

УДК 577.29

## НУКЛЕАЗЫ CRISPR-Cas II ТИПА: АЛГОРИТМ ПОИСКА И *in vitro* ХАРАКТЕРИСТИКА<sup>1</sup>

© 2023 г. А. А. Васильева<sup>a, b, \*</sup>, С. А. Алюкас<sup>c</sup>, П. А. Селькова<sup>a, b</sup>, А. Н. Арсениев<sup>a, b</sup>, В. Е. Чернова<sup>b</sup>,  
О. С. Мушарова<sup>a</sup>, Е. И. Климука<sup>a</sup>, М. А. Ходорковский<sup>b</sup>, К. В. Северинов<sup>a, d, \*\*</sup>

<sup>a</sup>Институт молекулярной генетики Российской академии наук, Москва, 123182 Россия

<sup>b</sup>Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, 195251 Россия

<sup>c</sup>ООО “Биотехнологический кампус”, Москва, 117437 Россия

<sup>d</sup>Waksman Institute of Microbiology, Piscataway, NJ, 08854 USA

\*e-mail: daucussativus7@gmail.com

\*\*e-mail: severik@waksman.rutgers.edu

Поступила в редакцию 25.11.2022 г.

После доработки 18.12.2022 г.

Принята к публикации 23.12.2022 г.

Использование компонентов систем адаптивного иммунитета бактерий CRISPR-Cas для внесения направленных изменений в ДНК открыло широкие перспективы для программируемого редактирования геномов высших организмов. Наиболее широко используемые генные редакторы основаны на Cas9-эффекторах CRISPR-Cas систем II типа. Cas9 способны специфически вносить двуниевые разрывы в участки ДНК, комплементарные последовательностям направляющих РНК. Несмотря на широкий спектр охарактеризованных Cas9, поиск новых вариантов Cas9 остается актуальной задачей, так как доступные Cas9-редакторы имеют ряд ограничений. В данной работе представлен алгоритм поиска и последующей характеристики новых Cas9-нуклеаз, разработанный в нашей лаборатории. Приведены подробные протоколы, описывающие биоинформатический поиск, клонирование и выделение рекомбинантных белков Cas9, проверку наличия нуклеазной активности *in vitro* и определение PAM-последовательности, необходимой для узнавания ДНК-мишеней. Рассмотрены потенциальные сложности, которые могут возникнуть в процессе работы, а также способы их преодоления.

**Ключевые слова:** CRISPR-система, Cas9, нуклеаза, редактирование генома, биоинформатический поиск

**DOI:** 10.31857/S0026898423030163, **EDN:** CHWIKZ

### ВВЕДЕНИЕ

CRISPR-Cas системы обеспечивают адаптивный иммунитет бактерий и архей. Полная CRISPR-Cas система состоит из CRISPR-кассеты и кластера генов, кодирующих Cas-белки [1]. CRISPR-кассета — это участок ДНК, в котором короткие повторы разделяются уникальными последовательностями — спейсерами. Несмотря на название CRISPR (Clustered Regularly Interspaced Short

Palindromic Repeats — кластеризованные регулярно разделенные короткие палиндромные повторы) в ряде случаев, в частности в системах II типа, рассматриваемых в данной работе, повторы не относятся к палиндромам. Некоторые CRISPR-спейсеры возникают за счет встраивания чужеродного генетического материала (фрагментов вирусных или плазмидных последовательностей) в CRISPR-кассету. В связи с этим многообразие спейсеров, накопленных в CRISPR-кассете конкретной бактерии, можно считать “памятью” об инфекциях, перенесенных ее предками. В результате транскрипции CRISPR-кассеты и процессинга образующейся некодирующей РНК формируются короткие CRISPR-РНК (CCRISPR-RNA, cgРНК), каждая из которых содержит последовательность одного из спейсеров и фланкирующую последовательность фрагмента повтора. Индивидуальные cgРНК связываются с Cas-нуклеазами,

<sup>1</sup> Дополнительная информация для этой статьи доступна по doi 10.31857/S0026898423030163 для авторизованных пользователей.

Сокращения: CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) — кластеризованные регулярно разделенные короткие палиндромные повторы; cgРНК (CCRISPR-RNA) — CRISPR-РНК; DR (Direct Repeat) — прямой повтор; ORF (Open Reading Frame) — открытая рамка считывания; PAM (Protospacer Adjacent Motif) — примыкающий к протоспейсеру мотив; transРНК (trans-activating CCRISPR-RNA) — *транс*-активирующая cgРНК.

образуя эффекторные комплексы. При появлении в клетке ДНК (а для некоторых систем РНК), содержащей участки, соответствующие одному из спейсеров CRISPR-кассеты (например, при повторной инфекции вирусом), эффекторный комплекс узнает ее за счет комплементарных взаимодействий с сгРНК. Активация нуклеазной активности белкового компонента эффекторного комплекса приводит к расщеплению чужеродной нуклеиновой кислоты в месте узнавания и последующей ее деградации [2–4].

Описанный общий механизм защитного действия систем CRISPR-Cas позволяет использовать эффекторные комплексы в качестве инструмента по направленному изменению последовательностей ДНК в клетках самых разных организмов, включая человека, что актуально для решения многих задач, стоящих перед биотехнологией и биомедициной. В этом случае используют эффекторный белок, запрограммированный не природной сгРНК, а специально подобранной направляющей РНК (RNA guide, РНК-гид). Участок направляющей РНК, соответствующий спейсеру сгРНК, комплементарен последовательности генома, в которую необходимо внести изменение.

Cas-белки можно разделить на две группы, исходя из их функций. Белки адаптационного модуля обеспечивают процесс адаптации – встраивания новых спейсеров в CRISPR-кассету. Белки эффекторного модуля отвечают за интерференцию – расщепление чужеродных нуклеиновых кислот. Исходя из строения эффекторов, CRISPR-Cas системы подразделяют на два класса [5]. В системах класса 1 эффекторы представляют собой комплексы, состоящие из нескольких Cas-белков. В системах класса 2 эффекторы – это односубъединичными белки. По наличию генов, кодирующих характерные дополнительные белки, системы класса 1 подразделяются на типы I, III и IV, а системы класса 2 – на типы II, V и VI. В свою очередь, каждый тип включает разнообразные подтипы [5].

Наличие лишь одного белка, достаточного для программируемого расщепления нуклеиновой кислоты, делает системы класса 2 наиболее удобными для практического применения. Эффекторы данного класса носят названия Cas9 (тип II), Cas12 (тип V) и Cas13 (тип VI). К наиболее изученным эффекторам класса 2 относятся несколько представителей типа II. Так, нуклеазу Cas9 из бактерии *Streptococcus pyogenes* (SpCas9) активно используют для редактирования геномов эукариот [6–8]. Однако геномные редакторы, основанные на охарактеризованных на сегодня Cas9-белках, не лишены недостатков. Во-первых, некоторые из этих нуклеаз имеют слишком большой размер, что затрудняет их доставку в клетки. Во-вторых, в

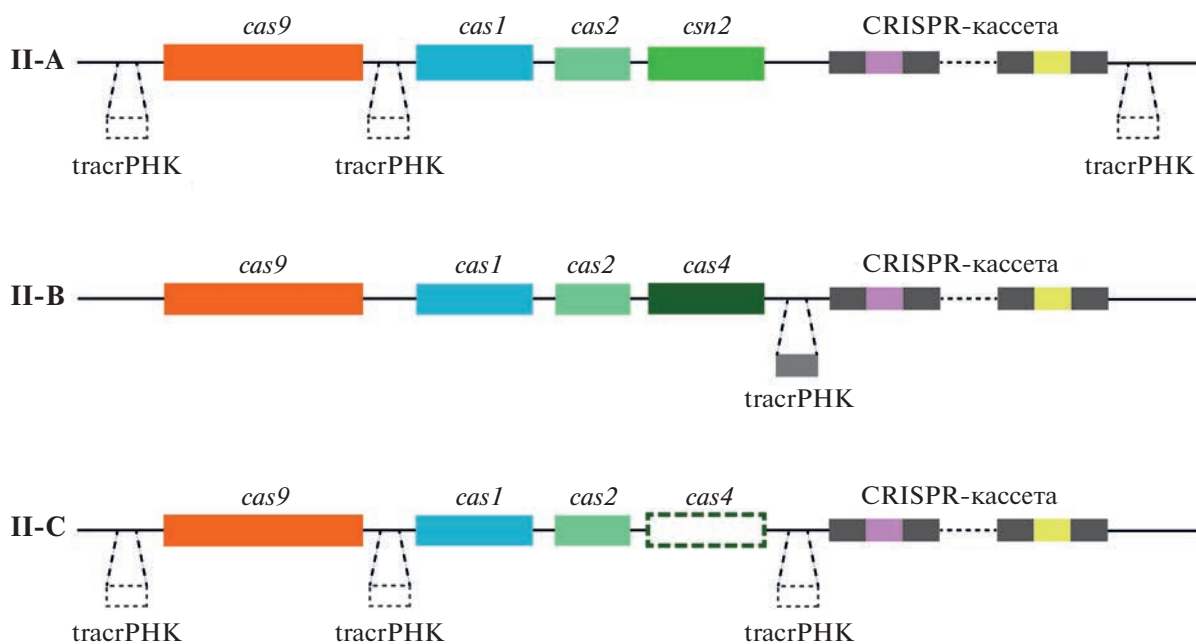
ряде случаев эффективность и специфичность редактирования оказывается невысокой. Наконец, мишени, доступные для конкретного Cas9-эффектора, требуют, кроме комплементарности к направляющей РНК, наличия специфической последовательности PAM (см. “Организация CRISPR-Cas систем II типа”), что ограничивает число мишеней, которые могут быть подвергнуты редактированию. Cas9-нуклеазы с новыми свойствами позволяют расширить спектр существующих инструментов геномного редактирования и подобрать оптимальные эффекторы для решения конкретных задач. Таким образом, поиск новых Cas9-нуклеаз и их характеристика – задача, актуальная как для фундаментальной, так и прикладной науки.

### ОРГАНИЗАЦИЯ CRISPR-Cas СИСТЕМ II ТИПА

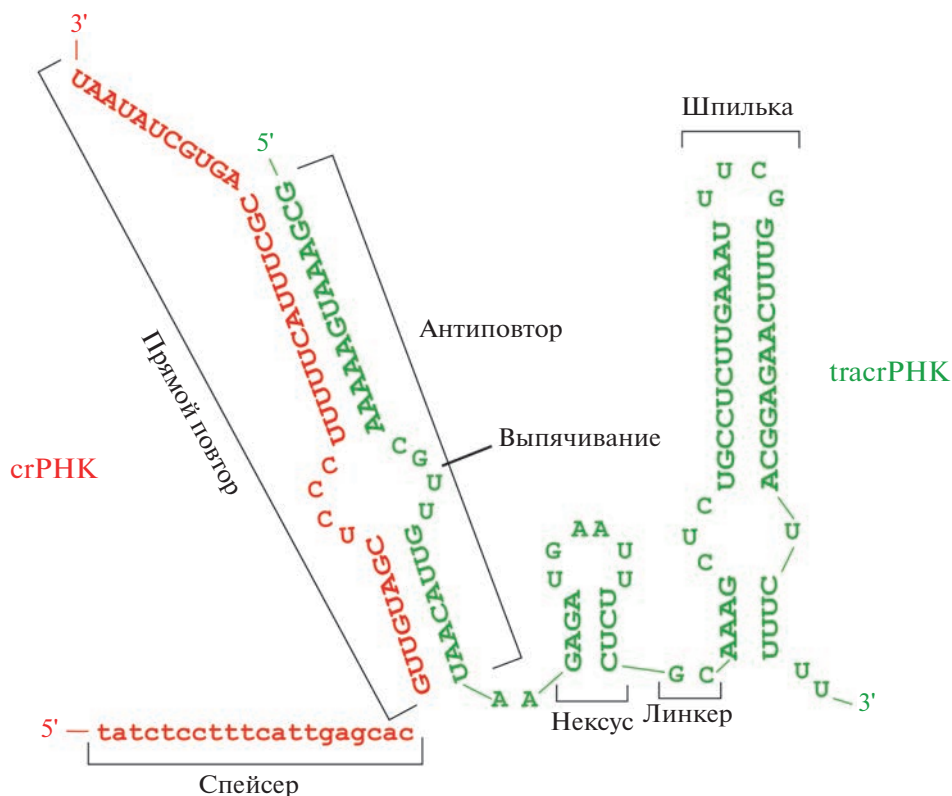
В системах CRISPR-Cas типа II роль эффектора выполняет нуклеаза Cas9. Состав адаптационного модуля зависит от конкретного подтипа (II-A, II-B или II-C) и может включать в себя белки Cas1, Cas2, Csn2, Cas4 (рис. 1). Подавляющее большинство CRISPR-кассет систем II типа имеет длину прямого повтора (Direct Repeat, DR) от 34 до 37 п.н. и длину спейсера – от 29 до 32 п.н. Число спейсеров обычно не превышает 30 (табл. S1, см. Дополнительные материалы).

Зрелые индивидуальные сгРНК содержат частичную последовательность спейсера (например, в случае хорошо изученной системы II-A бактерии *S. pyogenes* длина спейсеров в кассете составляет 30 п.н., а участок, соответствующий спейсеру в сгРНК, составляет лишь 20 нуклеотидов). Кроме того, сгРНК содержит фрагмент DR, примыкающий к 3'-концевой части спейсерного участка (рис. 2). Для процессинга первичного транскрипта CRISPR-кассеты и функционирования эффектора необходима дополнительная *транс*-активирующая сгРНК (*trans*-activating CRISPR-RNA, tracrPНК). Последовательность tracrPНК имеет участок, комплементарный фрагменту последовательности DR (антиповтор) (рис. 2). Молекулы tracrPНК образуют дуплексы с комплементарными участками DR в первичном транскрипте CRISPR-кассеты. Расщепление дуплексов клеточными РНКазами приводит к образованию индивидуальных сгРНК; tracrPНК остается связанной с молекулами сгРНК и становится частью эффекторного комплекса [9]. Как правило, ген tracrPНК находится поблизости от остальных компонентов локуса CRISPR-Cas (рис. 1).

Для узнавания и расщепления комплементарных сгРНК мишеней эффекторным комплексом Cas9 необходим мотив, примыкающий к прото-спейсеру (Protospacer Adjacent Motif, PAM) [10].



**Рис. 1.** Схема устройства локусов CRISPR-Cas систем подтипов II-A, II-B и II-C. Каждый локус содержит один ген *транс*-активирующей *сгРНК* (*trans-activating CRISPR-RNA*, *tracrPHK*); в системах, в которых местоположение этого гена варьирует, оно отмечено прямоугольниками, нарисованными пунктирными линиями. Для системы II-C ген белка Cas4 отмечен пунктиром, так как его присутствие опционально. Направление транскрипции отдельных элементов локуса не указано, поскольку оно может отличаться у разных представителей одного подтипа.



**Рис. 2.** Предсказанная *in silico* вторичная структура комплекса *сгРНК*-*тракрРНК* системы CRISPR-Cas II типа бактерии *Pasteurella pneumotropica* ATCC 35149. Последовательность *сгРНК* окрашена в красный цвет, *тракрРНК* – в зеленый.

РАМ фланкирует мишень с 3'-конца и узнается специальным доменом (РАМ Interacting Domain, PID) Cas9. Экспериментально охарактеризованные белки Cas9 узнают РАМ разной последовательности и длины (табл. S1, см. Дополнительные материалы). На данный момент четкой взаимосвязи между последовательностями PID различных Cas9-нуклеаз и конкретными последовательностями РАМ не просматривается. Существующие данные говорят лишь о наиболее предпочтительных вариантах взаимодействия определенных аминокислотных остатков PID и нуклеотидных оснований РАМ. Так, например, известно, что остатки аргинина в PID предпочтительно взаимодействуют с гуанинами, а остатки глутамина — с аденинами РАМ [11]. Подобного рода анализ помог предсказать наличие А-богатого РАМ у нуклеаз SmacCas9, что затем было подтверждено и экспериментально (SmacCas9: 5'-NAAN-3' РАМ) [12]. Однако вывести точную последовательность РАМ только лишь на основе анализа последовательности PID на данный момент не удастся. Более того, на примере белка Nme1-Cas9 и его ортологов было показано, что даже близкие по последовательности Cas9-нуклеазы могут значительно различаться по узнаваемым последовательностями РАМ (у Nme1Cas9 это 5'-NNNNGATT-3', а у близкого ортолога Nsp2Cas9 — 5'-NNNNCC-3') [13]. Следовательно, экспериментальное определение РАМ-последовательности — абсолютно необходимый шаг для характеристики не исследованных ранее белков Cas9 и принятии решения о перспективности их использования для геномного редактирования.

Таким образом, поиск и первичный (*in vitro*) анализ новых Cas9-нуклеаз включают следующие основные этапы:

1. Биоинформатический поиск генов, кодирующих Cas9-нуклеазы и соответствующие tracrRNA, и предсказание последовательности зрелой crRNA.
2. Гетерологическая экспрессия гена *cas9* в модельном штамме и получение рекомбинантного белка.
3. Нарботка crRNA и tracrRNA.
4. Демонстрация программируемого расщепления мишени рекомбинантной нуклеазой Cas9 и идентификация РАМ.

Полученная в результате выполнения этих этапов информация используется для принятия решения о перспективности дальнейших исследований, в частности оценки работы нового редактора в эукариотических клетках. Ниже приведено подробное описание алгоритма исследования для каждого из этапов первичного анализа на примере нуклеазы PpCas9, охарактеризованной в работе “PpCas9 from *Pasteurella pneumotropica* — a compact Type II-C Cas9 ortholog active in human cells” Я. Федоровой (I. Fedorova) и соавт. [14].

## БИОИНФОРМАТИЧЕСКИЙ ПОИСК НОВЫХ СИСТЕМ CRISPR-Cas II ТИПА

Ниже мы приводим пошаговое описание такого поиска.

1. Выбрать нуклеотидную последовательность, в которой будет проведен поиск системы CRISPR-Cas II типа.

Поиск можно вести как с полногеномными последовательностями, так и с метагеномными данными. Мы использовали как данные из общедоступных баз, так и результаты секвенирования, полученные в нашей лаборатории (использование собственных данных облегчает последующее патентование редакторов). В случае PpCas9 поиск проводили в находящейся в общем доступе последовательности генома бактерии *Rodentibacter pneumotropicus* ATCC 35149 (*[Pasteurella] pneumotropica*; GenBank Acc. No BBIX01000009.1).

2. Провести поиск CRISPR-кассеты.

Использовать программу CRT CRISPR Recognition Tool [15]. Минимальное число повторов в кассете (-minNR) — 2. Остальные параметры — по умолчанию.

Альтернативные программы для поиска: CRISPRCasFinder [16], PILER-CR [17], CRISPR-Detect [18], CRISPRidentify [19].

3. В обнаруженной в результате поиска нуклеотидной последовательности выбрать участок, окружающий CRISPR-кассету и содержащий 30000 п.н. с каждой ее стороны.

4. В выбранном участке определить все открытые рамки считывания (ORFs) длиной более 900 кодонов.

Использовать модель для генетического кода 11 (Бактерии, Археи и Пластиды растений). В качестве старт-кодонов использовать любой смысловой кодон.

5. Подготовить профиль HMM (Hidden Markov Model) для белков Cas9.

Для построения профиля использовать команду **hmmbuild** пакета HMMER [20] из множественного выравнивания последовательностей Cas9-нуклеаз. Выравнивание можно подготовить самостоятельно или использовать доступное в сторонних базах данных (Conserved Domain Database (CDD) или общий fasta-файл типа II в zip-архиве S1, см. Дополнительные материалы).

6. Использовать HMM-профиль Cas9 из п. 5 для поиска среди ORFs, предсказанных в п. 4, при помощи команды **hmmsearch** пакета HMMER.

7. Для более корректного определения подтипа гомологов Cas9, обнаруженных в п. 6, анализ повторяют, используя профили нуклеаз подтипов II и V (fasta-файлы в zip-архиве S1, см. Дополнительные материалы).

Дополнительная проверка с использованием профилей близкородственной группы (для типа II это системы типа V) рекомендуется для отсеивания ложноположительных совпадений с искомым профилем. Для дальнейшей работы выбирают те ORFs, для которых получен максимальный score с профилями интереса (II-A, II-B, или II-C).

#### 8. Проверить уникальность отобранных ORFs.

Для этого построить множественное выравнивание найденных ORFs и ранее описанных Cas9 (табл. S1, см. Дополнительные материалы). Вычислить попарные дистанции, используя матрицу BLOSUM62. Отобрать те ORFs, для которых не нашлось пары с дистанцией менее 0.5.

#### 9. Идентифицировать ген *tracrPНК*.

Существует ряд программ, предназначенных для поиска последовательностей *tracrPНК*. Например, алгоритм S. Dooley и др. [21] и CRISPR-*tracrRNA* [22]. Рекомендуется провести первичный поиск с использованием одной из этих программ с менее строгими, чем с заданными по умолчанию, параметрами (для CRISPR-*tracrRNA*: `-anti_repeat_similarity_threshold 0.3 -anti_repeat_coverage_threshold 0.3`). Отсеять предсказанные последовательности, которые находятся на расстоянии более чем 4000 п.н. от начала или конца ORF изучаемой Cas9-нуклеазы. Стоит отметить, что программы зачастую некорректно предсказывают начало и конец гена *tracrPНК*. В связи с этим рекомендуется использовать их только для определения приблизительного местоположения искомого гена, а затем вручную индивидуально оценивать каждый из предсказанных кандидатов, ориентируясь на следующие известные характеристики *tracrPНК*:

- Местоположение *tracrPНК* в локусе (зависит от подтипа системы, рис. 1).
- Последовательность *tracrPНК* не должна превышать длину в 150 п.н., длина антиповтора должна равняться приблизительно 24 п.н.
- “Хвостовая” часть *tracrPНК*, идущая за антиповтором, должна складываться в определенные вторичные структуры – так называемый нексус и следующие за ним шпильки, число которых может различаться (рис. 2). Терминальная шпилька *tracrPНК* вместе с примыкающей последовательностью из нескольких (чаще всего пяти) урацилов образует Rho-независимый терминатор транскрипции [23].

Существующие варианты вторичных структур *tracrPНК* подробно представлены в работе S. Dooley и соавт. [21]. Для предсказания и визуализации вторичной структуры РНК можно использовать программу RNAfold [24].

Рекомендуется отбирать кандидатов, для которых биоинформатически предсказываемая *tracrPНК* соответствует всем перечисленным ха-

рактеристикам. Такие кандидаты могут быть использованы для наработки РНК и подтверждения функции в экспериментах *in vitro*.

Если предсказать *tracrPНК* биоинформатически не удастся, можно использовать подход RNAseq (секвенирование РНК): ввести генетическую конструкцию с клонированным локусом изучаемой CRISPR-Cas системы в клетки *Escherichia coli*, а затем выделить из них РНК и провести высокопроизводительное секвенирование. Данный подход позволяет не только обнаружить *tracrPНК* (и *crPНК*) [25], но и корректно определить их границы [26].

10. Для дополнительной оценки обнаруженной CRISPR-Cas системы II типа проверить наличие остальных компонентов, характерных для конкретного подтипа.

Для этого повторить этапы 4, 5 и 6 с HMM-профилями Cas1, Cas2, Csn2, Cas4 (fasta-файлы в zip-архиве S1, см. Дополнительные материалы). Минимальное число кодонов при определении ORF (п. 4): 150 для Cas1, 50 для Cas2, 150 для Csn2 и 80 для Cas4. Этот шаг не считается обязательным, поскольку полное или частичное отсутствие данных компонентов не означает, что предсказанный Cas9-кандидат не обладает нуклеазной активностью.

#### 11. Предсказать структуру зрелой *crPНК*.

Для этого необходимо определить направление транскрипции CRISPR-кассеты. Это можно сделать визуально, поскольку DR уже охарактеризованных CRISPR-Cas систем II типа похожи между собой: например, DR систем II типа бактерий *Neisseria meningitidis* 8013 и *Otariodibacter oris* DSM 23800 различаются всего тремя позициями (табл. S1, см. Дополнительные материалы). Также определению правильного направления DR может помочь предсказанная *tracrPНК* – последовательность DR должна быть способна образовывать дуплекс с участком антиповтора (рис. 2). В случае затруднений при предсказании направления транскрипции CRISPR-кассеты рекомендуется протестировать оба направления DR в экспериментах *in vitro* или воспользоваться подходом RNAseq (см. выше).

В последовательность *crPНК*, как правило, входит лишь часть повтора, комплементарная антиповтору *tracrPНК*. При подборе *crPНК* для *in vitro* проверки можно начать с варианта с полноразмерным повтором, а затем проанализировать несколько укороченных вариантов.

Помимо DR зрелая *crPНК* содержит участок спейсера, комплементарного целевой мишени. Известно, что изменение длины данного участка значительно влияет на эффективность интерференции. Так, например, для нуклеаз SpCas9, SaCas9, Cje1Cas9 и RpCas9 наибольшая эффективность в эукариотических клетках наблюдается

при длинах спейсера 20, 21, 22, 24 п.н. соответственно [14, 27, 28]. В связи с этим рекомендуется тестировать сгРНК с несколькими вариантами длин спейсера.

Программные коды, реализующие отдельные этапы биоинформатического поиска, можно скачать из репозитория GitHub по ссылке [https://github.com/CRISPRNanoBio/Cas9\\_Pipeline](https://github.com/CRISPRNanoBio/Cas9_Pipeline).

По результатам биоинформатического поиска (рис. S1, см. Дополнительные материалы) для дальнейшей экспериментальной работы отбирают те системы-кандидаты, которые соответствуют следующим критериям:

- ORF Cas9 уникальна, то есть кодируемая ею аминокислотная последовательность непохожа на ранее охарактеризованные нуклеазы.
- Система содержит в себе все обязательные компоненты: Cas9, CRISPR-кассета, tracrPНК. Наличие других компонентов системы желательное, но не относится к обязательным условиям.

Организация локуса CRISPR-Cas системы II типа бактерии *Pasteurella pneumotropica* ATCC 35149, определенная по результатам описанного выше биоинформатического анализа, представлена в табл. S2. Дополнительные материалы.

### КЛОНИРОВАНИЕ ГЕНА НУКЛЕАЗЫ-КАНДИДАТА

На следующем этапе работы необходимо получить генетическую конструкцию, несущую ген нуклеазы-кандидата, для дальнейшего получения ее рекомбинантной версии.

1. Получить нуклеотидную последовательность гена нуклеазы-кандидата.

В случае недоступности образцов ДНК с выбранной для изучения системой, например, если система была обнаружена в открытых базах метагеномных данных, необходимо заказать синтетическую последовательность гена нуклеазы-кандидата.

2. Выбрать вектор для клонирования.

Для получения рекомбинантных белков Cas9 рекомендуется использовать экспрессионные векторы, позволяющие проводить дальнейшую хроматографическую очистку целевого белка аффинным методом. Комбинация аффинной хроматографии с последующей очисткой гель-фильтрацией позволяет получить препарат нуклеазы достаточной чистоты для проведения экспериментов *in vitro*.

Для этих целей хорошо подходит вектор pET21a (рис. S2, см. Дополнительные материалы), так как к полилинкеру, используемому для клонирования, примыкает последовательность, кодирующая шесть гистидинов, что позволяет проводить очистку рекомбинантных белков на металл-хелат-

ном аффинном сорбенте (Ni-NTA-агарозе). Транскрипция целевого гена происходит с промотора РНК-полимеразы фага T7. Транскрипция гена РНК-полимеразы и целевого гена регулируется репрессором лактозного оперона.

3. Проанализировать последовательности гена нуклеазы-кандидата и вектора pET21a и выбрать оптимальные для клонирования сайты рестрикции.

Так, например, для гена *ppcas9* были выбраны сайты эндонуклеаз рестрикции XhoI и NdeI.

4. Провести клонирование целевого гена.

Альтернативой классическому подходу с использованием рестрикции и лигирования может быть клонирование по Гибсону, при котором у вектора и вставки есть частично перекрывающиеся последовательности на концах. Использовать NEBuilder HiFi DNA Assembly Master Mix (NEB) согласно протоколу производителя.

5. Верифицировать вставку целевого гена нуклеазы-кандидата в вектор.

Для этого использовать метод рестрикционно-го анализа: контрольный вектор pET21a и полученную на его основе рекомбинантную плазмиду обработать парой эндонуклеаз рестрикции.

В случае обработки вектора pET21a эндонуклеазами рестрикции HindIII и XbaI образуются фрагменты длиной 5300 и 103 п.н.; обработка вектора pET21a с клонированным геном *ppcas9* (pET21a\_PpCas9) этими эндонуклеазами рестрикции приводит к образованию трех фрагментов длиной 5400, 2500 и 630 п.н. (рис. S3, см. Дополнительные материалы).

Для дополнительной верификации провести секвенирование клонированного гена по Сэнгеру.

### НАРАБОТКА РЕКОМБИНАНТНОЙ НУКЛЕАЗЫ-КАНДИДАТА

Получение рекомбинантной версии нуклеазы-кандидата включает в себя следующие основные этапы:

1. Экспрессия гена нуклеазы-кандидата в клетках продуцентах
2. Проверка продукции нуклеазы
3. Проверка растворимости нуклеазы
4. Лизис клеток-продуцентов
5. Аффинная хроматография
6. Гель-фильтрация

Ниже приведены подробные протоколы каждого из этапов.

### *Экспрессия гена нуклеазы-кандидата в клетках-продуцентах*

1. Трансформировать клетки *E. coli* Rosetta (DE3) плазмидой pET21a\_PpCas9 стандартными методами и высеять на чашки Петри, содержащие ампициллин и хлорамфеникол.

2. Перенести единичную колонию в 5 мл среды LB, содержащей ампициллин и хлорамфеникол, и растить в течение ночи при 37°C.

3. Развести 5 мл ночной культуры в 500 мл среды LB, содержащей ампициллин.

4. Инкубировать клеточные культуры при температуре 37°C и равномерной аэрации на вращающейся качалке до достижения оптической плотности 0.6 при длине волны 600 нм.

5. Индуцировать транскрипцию клонированного целевого гена добавлением изопропил- $\beta$ -D-1-тиогалактопиранозид (ИПТГ) до концентрации 1 мМ.

6. Продолжить рост клеточных культур при температуре 20–24°C в течение 4–6 ч.

7. Собрать клетки центрифугированием в настольной центрифуге на скорости 4000–5000 об/мин (~ 3500 g) в течение 30 мин.

Полученные осадки можно сразу использовать для выделения и очистки белка либо заморозить и хранить при температуре –20°C.

### *Проверка продукции нуклеазы-кандидата*

1. Отобрать по 1 мл клеточных культур до внесения индуктора и после инкубации с индуктором (ИПТГ) (п. 4 и п. 6 “*Экспрессия гена нуклеазы-кандидата в клетках-продуцентах*”).

Для корректного сравнения образцов и детекции продукта экспрессии гена важно, чтобы образцы до и после индукции содержали одинаковое число клеток.

2. Собрать клетки центрифугированием в настольной мини-центрифуге на скорости 13000 об/мин (~12000 g) в течение 10 мин.

3. Удалить надосадочную жидкость.

4. Ресуспендировать осадки в небольшом объеме (~100 мкл) фосфатно-солевого буфера (PBS).

5. Отобрать из обеих проб по 15 мкл и проанализировать наличие индуцированного белка при помощи электрофореза в денатурирующем 10%-ном ПААГ по методике Лэммли [29].

Результаты проверки продукции нуклеазы PpCas9 представлены на рис. S4 Дополнительных материалов. Продукт, соответствующий ожидаемой молекулярной массе целевого белка (121 кДа), обнаружен только в образце “После индукции”.

### *Проверка растворимости нуклеазы-кандидата*

Для хроматографической очистки белка Cas9 необходимо убедиться в его растворимости.

1. Подготовить лизисный буфер.

Состав: 50 мМ Трис, 500 мМ NaCl, 10 мМ имидазол, 1 мМ  $\beta$ -меркаптоэтанол, pH 7.5 (при комнатной температуре).

2. Отобрать 1 мл клеточной культуры после инкубации с индуктором (ИПТГ) (п. 6 “*Экспрессия гена нуклеазы-кандидата в клетках-продуцентах*”).

6. Собрать клетки центрифугированием в настольной мини-центрифуге на скорости 13000 об/мин (~12000 g) в течение 10 мин.

7. Удалить надосадочную жидкость.

3. Ресуспендировать осадок в 500 мкл лизисного буфера с добавлением лизоцима (концентрация 1 мг/мл).

4. Инкубировать полученную клеточную суспензию на льду в течение 15 мин.

5. Разрушить клеточные оболочки методом ультразвуковой дезинтеграции на ледяной бане.

Ультразвуковую дезинтеграцию проводили на приборе LABSONIC (“Sartorius AG”, Германия): амплитуда ультразвука 30%, интервал времени активности 0.2 с.

Продолжительность одного цикла обработки – 20 с, с последующим охлаждением в течение 40 с, общее время обработки – 10 мин.

6. Центрифугировать в настольной центрифуге с охлаждением на скорости 13000 об/мин (~16000 g) в течение 20 мин при 4°C.

7. Отобрать надосадочную жидкость.

8. К осадку добавить 500 мкл лизисного буфера, ресуспендировать.

9. Отобрать из ресуспендированного осадка и из надосадочной жидкости по 15 мкл и проанализировать наличие индуцированного белка методом электрофореза в денатурирующем 10%-ном ПААГ по методике Лэммли [29].

Результаты проверки для нуклеазы PpCas9 представлены в Дополнительных материалах, рис. S5. В пробе “Надосадочная жидкость” присутствует продукт, соответствующий ожидаемой молекулярной массе целевого белка, что свидетельствует о том, что полученный белок PpCas9 растворим.

### *Лизис клеток-продуцентов*

1. Осадок, полученный из 500 мл жидкой клеточной культуры, содержащей индуцированный белок, разморозить на льду в течение 30 мин.

2. Ресуспендировать в 15 мл лизисного буфера с добавлением лизоцима (концентрация 1 мг/мл).

3. Инкубировать на льду в течение 30 мин.

4. Разрушить клеточные оболочки методом ультразвуковой дезинтеграции на ледяной бане.

Параметры прибора аналогичны таковым на этапе “Проверка растворимости нуклеазы-кандидата” п. 5. Общее время обработки увеличено и составляет 40 мин.

5. Центрифугировать в настольной центрифуге с охлаждением на скорости 11000 об/мин (~16000 g) в течение 60 мин при 4°C.

6. Надосадочную жидкость профильтровать через шприцевой фильтр с целлюлозно-нитратной мембраной (размер пор 0,4 мкм).

Для выделения функционально активного белка проводят двухстадийную очистку лизата методами аффинной хроматографии и гель-фильтрации. Очистку нуклеазы PpCas9 проводили на приборе АКТА (“GE Healthcare”, США). Можно использовать оборудование других производителей. Для аффинной хроматографии также возможно использование гравитационных колонок.

#### *Аффинная хроматография*

1. Полученный после фильтрации лизат нанести на колонку HisTrap HP 1 mL (“GE Healthcare”), содержащую иммобилизованные ионы никеля, на скорости 0,5 мл/мин.

2. Промыть колонку от неспецифически связанных белков лизисным буфером (5–10 объемов колонки).

3. Промыть колонку лизисным буфером, содержащим 30 мМ имидазол (15 объемов колонки).

4. Элюировать целевой белок лизисным буфером, содержащим 300 мМ имидазол (5 объемов колонки).

#### *Гель-фильтрация*

1. Выбрать фракции, содержащие наибольшее количество белка, и объединить их.

В рассматриваемом примере нуклеаза PpCas9 выходила во фракциях 3 и 4 (рис. S6, см. Дополнительные материалы).

2. При помощи центрифужных концентраторов Amicon (“Merck Millipore”, США) с фильтром на 100 кДа сконцентрировать объединенные фракции до объема 600 мкл.

3. Нанести образец на гель-фильтрационную колонку Superdex 200 Increase 10/300 GL (“Sigma-Aldrich”, США), уравновешенную буфером: 50 мМ Трис-HCl (pH 7,5 при комнатной температуре), 500 мМ NaCl, 1 мМ DTT.

4. Ориентируясь на показания интенсивности поглощения при длине волны 280 нм и известные калибровочные данные используемой хроматографической колонки, собрать фракции элюата, содержащие белок.

5. Проверить наличие и чистоту целевого белка в собранных фракциях методом электрофореза в денатурирующем 10%-ном ПААГ по методике Лэммли [29].

В рассматриваемом примере нуклеазы PpCas9 для электрофореза были взяты фракции 4–6 (рис. S7, см. Дополнительные материалы).

6. Объединить фракции и сконцентрировать при помощи центрифужных концентраторов с фильтром на 100 кДа до конечной концентрации целевого белка 1–3 мг/мл (для используемых объемов жидкой клеточной культуры, указанных выше, конечный объем на данном этапе, как правило, составляет 250–300 мкл).

7. К концентрату добавить глицерин до конечной концентрации 10%.

8. Полученный образец белка разделить на аликвоты объемом 5–10 мкл, распределить их по микробиркам и заморозить в жидком азоте.

9. Очищенный белок хранить при температуре –80°C. Остатки от использованных образцов повторно не замораживать.

Описанные протоколы получения рекомбинантного белка применимы не только для ортологов Cas9-нуклеаз, но и для представителей других типов: например, Cas12a и Cas12e [30].

**Возможные проблемы.** Самая часто встречающаяся проблема связана с нерастворимостью целевой нуклеазы в лизисном буфере (раздел “Проверка растворимости нуклеазы-кандидата”). В некоторых случаях для повышения растворимости целевого белка рекомендуется проводить индукцию при более низких температурах (раздел “Экспрессия гена нуклеазы-кандидата в клетках-производителях”, п. 6). Также решению данной проблемы может способствовать наработка белкового продукта гена нуклеазы-кандидата, слитого с белком, связывающим мальтозу (Maltose-Binding Protein; MBP) [31].

### ДЕМОНСТРАЦИЯ ПРОГРАММИРУЕМОГО РАСЩЕПЛЕНИЯ МИШЕНИ НУКЛЕАЗОЙ-КАНДИДАТОМ И ИДЕНТИФИКАЦИЯ РАМ

#### *In vitro* расщепление РАМ-библиотеки нуклеазой-кандидатом

Как было сказано ранее, для узнавания и расщепления мишени нуклеазой Cas9 последняя должна быть фланкирована определенной последовательностью РАМ (см. “Организация CRISPR-Cas систем II типа”). Понятно, что для нуклеазы кандидата она неизвестна. Поэтому для проверки наличия нуклеазной активности используется специальный набор последовательностей, в котором участок, соответствующий РАМ, заменен случайными нуклеотидами. Такой набор после-



довательностей далее обозначен как “РАМ-библиотека”. Часть последовательностей РАМ-библиотеки будет содержать участок случайных нуклеотидов, совпадающий с РАМ. В результате взаимодействия с эффекторным комплексом нуклеазы такие последовательности подвергнутся расщеплению.

Мы использовали *in vitro* РАМ-библиотеку, содержащую 7 случайных нуклеотидов (7N), так как РАМ-последовательность известных Cas9 чаще всего короче 7 н. (табл. S1, см. Дополнительные материалы).

Анализ реакции расщепления РАМ-библиотеки проводят электрофорезом в агарозном геле. При наличии нуклеазной активности Cas9-кандидата на агарозном геле будут видны фрагменты нескольких размеров. Одни соответствуют последовательностям, которые не подверглись расщеплению, вторые – расщепленным последовательностям. Используемая 7N РАМ-библиотека содержит последовательности длиной 374 п.н. Выбранный нами участок спейсера сгРНК позволяет расщеплять последовательности такой библиотеки на фрагменты длиной приблизительно 49 и 325 п.н. (рис. S8a, см. Дополнительные материалы).

Далее приводится подробный протокол проверки активности нуклеазы-кандидата с использованием 7N РАМ-библиотеки.

#### 1. Подготовить 7N *in vitro* РАМ-библиотеку.

В представленной здесь работе за основу РАМ-библиотеки был взят фрагмент гена *grin2b* человека. Аналогичным образом можно использовать любые нуклеотидные последовательности. Для более удобной визуализации результата на агарозном геле рекомендуется использовать последовательности длиной более 300 п.н.

Семь случайных нуклеотидов вводят в нужную последовательность при помощи амплификации с использованием прямого и обратного праймеров для 7N-библиотеки. Прямой праймер при этом включает участок из 7 случайных нуклеотидов (табл. S2, см. Дополнительные материалы).

Провести очистку целевого фрагмента из агарозного геля.

#### 2. Синтезировать сгРНК и trасгРНК (табл. S3, см. Дополнительные материалы), используя набор для проведения *in vitro* транскрипции HiScribe T7 High Yield RNA Synthesis Kit (“NEB”, США) согласно протоколу производителя.

Последовательность DR сгРНК и последовательность trасгРНК были определены ранее (см. “Биоинформатический поиск новых систем CRISPR-Cas II типа” п. 9 и п. 11). В качестве спейсера сгРНК выбирают область, примыкающую к 7N-участку *in vitro* РАМ-библиотеки с 5'-конца.

Матрицы для синтеза сгРНК и trасгРНК должны содержать последовательность промотора T7 РНК-полимеразы (табл. S2, см. Дополнительные материалы). Матрица может быть либо полностью двухцепочечная, либо одноцепочечная с двухцепочечной областью промотора T7 РНК-полимеразы.

#### 3. Провести очистку полученных в результате *in vitro* транскрипции сгРНК и trасгРНК при помощи набора для очистки РНК – RNA Clean & Concentrator-25 (“Zymo Research”, США) согласно протоколу производителя.

#### 4. Спектрофотометрически измерить концентрации синтезированных сгРНК, trасгРНК и 7N *in vitro* РАМ-библиотеки.

#### 5. Провести *in vitro* реакцию расщепления 7N РАМ-библиотеки изучаемой нуклеазой в объеме 40 мкл.

Для этого объединить: вода (до 40 мкл), 1× CutSmart-буфер (“NEB”), 0.5 мМ DTT, 2 мкМ сгРНК, 2 мкМ trасгРНК, 400 нМ рекомбинантный белок Cas9, 100 нМ 7N *in vitro* РАМ-библиотека.

#### 6. Провести аналогичную реакцию, не добавляя сгРНК (контроль).

#### 7. Обе реакционные смеси инкубировать при 37°C в течение 30 мин.

#### 8. Добавить по 1 мкл протеиназы К (“Thermo Fisher Scientific”, США) для деградации белков, которые могут препятствовать движению ДНК в агарозном геле. Инкубировать смесь при температуре 37°C в течение 20 мин.

#### 9. Контроль и образец проанализировать методом геле-электрофореза в 2%-ном агарозном геле.

**Возможные проблемы.** Мы рекомендуем использовать 7N *in vitro* РАМ-библиотеку. Однако в ряде случаев (например, нуклеаза NmeI Cas9 и ее ближайшие ортологи, а также CjCas9) участок РАМ оказывается длиннее 7 п.н. (5'-NNNN-GATT-3' [32] и 5'-NNNNRYAC-3' [27] соответственно). В таком случае существует вероятность, что РАМ-библиотека не будет содержать последовательностей, соответствующих РАМ исследуемой нуклеазы. Отрицательный результат расщепления РАМ-библиотеки при этом ошибочно будет интерпретирован как отсутствие нуклеазной активности белка-кандидата. В таком случае можно использовать библиотеки с более длинным участком случайных нуклеотидов (8N, 9N, 10N и т.д.). Однако на следующем этапе работы (“Идентификация РАМ”) при использовании таких библиотек потребуется увеличить покрытие высокоэффективного секвенирования.

Эффективность нуклеазной активности белка Cas9 сильно зависит от условий проведения реакции. Мы рекомендуем проводить реакцию расщепления в присутствии ионов магния и калия (входят в состав буфера CutSmart) при температу-

ре 37°C. При отсутствии расщепления РАМ-библиотеки рекомендуется проверить эффективность гидролиза при различных температурах и в присутствии различных ионов (например, марганца или меди). Так, для нуклеазы PpCas9 наиболее эффективное расщепление РАМ-библиотеки зарегистрировано при 42°C (см. [14]).

Также решению проблемы отсутствия расщепления последовательностей библиотеки может способствовать изменение концентраций основных компонентов: белка, сгРНК и/или tracrРНК.

### Идентификация РАМ

Определить консенсус РАМ-последовательности можно по результатам анализа продуктов расщепления РАМ-библиотеки. К сожалению, на практике не удастся воссоздать идеальную ситуацию, в которой все последовательности, содержащие РАМ нуклеазы, расщепляются на 100% и могут быть полностью отделены от последовательностей, не подвергшихся расщеплению. В связи с этим используют подход, включающий анализ двух РАМ-библиотек: исходной (контроль) и библиотеки, часть последовательностей которой подверглась расщеплению нуклеазой (образец). Обе библиотеки проходят одинаковую пробоподготовку, а полученные на них результаты – сравнительный анализ. По результатам этого сравнения отбирают те последовательности, которые вероятнее всего соответствуют РАМ.

Данный подход можно реализовать двумя способами. В первом случае сначала нужно отобрать из образца последовательности, которые подверглись расщеплению. Полученный набор последовательностей далее сравнивают с контролем и отбирают те, доля которых от всех последовательностей в образце достоверно увеличилась по сравнению с их долей в контроле. Такой подход получил название “тест обогащения”. Во втором случае из образца, наоборот, отбирают те последовательности, которые не подверглись расщеплению, а при последующем сравнении с контролем ищут последовательности, доля которых достоверно уменьшилась в образце. Этот подход получил название “деплеционный тест” [33]. В представленной работе мы реализовали один из вариантов *in vitro* деплеционного теста (рис. S8б, см. Дополнительные материалы).

### 7N *in vitro* деплеционный тест

Этот анализ проводят с использованием образца и контроля, полученных на этапе “*In vitro* расщепление РАМ-библиотеки нуклеазой-кандидатом”, п. 9.

1. Часть образца, которая не подверглась расщеплению, и контроль экстрагировать из геля. Длины экстрагируемых последовательностей 374 п.н.

2. Очищенные последовательности подготовить для анализа высокоэффективным секвенированием с помощью двухстадийной ПЦР (праймеры см. в табл. S2 Дополнительных материалов). Праймеры первого этапа содержат последовательности, которые отжигаются к участку РАМ-библиотеки длиной 92 п.н., включающему 7 случайных нуклеотидов и 20 нуклеотидов мишени, комплементарной сгРНК. Дополнительно в праймеры включены фрагменты адаптеров Illumina (“Illumina”, США). Полученные на этом этапе ампликоны используют в качестве матрицы для второй ПЦР. На втором этапе адаптеры полностью достраивают и вносят последовательности, содержащие 6 или 8 рандомизированных нуклеотидов (баркоды).

3. Последовательности проанализировать на платформе Illumina одиночными или парными прочтениями с минимальной длиной прочтения 75 п.н. с каждой стороны и минимальным покрытием 10 млн прочтений на образец.

Парные прочтения в данном случае будут практически полностью перекрывать друг друга, что позволит исправить возможные ошибки секвенирования. Для дальнейшей работы пары прочтений необходимо объединить и построить соответствующие консенсусы.

4. Из полученных прочтений (либо консенсусов прямых и обратных прочтений) вырезать участки, соответствующие положению 7 случайных нуклеотидов.

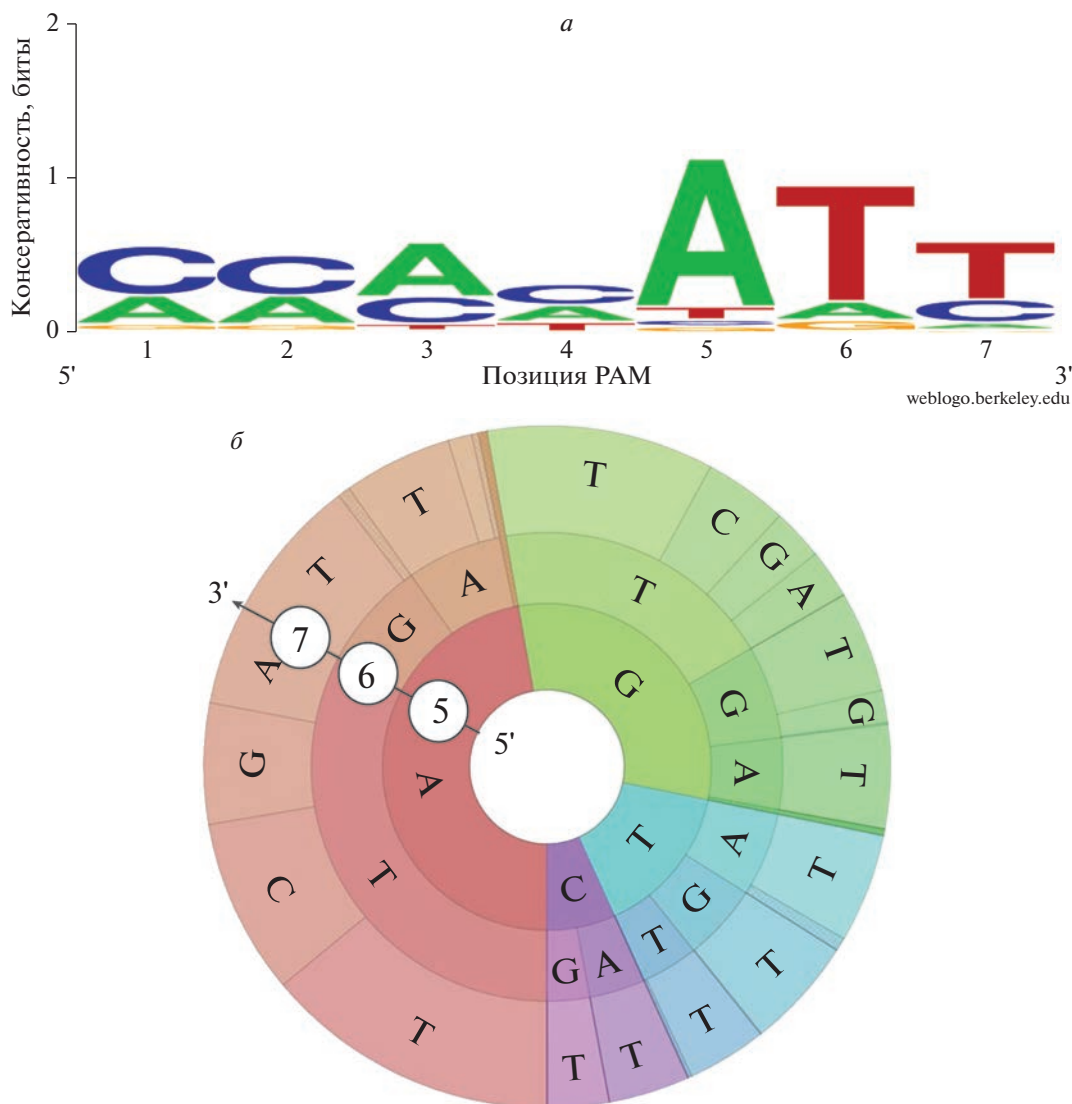
5. Сравнить наборы последовательностей для контроля и образца и визуализировать полученный результат.

### Визуализация результата *in vitro* деплеционного теста с помощью сервиса WebLogo [34]

В случае деплеционного теста РАМ-лого представляет собой консенсус последовательностей, частота встречаемости которых уменьшилась в образце по сравнению с контролем (деплецированные последовательности). Те позиции на полученном лого, которые характеризуются наибольшей консервативностью, выраженной в битах, рассматривают как возможные значимые позиции РАМ.

1. При помощи критерия  $\chi^2$  сравнить наборы последовательностей, соответствующих 7N-участку РАМ-библиотеки, из контроля и образца и отобрать те, которые достоверно деплецировались.

Мы рекомендуем использовать пороговое значение  $p\text{-value} \leq 10^{-12}$ . Однако получаемые в этом тесте данные содержат большое число как ложноположительных, так и ложноотрицательных результатов. Именно поэтому возможно увеличение или наоборот уменьшение порогового значения  $p\text{-value}$  для получения лого с высококонсервативными отдельными позициями РАМ.



**Рис. 3.** Результаты 7N *in vitro* деплеционного теста для нуклеазы PpCas9. Визуализация результатов в виде PAM-лого для позиций 1–7 PAM (а) и в виде PAM-колеса для позиций 5–7 PAM (б). Адаптировано из работы [14, Creative Commons Attribution License].

2. Построить PAM-лого из деплецированных последовательностей с использованием сервиса WebLogo (<https://weblogo.berkeley.edu/logo.cgi>).

По лого, полученном по результатам 7N *in vitro* деплеционного теста для нуклеазы PpCas9, можно предположить значимость позиций 5, 6, 7 от конца мишени (рис. 3а).

Аналогично для нуклеазы SauriCas9 по лого была выявлена консервативность позиций 3 и 4 PAM и преобладание доли G в них ([35]).

#### *Визуализация результата *in vitro* деплеционного теста в виде PAM-колеса [36]*

PAM-колесо (PAM wheel) строят с использованием всех вариантов последовательностей, соответствующих 7N-участку PAM-библиотеки, и ин-

формации по частоте их встречаемости в образце и контроле. Это отличает PAM-колесо от PAM-лого, для построения которого берут только деплецированные последовательности.

Однако, у PAM-колеса есть ограничение – результаты удобно визуализировать только для небольшого числа позиций PAM (до четырех). Таким образом, этот инструмент можно считать дополнением к уже построенному PAM-лого.

1. Из имеющихся наборов последовательностей, соответствующих 7N-участку PAM-библиотеки, необходимо отобрать только интересующие позиции (до четырех), которые могут иметь значимость согласно PAM-лого. После этого необходимо уточнить, сколько раз встречается каждый новый вариант последовательности.

Так, для нуклеазы PrCas9 отбирают позиции 5, 6 и 7. Например, если в изначальном наборе присутствует одна последовательность **CCCCATT** и одна последовательность **GGGGATT**, в новый набор необходимо внести две последовательности **ATT**.

2. Для каждого нового варианта последовательности, соответствующей участку случайных нуклеотидов, рассчитать коэффициент деплеции ( $D_{\text{coef}}$ ):

$$3. D_{\text{coef}} = \log_2 \frac{N_{\text{sample}} N_{\text{seq\_var.control}}}{N_{\text{control}} N_{\text{seq\_var.sample}}},$$

где  $N_{\text{sample}}$  – общее число последовательностей в образце;  $N_{\text{control}}$  – общее число последовательностей в контроле;  $N_{\text{seq\_var.sample}}$  – число последовательностей определенного варианта в образце;  $N_{\text{seq\_var.control}}$  – число последовательностей определенного варианта в контроле.

В исходном методе используют “тест обогащения” и рассчитывают соответствующий коэффициент обогащения. Таким образом,  $D_{\text{coef}}$  рассчитывается как коэффициент, обратный рассчитанному Leenay и др. [36].

4. Полученные значения коэффициентов записать в таблицу специального формата в файл KrgonaExcelTemplate (табл. S4, см. Дополнительные материалы), в котором проводят построение РАМ-колеса.

Дополнительная проверка результатов с построением РАМ-колеса для нуклеазы PrCas9 подтверждает вероятную значимость позиций 5, 6 и 7. Также можно предположить, что присутствие нуклеотидов А либо G равновероятно допустимо в позиции 5 РАМ (рис. 3б). По результатам построения РАМ-лого и РАМ-колеса для нуклеазы PrCas9 можно сделать вывод о том, что ее РАМ соответствует консенсусу 5'-NNNNR~~TT~~-3'.

Для нуклеазы FrCas9, с РАМ-последовательностью 5'-NNTA-3', при визуализации РАМ-колеса для позиций 2, 3 и 4 РАМ четко видно преобладание Т в третьем положении и А в четвертом ([37]).

*Экспериментальное подтверждение предсказанной РАМ-последовательности*

Проверку значимости отдельных нуклеотидных позиций проводят по представленной ниже методике.

1. Среди наиболее деплецированных последовательностей, отобранных для построения РАМ-лого, выбрать одну.

Для нуклеазы PrCas9 была выбрана последовательность **СААСАТТ** (рис. 4).

Далее необходимо проверить, что выбранная последовательность распознается нуклеазой как РАМ.

2. Синтезировать ДНК-мишень по аналогии с получением 7N-библиотеки для *in vitro* деплеционного теста, заменив в прямом праймере рандомизированный участок случайных нуклеотидов на последовательность, обратнoкомплементарную выбранному варианту РАМ.

3. Провести *in vitro* реакцию расщепления полученного фрагмента в объеме 20 мкл.

Для этого объединить: вода (до 20 мкл), 1× CutSmart-буфер (“NEB”), 0.5 мМ ДТТ, 2 мкМ crРНК, 2 мкМ tracrРНК, 400 нМ рекомбинантный белок Cas9, 40 нМ фрагмент ДНК с вариантом РАМ из п. 1.

4. Провести контрольную реакцию без добавления crРНК.

5. Обе реакционные смеси инкубировать при 37°C в течение 30 мин (либо при той температуре, которая обеспечивает наиболее эффективное расщепление мишеней нуклеазой).

6. Добавить по 1 мкл протеиназы К (“Thermo Fisher Scientific”). Инкубировать при 37°C в течение 20 мин.

7. Продукты реакций проанализировать методом гель-электрофореза в 2%-ном агарозном геле.

Дальнейшие этапы проводить, если фрагмент ДНК, содержащий выбранный вариант последовательности РАМ, подвергся расщеплению. В случае отсутствия расщепления необходимо вернуться к п. 1 и выбрать другую последовательность.

8. Составить набор последовательностей, которые могут быть модификациями выбранного варианта.

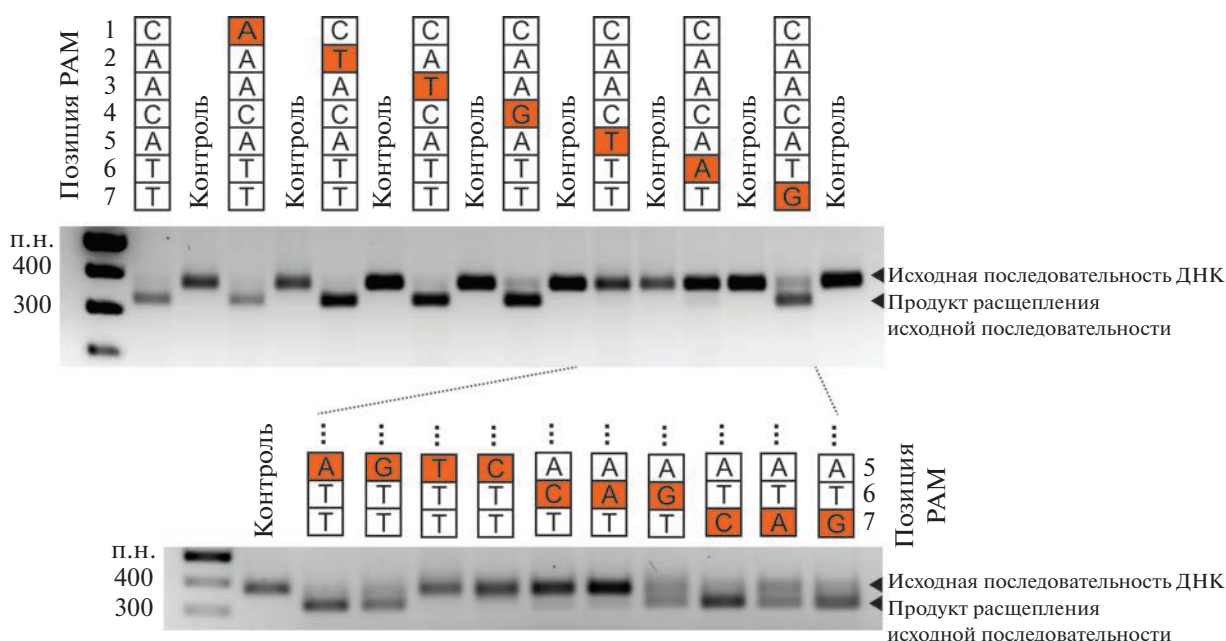
Каждая последовательность такого набора включает замену одной нуклеотидной позиции; при этом пуриновые основания заменить на пиримидиновые и наоборот.

Для PrCas9 был создан следующий набор из 7 последовательностей:

- СААСАТТ** (исходный вариант)
- АААСАТТ**
- СТАСАТТ**
- САТСАТТ**
- СААГАТТ**
- СААСТТТ**
- СААСААТ**
- СААСАТG**

9. Для набора полученных последовательностей повторить п. 2–7.

По отсутствию расщепления нуклеазой последовательности из данного набора можно судить о значимости соответствующей позиции РАМ.



**Рис. 4.** Проверка значимости отдельных нуклеотидных позиций в PAM-последовательности PpCas9. Для позиций 5, 6 и 7 проведена дополнительная проверка влияния замен на все четыре варианта нуклеотидов (нижняя часть рисунка). Замены в позиции 5 на Т или С и любые замены в позиции 6 блокируют или значительно снижают расщепление соответствующей последовательности нуклеазой. Замены в позиции 7 снижают эффективность расщепления. Адаптировано из работы [14, Creative Commons Attribution License].

Дополнительно для наиболее значимых позиций PAM-последовательности (или для всех позиций) можно проверить, как замены на все возможные варианты нуклеотидов влияют на эффективность распознавания и расщепления нуклеазой мишени.

Для нуклеазы PpCas9 вышеописанная проверка показала, что значимы нуклеотиды в позициях 5, 6 и 7 PAM. Также показано, что замена А на G в позиции 5 не влияет на эффективность расщепления. Таким образом, удалось подтвердить, что PAM нуклеазы PpCas9 соответствует консенсусу 5'-NNNNRTT-3', предсказанному по результатам деплеционного теста (рис. 4).

Для дополнительной проверки соответствия PAM-последовательности нуклеазы предсказанному консенсусу также рекомендуется проверить эффективность расщепления различных мишеней, которые фланкированы этим консенсусом ([14]). Реакцию проводят с любым фрагментом ДНК, содержащим участки, совпадающие с консенсусом PAM (условия реакции аналогичны тем, которые используются при проверке значимости отдельных нуклеотидных позиций).

## ЗАКЛЮЧЕНИЕ

Мы многократно проверили приведенные протоколы на практике и доказали их эффективность. Используя их, мы успешно охарактеризо-

вали ряд новых CRISPR-Cas редакторов: нуклеазы PpCas9 из бактерии *Pasteurella pneumotropica* [14], CcCas9 из бактерии *Clostridium cellulolyticum* H10 [38], DfCas9 из бактерии *Deftluviimonas* sp. 20V17 [14] и CoCas9 из бактерии *Capnocytophaga ochracea* (в печати).

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (грант № 075-15-2021-1062). Исследования А.А.В., П.А.С. и А.Н.А. также поддержаны грантом Российского научного фонда (грант № 21-14-00122).

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

Авторы заявляют об отсутствии конфликта интересов.

## СПИСОК ЛИТЕРАТУРЫ

1. Jansen R., van Embden J.D.A., Gastra W., Schouls L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575.
2. Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., Horvath P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. **315**, 1709–1712.
3. Marraffini L.A., Sontheimer E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*. **322**, 1843–1845.

4. Garneau J.E., Dupuis M.-È., Villion M., Romero D.A., Barrangou R., Boyaval P., Fremaux C., Horvath P., Magadán A.H., Moineau S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. **468**, 67–71.
5. Makarova K.S., Wolf Y.I., Iranzo J., Shmakov S.A., Alkhnbashi O.S., Brouns S.J.J., Charpentier E., Cheng D., Haft D.H., Horvath P., Moineau S., Mojica F.J.M., Scott D., Shah S.A., Siksnys V., Terns M.P., Venclovas Č., White M.F., Yakunin A.F., Yan W., Zhang F., Garrett R.A., Backofen R., van der Oost J., Barrangou R., Koonin E.V. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83.
6. Cong L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., Hsu P.D., Wu X., Jiang W., Marraffini L.A., Zhang F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*. **339**, 819–823.
7. Jinek M., East A., Cheng A., Lin S., Ma E., Doudna J. (2013) RNA-programmed genome editing in human cells. *Elife*. **2**, e00471.
8. Mali P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*. **339**, 823–826.
9. Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Piszczak Z.A., Eckert M.R., Vogel J., Charpentier E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. **471**, 602–607.
10. Shah S.A., Erdmann S., Mojica F.J.M., Garrett R.A. (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* **10**, 891–899.
11. Luscombe N.M., Laskowski R.A., Thornton J.M. (2001) Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res.* **29**, 2860–2874.
12. Chatterjee P., Lee J., Nip L., Koseki S.R.T., Tysinger E., Sontheimer E.J., Jacobson J.M., Jakimo N. (2020) A Cas9 with PAM recognition for adenine dinucleotide. *Nat. Commun.* **11**, 2474.
13. Wei J., Hou L., Liu J., Wang Z., Gao S., Qi T., Gao S., Sun S., Wang Y. (2022) Closely related type II-C Cas9 orthologs recognize diverse PAMs. *Elife*. **11**, e77825.
14. Fedorova I., Vasileva A., Selkova P., Abramova M., Arseniev A., Pobegalov G., Kazalov M., Musharova O., Goryanin I., Artamonova D., Zyubko T., Shmakov S., Artamonova T., Khodorkovskii M., Severinov K. (2020) PpCas9 from *Pasteurella pneumotropica* – a compact type II-C Cas9 ortholog active in human cells. *Nucleic Acids Res.* **48**, 12297–12309.
15. Bland C., Ramsey T.L., Sabree F., Lowe M., Brown K., Kyrpides N.C., Hugenholtz P. (2007) CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*. **8**, 209.
16. Couvin D., Bernheim A., Toffano-Nioche C., Touchon M., Michalik J., Néron B., Rocha E.P.C., Vergnaud G., Gautheret D., Pourcel C. (2018) CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251.
17. Edgar R.C. (2007) PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics*. **8**, 18.
18. Biswas A., Staals R.H.J., Morales S.E., Fineran P.C., Brown C.M. (2016) CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics*. **17**, i356–i356.
19. Mitrofanov A., Alkhnbashi O.S., Shmakov S.A., Makarova K.S., Koonin E.V., Backofen R. (2021) CRISPRidentify: identification of CRISPR arrays using machine learning approach. *Nucleic Acids Res.* **49**, e20.
20. Mistry J., Finn R.D., Eddy S.R., Bateman A., Punta M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121.
21. Dooley S.K., Baken E.K., Moss W.N., Howe A., Young J.K. (2021) Identification and evolution of Cas9 tracrRNAs. *CRISPR J.* **4**, 438–447.
22. Mitrofanov A., Ziemann M., Alkhnbashi O.S., Hess W.R., Backofen R. (2021) CRISPRtracrRNA: robust approach for CRISPR tracrRNA detection. *Bioinformatics*. **38**, ii42–ii48.
23. Livny J., Waldor M.K. (2007) Identification of small RNAs in diverse bacterial species. *Curr. Opin. Microbiol.* **10**, 96–101.
24. Lorenz R., Bernhart S.H., zu Siederdissen C.H., Tafer H., Flamm C., Stadler P.F., Hofacker I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26.
25. Harrington L.B., Burstein D., Chen J.S., Paez-Espino D., Ma E., Witte I.P., Cofsky J.C., Kyrpides N.C., Banfield J.F., Doudna J.A. (2019) Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*. **362**, 839–842.
26. Zhang Y., Heidrich N., Ampattu B.J., Gunderson C.W., Seifert H.S., Schoen C., Vogel J., Sontheimer E.J. (2013) Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol. Cell*. **50**, 488–503.
27. Ran F.A., Cong L., Yan W.X., Scott D.A., Gootenberg J.S., Kriz A.J., Zetsche B., Shalem O., Wu X., Makarova K.S., Koonin E.V., Sharp P.A., Zhang F. (2015) *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature*. **520**, 186–191.
28. Kim E., Koo T., Park S.W., Kim D., Kim K., Cho H.-Y., Song D.W., Lee K.J., Jung M.H., Kim S., Kim J.H., Kim J.H., Kim J.-S. (2017) *In vivo* genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500.
29. Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227**, 680–685.
30. Selkova P., Vasileva A., Pobegalov G., Musharova O., Arseniev A., Kazalov M., Zyubko T., Shcheglova N., Artamonova T., Khodorkovskii M., Severinov K., Fedorova I. (2020) Position of *Deltaproteobacteria* Cas12e nuclease cleavage sites depends on spacer length of guide RNA. *RNA Biol.* **17**, 1472–1479.
31. Sun P., Tropea J.E., Waugh D.S. (2011) Enhancing the solubility of recombinant proteins in *Escherichia coli* by

- using hexahistidine-tagged maltose-binding protein as a fusion partner. *Methods Mol. Biol.* **705**, 259–274.
32. Esvelt K.M., Mali P., Braff J.L., Moosburner M., Yaung S.J., Church G.M. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods.* **10**, 1116–1121.
  33. Zetsche B., Gootenberg J.S., Abudayyeh O.O., Slaymaker I.M., Makarova K.S., Essletzbichler P., Volz S.E., Joung J., van der Oost J., Regev A., Koonin E.V., Zhang F. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell.* **163**, 759–771.
  34. Crooks G.E., Hon G., Chandonia J.-M., Brenner S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.
  35. Hu Z., Wang S., Zhang C., Gao N., Li M., Wang D., Wang D., Liu D., Liu H., Ong S.-G., Wang H., Wang Y. (2020) A compact Cas9 ortholog from *Staphylococcus auricularis* (SauriCas9) expands the DNA targeting scope. *PLoS Biol.* **18**, e3000686.
  36. Leenay R.T., Maksimchuk K.R., Slotkowski R.A., Agrawal R.N., Gomaa A.A., Briner A.E., Barrangou R., Beisel C.L. (2016) Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell.* **62**, 137–47.
  37. Cui Z., Tian R., Huang Z., Jin Z., Li L., Liu J., Huang Z., Xie H., Liu D., Mo H., Zhou R., Lang B., Meng B., Weng H., Hu Z. (2022) FrCas9 is a CRISPR/Cas9 system with high editing efficiency and fidelity. *Nat. Commun.* **13**, 1425.
  38. Fedorova I., Arseniev A., Selkova P., Pobegalov G., Goryanin I., Vasileva A., Musharova O., Abramova M., Kazalov M., Zyubko T., Artamonova T., Artamonova D., Shmakov S., Khodorkovskii M., Severinov K. (2020) DNA targeting by *Clostridium cellulolyticum* CRISPR-Cas9 type II-C system. *Nucleic Acids Res.* **48**, 2026–2034.

## Type II CRISPR-Cas System Nucleases: a Pipeline for Prediction and *in vitro* Characterization

**A. A. Vasileva<sup>1, 2, \*</sup>, S. A. Aliukas<sup>3</sup>, P. A. Selkova<sup>1, 2</sup>, A. N. Arseniev<sup>1, 2</sup>, V. E. Chernova<sup>2</sup>,  
O. S. Musharova<sup>1</sup>, E. I. Klimuk<sup>1</sup>, M. A. Khodorkovskii<sup>2</sup>, and K. V. Severinov<sup>1, 4, \*\*</sup>**

<sup>1</sup>*Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, 123182 Russia*

<sup>2</sup>*Peter the Great St. Petersburg Polytechnic University, St. Petersburg, 195251 Russia*

<sup>3</sup>*LLC “Biotech campus”, Moscow, 117437 Russia*

<sup>4</sup>*Waksman Institute of Microbiology, Piscataway, NJ, 08854 USA*

\**e-mail: daucussativus7@gmail.com*

\*\**e-mail: severik@waksman.rutgers.edu*

The use of CRISPR-Cas bacterial adaptive immunity systems components for targeted DNA changing has opened broad prospects for programmable genome editing of higher organisms. The most widely used gene editors are based on the Cas9 effectors of the type II CRISPR-Cas systems. In complex with guide RNAs, Cas9 proteins are able to directionally introduce double-strand breaks into DNA regions complementary to guide RNA sequences. Despite the wide range of characterized Cas9s, the search for new Cas9 variants remains an actual task, since the available Cas9 editors have several limitations. This paper presents a workflow for the search and subsequent characterization of new Cas9 nucleases developed in our laboratory. Detailed protocols describing the bioinformatical search, cloning and isolation of recombinant Cas9 proteins, testing for nuclease activity *in vitro*, and determining the PAM sequence required for recognition of DNA targets, are presented. Potential difficulties that may arise, as well as ways to overcome them, are considered.

**Keywords:** CRISPR, Cas9, nuclease, genome editing, bioinformatics search