которых изменился

Белок	Краткая характеристика	2D-анализ: сравнение белков линии <i>CBS</i> <sup>-/-</sup> ; <i>CSE</i> <sup>-/-</sup> и 58492
#1. <b>Jafrac1</b> (Thioredoxin peroxidase 1, CG1633)	Тиолспецифическая пероксидаза, катализирует восстановление пероксида водорода и органических гидропероксидов до воды и спиртов соответственно	Уровень экспрессии увеличивается в 1.2 раза, сдвигается точка изофокусирования сдвигается в кислую сторону
#2. <b>Fer1HCH</b> (Ferritin 1 heavy chain homologue, CG2216)	Одна из двух субъединиц основного комплекса хранения железа – молекулы ферритина 1	
#3. <b>Pros<math>\beta</math>4</b> (Proteasome beta4 subunit, GC17331)	Субъединица протеасомы- $\beta$ 4	Изменяется точка изофокусирования (сдвигается в кислую сторону)
#4. <b>MFS14</b> (Lethal(2)-37Cc, CG15095)	Транспортер 14 надсемейства основных посредников (MFS14) участвует в реакции на гипоксию.	Уровень экспрессии увеличивается в 1.2 раза
#5. <b>Hsp26</b> (Heat shock protein 26, CG4183)	Белок теплового шока 26 (Hsp26), участвует в белковом фолдинге	Уровень экспрессии снижается 0.74 раза
#6. <b>Cp7Fc</b> (Chorion protein c at 7F (Cp7Fc), isoform A)	Белок хориона С в 7F (Cp7Fc), принадлежит к семейству из девяти белков хориона. Экспрессируется в фолликулярных клетках	Уровень экспрессии увеличивается в 1.9 раза
#7. <b>Txl</b> (Thioredoxin- like, CG5495)	Тиоредоксин-подобный белок, участвует, согласно предварительному анализу, в окислительно-восстановительном гомеостазе клетки	Уровень экспрессии увеличивается в 1.5 раза
#8. <b>COP9 signalosome subunit 5, CSN5</b> (Metalloprotease component of the 26S proteasome, CG14884)	CSN5 – пятая субъединица сигнасомы COP9. Сигнасома играет центральную роль в регуляции E3-куллинового кольца убиквитинлигаз	Уровень экспрессии снижается 0.75 раза



**Рис. 7.** *a* – ОксиБлот-анализ (см. Экспериментальную часть). 1 – 58492; 2 – *CSE*<sup>-/-</sup>; 3, 4 – *CBS*<sup>-/-</sup>; 5 – *CBS*<sup>-/-</sup>, *CSE*<sup>-/-</sup>(1); 6 – *CBS*<sup>-/-</sup>, *CSE*<sup>-/-</sup>(2); 7 – *MST*<sup>-/-</sup>. *b* – Общее количество окисленных белков, нормализация по актину. 1 – 58492; 2 – *CSE*<sup>-/-</sup>; 3, 4 – *CBS*<sup>-/-</sup>; 5 – *CBS*<sup>-/-</sup>, *CSE*<sup>-/-</sup>(1); 6 – *CBS*<sup>-/-</sup>, *CSE*<sup>-/-</sup>(2); 7 – *MST*<sup>-/-</sup>. Объединены данные пяти независимых экспериментов.

тидневных самок, использовали для определения общего количества окисленных белков (рис. 7).

Из рис. 7 видно, что по количеству окисленных белков линии мух с делециями генов серного метаболизма не отличаются статистически значимо от контрольной линии, что может свидетельствовать об адаптации мух к сниженной продукции сероводорода и предотвращении окисления белков.

Причины и механизмы снижения количества протеасом требуют дальнейшего изучения. Мы полагаем, что одной из возможных причин является дисбаланс количества свободных аминокислот (избыток метионина и недостаток цистеина) у мух с делециями генов *chs* и *cse*. Ранее на клетках млекопитающих показали, что аминокислотное голодание значительно ускоряет деградацию 26S протеасомы, что может приводить к снижению числа таких комплексов в клетке [33, 34].

Полученные нами результаты показали, что нарушение пути транссульфурации и трансметилирования при делеции генов *chs* и *cse* имеет значимые последствия на уровне протеома. У личинок третьего возраста выявлено снижение синтеза и накопления белка ГВР2, запасящего метионин. В яйчниках обнаружены изменения в характере экспрессии белков, участвующих в защите организма от окислительного стресса и гипоксии. Особый интерес вызывает снижение количества каталитически-активных и структурных субъединиц протеасом у мух с делециями генов *chs* и *cse*, что требует дальнейшего изучения.

Благодарим А.В. Морозова за консультации и помощь в постановке экспериментов, М. Сереб-

рякову за проведение масс-спектрометрического анализа пептидов и С. Сорокину за техническую помощь.

Работа проведена при поддержке Российского научного фонда (№ 2017-74- 30030-П).

В работе использованы лабораторные линии *Drosophila melanogaster*, соблюдены все применимые международные стандарты ведения и использования линий.

Авторы заявляют об отсутствии конфликта интересов.

## СПИСОК ЛИТЕРАТУРЫ

1. Mota-Martorell N., Jové M., Borrás C., Berdún R., Obis E., Sol J., Cabré R., Pradas I., Galo-Licona J.D., Puig J., Viña J., Pamplona R. (2020) Methionine trans-sulfuration pathway is upregulated in long-lived humans. *Free Rad. Biol. Med.* **162**, 38–52.
2. Parkhitko A.A., Jouandin P., Mohr S.E., Perrimon N. (2019) Methionine metabolism and methyltransferases in the regulation of aging and lifespan extension across species. *Aging Cell.* **18**, e13034.
3. Perridon B.W., Leuvenink H.G., Hillebrands J.L., van Goor H., Bos E.M. (2016) The role of hydrogen sulfide in aging and age-related pathologies. *Aging.* **8**, 2264–2289.
4. Sokolov A.S., Nekrasov P.V., Shaposhnikov M.V., Moskalev A.A. (2021) Hydrogen sulfide in longevity and pathologies: inconsistency is malodorous. *Ageing Res. Rev.* **67**, 101262.
5. Tabibzadeh S. (2021) Signaling pathways and effectors of aging. *Front. Biosci.* **26**, 50–96.
6. Xiao Q., Ying J., Xiang L., Zhang C. (2018) The biologic effect of hydrogen sulfide and its function in various diseases. *Medicine.* **97**, e13065.

7. Kimura Y., Goto Y., Kimura H. (2010) Hydrogen sulfide increases glutathione production and suppresses oxidative stress in mitochondria. *Antioxid. Redox Signal.* **12**, 1–13.
8. Kabil O., Banerjee R. (2014) Enzymology of H<sub>2</sub>S biogenesis, decay and signaling. *Antioxid. Redox Signal.* **20**, 770–782.
9. Paul B.D., Snyder S.H. (2012) H<sub>2</sub>S signaling through protein sulfhydration and beyond. *Nat. Rev. Mol. Cell Biol.* **13**, 499–507.
10. Mudd S.H., Levy H.L., Kraus J.P. (2001) Disorders of transsulfuration. In: *The Online Metabolic and Molecular Bases of Inherited Disease*. New York: McGraw-Hill, 2007–2056.
11. Guzmán M.A., Navarro M.A., Carnicer R., Sarría A.J., Acín S., Arnal C., Muniesa P., Surra J.C., Arbonés-Mainar J.M., Maeda N., Osada J. (2006) Cystathionine  $\beta$ -synthase is essential for female reproductive function. *Hum. Mol. Genet.* **21**, 3168–3176.
12. Shirozu K., Tokuda K., Marutani E., Lefér D., Wang R., Ichinose F. (2014) Cystathionine  $\gamma$ -lyase deficiency protects mice from galactosamine/lipopolysaccharide induced acute liver failure. *Antioxid. Redox Signal.* **20**, 204–216.
13. Badié A., Chambers S.T., Gaddam R.R., Bhatia M. (2016) Cystathionine  $\gamma$ -lyase gene silencing with siRNA in monocytes/macrophages attenuates inflammation in cecal ligation and puncture induced sepsis in the mouse. *J. Biosci.* **41**, 87–95.
14. Gaddam R.R., Fraser R., Badié A., Chambers S., Cogger V.C., Le Couteur D.G., Ishii I., Bhatia M. (2016) Cystathionine gamma-lyase gene deletion protects mice against inflammation and liver sieve injury following polymicrobial sepsis. *PLoS One.* **11**, e0160521.
15. Kolluru G.K., Bir S.C., Yuan S., Shen X., Pardue S., Wang R., Kevil C.G. (2015) Cystathionine gamma-lyase regulates arteriogenesis through no-dependent monocyte recruitment. *Cardiovasc. Res.* **107**, 590–600.
16. Yuan S., Yurdagül A., Jr, Peretik J.M., Alfai M., Al Yafeai Z., Pardue S., Kevil C.G., Orr A. W. (2018) Cystathionine  $\gamma$ -lyase modulates flow-dependent vascular remodeling. *Arterioscler. Thromb. Vasc. Biol.* **38**, 2126–2136.
17. Snijder P.M., Baratashvili M., Grzeschik N.A., Leuvenink H.G.D., Kuijpers L., Huitema S., Schaap O., Giermans B.N.G., Kuipers J., Miljkovic J.L., Mitrovic A., Bos E.M., Szabó C., Kampinga H.H., Dijkers P.F., Bos E.M., Szabó C., Kampinga H.H., Dijkers P.F., Dunnen WFAD, Filipovic M.R., Goor H.V., Sibon OCM. (2016) Overexpression of cystathionine  $\gamma$ -lyase suppresses detrimental effects of spinocerebellar ataxia type 3. *Mol. Med.* **21**, 758–768.
18. Zatschina O., Karpov D., Chuvakova L., Rezykh A., Funikov S., Sorokina S., Zakluta A., Garbuz D., Shilova V., Evgen'ev M. (2020) Genome-wide transcriptional effects of deletions of sulphur metabolism genes in *Drosophila melanogaster*. *Redox Biol.* **36**, 101654.
19. Shaposhnikov M.V., Zakluta A.S., Zemskaya N.V., Guvatova Z.G., Shilova V.Y., Yakovleva D.V., Gorbunova A.A., Koval L.A., Ulyasheva N.S., Evgen'ev M.B., Zatschina O.G., Moskalev A.A. (2022) Deletions of the cystathionine- $\beta$ -synthase (CBS) and cystathionine- $\gamma$ -lyase (CSE) genes, involved in the control of hydrogen sulfide biosynthesis, significantly affect lifespan and fitness components of *Drosophila melanogaster*. *Mech. Ageing Dev.* **203**, 111656.
20. O'Farrell P.Z., Goodman H.M., O'Farrell P.H. (1977). High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell.* **12**, 1133–1141.
21. de Jong A., Schuurman K.G., Rodenko B., Ovaas H., Berkers C.R. (2012) Fluorescence-based proteasome activity profiling. *Meth. Mol. Biol.* **803**, 183–204.
22. Zatschina O.G., Kechko O.I., Mitkevich V.A., Kozin S.A., Yurinskaya M.M., Vinokurov M.G., Serebryakova M.V., Rezykh A.P., Evgen'ev M.B., Makarov A.A. (2018) Amyloid- $\beta$  with isomerized Asp7 cytotoxicity is coupled to protein phosphorylation. *Sci. Rep.* **8**, 3518.
23. Meghlaoui G.K., Veuille M. (1997) Selection and methionine accumulation in the fat body protein 2 gene (FBP2), a duplicate of the *Drosophila* alcohol dehydrogenase (ADH) gene. *J. Mol. Evol.* **44**, 23–32.
24. Zatschina O.G., Chuvakova L.N., Nikitina E.A., Rezykh A.P., Zakluta A.S., Sarantseva S.V., Surina N.V., Ksenofontov A.L., Baratova L.A., Shilova V.Y., Evgen'ev M.B. (2022) Genes responsible for H<sub>2</sub>S production and metabolism are involved in learning and memory in *Drosophila melanogaster*. *Biomolecules.* **12**, 751.
25. Lee K.S., Iijima-Ando K., Iijima K., Lee W.J., Le J.H., Yu K., Lee D.S. (2009) JNK/FOXO-mediated neuronal expression of fly homologue of peroxiredoxin II reduces oxidative stress and extends life span. *J. Biol. Chem.* **284**, 29454–29461.
26. Azad P., Zhou D., Russo E., Haddad G.G. (2009) Distinct mechanisms underlying tolerance to intermittent and constant hypoxia in *Drosophila melanogaster*. *PLoS One.* **4**, e5371.
27. Raynes R., Pomatto L.C., Davies K.J. (2016) Degradation of oxidized proteins by the proteasome: distinguishing between the 20S, 26S, and immunoproteasome proteolytic pathways. *Mol. Aspects Med.* **50**, 41–55.
28. Lefaki M., Papaevgeniou N., Chondrogianni N. (2017) Redox regulation of proteasome function. *Redox Biol.* **13**, 452–458.
29. Aiken C.T., Kaake R.M., Wang X., Huang L. (2011) Oxidative stress-mediated regulation of proteasome complexes. *Mol. Cell. Proteomics.* **10**, R110.006924.
30. Морозов А.В., Буров А.В., Астахова Т.М., Спаская Д.С., Маргулис Б.А., Карпов В.Л. (2019) Динамика функциональной активности и экспрессии субъединиц протеасом в условиях адаптации клетки к тепловому шоку. *Молекуляр. биология.* **53**, 638–647.
31. Jung T., Höhn A., Grune T. (2014) The proteasome and the degradation of oxidized proteins: Part II – protein oxidation and proteasomal degradation. *Redox Biol.* **2**, 99–104.
32. Höhn T.J., Grune T. (2014) The proteasome and the degradation of oxidized proteins: part III – redox regulation of the proteasomal system. *Redox Biol.* **2**, 388–394.
33. Cohen-Kaplan V., Livneh I., Avni N., Fabre B., Ziv T., Kwon Y.T., Ciechanover A. (2016) p62- and ubiquitin-dependent stress-induced autophagy of the mammalian 26S proteasome. *Proc. Natl. Acad. Sci. USA.* **113**, E7490–E7499.
34. Hoeller D., Dikic I. (2016) How the proteasome is degraded. *Proc. Natl. Acad. Sci. USA.* **113**, 13266–13268.

## The Effect of the Knockout of Major Transsulfuration Genes on the Pattern of Protein Synthesis in *D. melanogaster*

A. S. Zakluta<sup>1</sup>, V. Y. Shilova<sup>1</sup>, and O. G. Zatsepina<sup>1, \*</sup>

<sup>1</sup>Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia

\*e-mail: olzacepina@yandex.ru

The enzymes involved in the transsulfuration pathway and hydrogen sulfide production – cystathionine- $\beta$ -synthase (CBS), cystathionine- $\gamma$ -lyase (CSE) and 3-mercaptopyruvate sulfurtransferase (3-MST) – play an important cytoprotective role in the functioning of the organism. Using CRISPER/Cas9 technology, we obtained *Drosophila* lines with deleted *cbs*, *cse*, and *mst* genes as well as with double deletion of *cbs* and *cse* genes. We analyzed the effect of these mutations on the pattern of protein synthesis in the salivary glands of third instar larvae and in the ovaries of mature flies. In the salivary glands of lines with *cbs* and *cse* deletions, a decrease was found in the accumulation of the FBP2 storage protein containing 20% methionine amino acid residues. In the ovaries, changes were detected in the level of expression and isofocusing points of proteins involved in cell protection against oxidative stress, hypoxia, and protein degradation. It was shown that in the lines with deletions of transsulfuration enzymes the proteins have a similar degree of oxidation to that of the control line. A decrease in the total number of proteasomes and their activity was found in the lines with deletions of the *cbs* and *cse* genes.

**Keywords:** genes of transsulfuration pathway, CRISPER/Cas9 technology, *Drosophila*, proteasomes