

УДК 577.2.08

CRISPR/CAS9 – КОРОЛЬ ГЕНОМНОГО РЕДАКТИРОВАНИЯ

© 2017 г. А. В. Банников^{а, *}, А. В. Лавров^{а, б}

^аМедико-генетический научный центр, Москва, 115478

^бРоссийский национальный исследовательский университет им. Н.И. Пирогова, Москва, 117997

*e-mail: broyler3@mail.ru

Поступила в редакцию 15.06.2016 г.

Принята к печати 27.10.2016 г.

С открытием CRISPR/Cas9 появилась надежда на долгожданное обретение эффективного, надежного и доступного метода редактирования генома. Доступность CRISPR/Cas9 не вызывает сомнений, но эффективность и надежность до сих пор остаются предметом изучения. В представленном обзоре описаны принципы работы нуклеаз системы Cas. Рассмотрены также основные аспекты, которые необходимо учитывать при планировании работ с использованием CRISPR/Cas9 – методы оценки эффективности и специфичности Cas9, подходы к выбору sgRNA, генно-инженерные варианты Cas9, использование гомологичной рекомбинации и негомологичного соединения концов при редактировании ДНК.

Ключевые слова: CRISPR/Cas9, геномное редактирование, sgRNA

DOI: 10.7868/S0026898417040036

ВВЕДЕНИЕ

Впервые элементы системы CRISPR/Cas были описаны в 1987 году [1]. Долгое время роль обнаруженных повторяющихся последовательностей в ДНК бактерий оставалась неизвестной. В 2007 году установили, что нуклеазы семейства Cas и элементы CRISPR участвуют в противовирусном иммунитете бактерий [2]. Позднее аналогичную систему обнаружили у архей [3]. Эти работы стали основой для масштабного биоинформатического поиска подобных повторов в геномах различных видов организмов [4]. CRISPR/Cas распространена достаточно широко – она обнаружена у 48% бактерий и 84% архей.

Нуклеазы служат основным инструментом генной терапии. Идея генной терапии проста – в заданное место генома вносят двухцепочечный разрыв (DSB), репарация которого осуществляется с использованием двух основных механизмов – негомологичного соединения концов (non homologous end joining – NHEJ) и гомологичной репарации (homologous dependent repair – HDR). При NHEJ концы разрыва сшиваются друг с другом, при этом возможно появление делеций и вставок одного или нескольких нуклеотидов. При HDR используется шаблон, по которому копируется

локус вокруг DSB. Таким шаблоном может быть неповрежденная вторая хромосома или экзогенная ДНК, например плазида или олигонуклеотид. Таким образом, формируя DSB в определенном месте, а затем репарируя его с помощью экзогенной ДНК, можно исправлять патогенные мутации или изменять геном, внося новые последовательности в место разрыва. Именно это составляет сущность генной терапии. Специфические программируемые нуклеазы используются для редактирования генома достаточно давно, но Cas9 обходит всех конкурентов по одному важному параметру – цене. Перепрограммирование ранее известных нуклеаз требует дизайна новой аминокислотной последовательности, трудоемкого процесса без гарантии успеха. Возможности и преимущества разных нуклеаз обсуждены ранее [5]. Для перепрограммирования Cas9 нужно синтезировать лишь нуклеотидную последовательность (ДНК или РНК в зависимости от метода доставки) длиной около 20 н., что можно сделать быстро, дешево и с высокой точностью.

ПРИРОДНЫЕ СИСТЕМЫ CRISPR/CAS

Системы CRISPR/Cas делятся на шесть типов I–VI. Два основных компонента системы – оперон,

Сокращения: NGS (Next Generation Sequencing) – массовое параллельное секвенирование; NHEJ (Non Homologous End Joining) – негомологичное соединение концов ДНК; HDR (Homologous Dependent Repair) – гомологичная репарация ДНК; DSB (Double Stranded Break) – двухцепочечный разрыв ДНК; CRISPR (Clustered Regularly-Interspaced Short Palindromic Repeats) – регулярно перемежающиеся короткие палиндромные повторы; PAM (Protospacer Adjacent Motif) – мотив, прилежащий к протоспейсеру; crRNA – CRISPR РНК; sgRNA (single guide RNA) – направляющая РНК; tracrRNA (trans-activating crRNA) – трансактивирующая crРНК; Cas (CRISPR associated protein) – белок ассоциированный с CRISPR; spCas9 – Cas9 *Streptococcus pyogenes*.

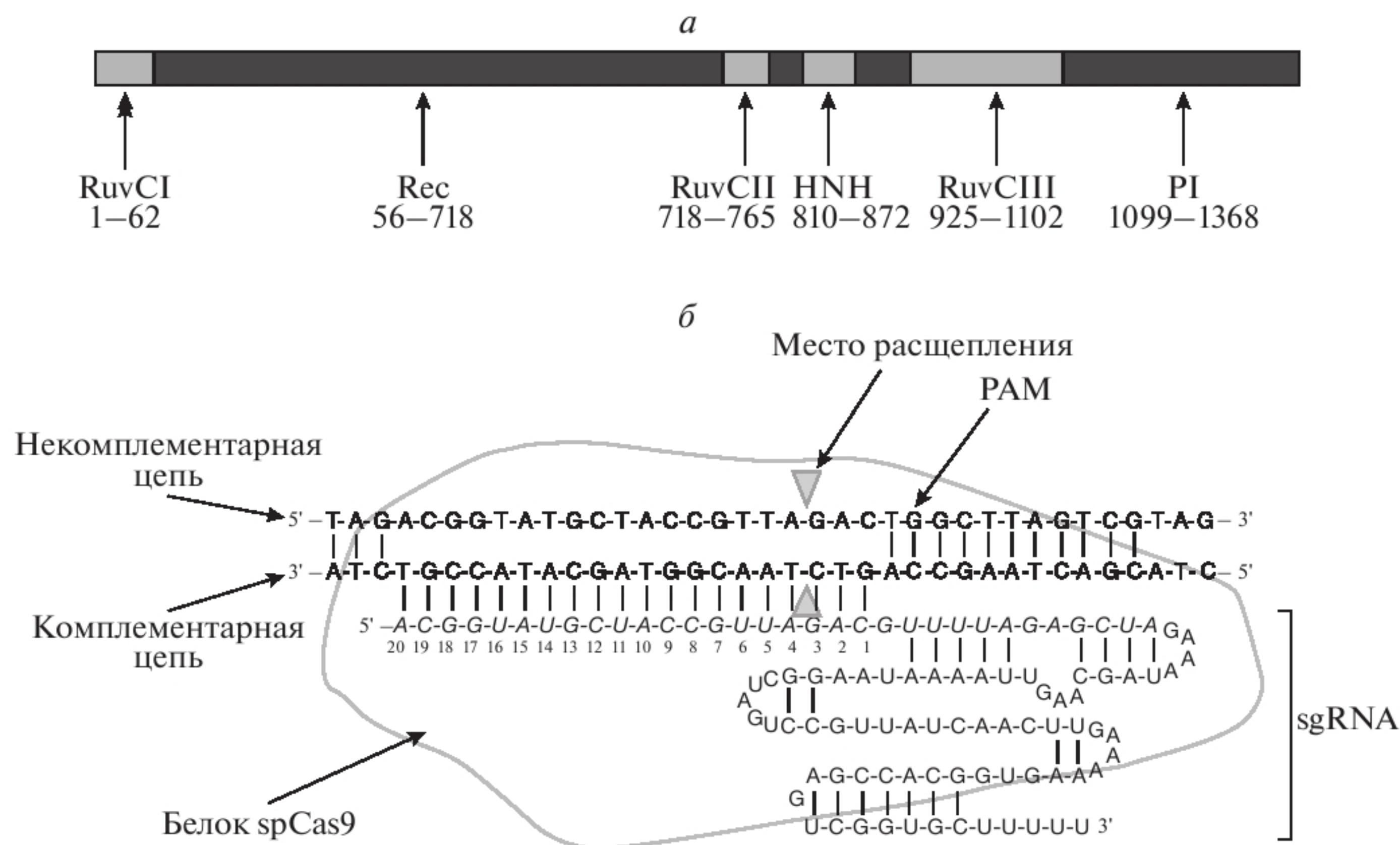


Рис. 1. Доменная структура spCas9 (*a*) и схема связывания spCas9 с целевой ДНК (*б*).

с которого синтезируются белок/ки Cas, и CRISPR – регулярно перемежающиеся короткие палиндромные повторы, с которых синтезируются РНК, направляющие нуклеазную активность, так называемые sgRNA [4]. В разных системах один или несколько белков Cas выполняют следующие основные функции: адаптерную – для связывания с sgRNA; РНКазную – для созревания sgRNA; и нуклеазную – для расщепления целевой ДНК. Все белки, обозначенные Cas, принадлежат к 10 семействам, Cas1–10.

Локус CRISPR состоит из палиндромных повторов и спейсеров. Спейсеры образуются в результате включения в геном хозяина фрагментов вирусных или плазмидных геномов, которые до встраивания называют протоспейсерами. С этого массива синтезируется sgRNA, содержащая спейсер. Затем sgRNA присоединяется непосредственно к белкам Cas систем I и III или с участием вспомогательной tracrRNA к белку Cas системы II. Специфичность взаимодействия комплекса Cas-РНК с ДНК обеспечивается гибридизацией между спейсером sgRNA и комплементарной так называемой целевой ДНК. Таким образом, нуклеаза Cas направляется на протоспейсер, присутствующий в вирусном геноме. В зависимости от типа Cas протоспейсер должен быть фланкирован с 5'- или 3'-конца мотивом PAM (protospacer adjacent motif). Обычно PAM, абсолютно необходимый для связывания Cas с ДНК, состоит из 3–5 н. Cas9 из *Streptococcus pyogenes* (spCas9), узнает PAM NGG, а Cas9 из *Staphylococcus aureus* (saCas9) – NNGRRT (рис. 1) [6]. PAM не обладают полной специфичностью, например, spCas9 узнает также PAM NAG, хотя и менее эффективно, чем NGG. Соот-

ветственно, PAM присутствует в вирусной ДНК, но этого мотива нет в массиве CRISPR, что обеспечивает защиту генома бактерий от расщепления Cas. Такой механизм используют системы I и II. В системе III ДНК, кодирующая спейсеры, фланкирована специальными повторами, защищающими ее от расщепления.

Системы I и III сходны по механизму действия и составу. Вокруг sgRNA собирается комплекс из адаптерных белков. В системе III нуклеаза (Cas10) присоединяется сразу, а в системе I (Cas3) – только после связывания с ДНК. Трехмерные структуры комплексов CRISPR/Cas I и III также очень сходны. Наиболее примечательной чертой этих структур является взаимодействие адаптерного белка Cas7 со спейсером, при котором несколько белков Cas7 полимеризуются с образованием цепи, которая присоединяется параллельно к спейсеру sgRNA. В комплексе с Cas7 sgRNA достаточно сильно изогнута, поэтому при взаимодействии с ДНК гибридизация происходит постепенно, примерно по 5 н. (по мере того как Cas7 теряет контакт с sgRNA). Считается, что такой механизм имеет большую точность (меньшую чувствительность к отсутствию комплементарности отдельных нуклеотидов), нежели единовременная гибридизация [6].

Система CRISPR/Cas II значительно отличается от систем I и III. В состав системы II входит только один белок, Cas9, который взаимодействует с sgRNA, а также имеет два нуклеазных домена, RuvC и HNH (рис. 1). В структуре Cas9 можно различить две достаточно хорошо обособленные доли, одна из которых содержит все нуклеазные домены, за исключением части HNH, а

другая – связывающие. crRNA присоединяется между двумя этими долями [7]. Спейсер взаимодействует с богатой аргинином α -спиралью, соединяющей две доли Cas9. Это взаимодействие наиболее сильно в начале спейсера, что, вероятно, служит одной из причин особенной важности так называемой “seed sequence” – 4 н. на 3'-конце спейсера, для узнавания нуклеотидной последовательности ДНК. В данном случае более сильное взаимодействие обеспечивает более точное позиционирование этих нуклеотидов при контакте с ДНК, с них начинается процесс гибридизации, который, вероятно, протекает постепенно, как и в системах I и III. Направляющая РНК (guide RNA – gRNA) системы II состоит из двух отдельных молекул РНК, crRNA и tracrRNA, которые присоединяются к Cas9 только после гибридизации друг с другом.

CRISPR/Cas IV, V и VI открыты относительно недавно и менее изучены. Система IV не содержит CRISPR и направляется молекулой ДНК [8]. Система V использует РНК длиной 41–42 н. для направления нуклеазной активности; эта РНК соответствует crRNA Cas9 [9]. В системе VI также используется короткая направляющая РНК, но расщепляется не ДНК, а РНК [10].

Более подробно принципы функционирования CRISPR-иммунитета и варианты систем разных микроорганизмов описаны в обзоре [11].

В настоящее время для геномного редактирования широко используют только Cas9. Остальные системы не так хорошо изучены, но имеют некоторые особенности, позволяющие в перспективе видеть их равными Cas9.

МОЛЕКУЛЯРНЫЕ МЕХАНИЗМЫ ФУНКЦИОНИРОВАНИЯ Cas9

Важным достижением в области геномного редактирования стало создание sgRNA (single guide RNA) [12]. Части crRNA и tracrRNA были искусственно объединены в единую РНК, которую назвали sgRNA и использовали для направления Cas9 без уменьшения ее активности. sgRNA состоит примерно из 80 н., из которых 20 спейсер. Таким образом, замена всего 20 н. позволяет направлять Cas9 на новый сайт ДНК.

В эукариотических клетках Cas9 экспрессируют, как правило, со стандартных промоторов, цитомегаловирусного (CMV) для белка и U6 (промотор полимеразы III) для sgRNA. Гены транскрибируются в ядре, после чего мРНК Cas9 экспортируется в цитозоль, где транслируется. sgRNA остается в ядре, так как не имеет сигналов экспорта. Cas9 содержит сигнал ядерной локализации – NLS (nuclear localization signal), благодаря которому белок возвращается в ядро. На данный момент известно, что NLS не требуется для ядерного импорта Cas9 [13, 14]. Это объясняют вы-

соким положительным зарядом самой нуклеазы. Однако, например, получена аллостерически регулируемая 5-гидрокситамоксифеном Cas9, активность которой в отсутствие 5-гидрокситамоксифена была ниже, чем у белка дикого типа [14]. В отсутствие NLS активность аллостерической Cas9 снижалась еще сильнее. По-видимому, ядерный транспорт Cas9 становится лимитирующим лишь при низкой активности или низкой концентрации нуклеазы. Попав в ядро, Cas9 соединяется с sgRNA, в результате чего образуется рибонуклеопротеиновый комплекс, способный специфически расщеплять ДНК.

Cas9 находит сайты связывания с ДНК, сталкиваясь с ней в процессе диффузии. Скольжения Cas9 по ДНК не наблюдали, что не удивительно, учитывая высокую специфичность взаимодействия Cas9 с ДНК, в первую очередь, с PAM. Тем не менее, Cas9 может связываться с ДНК и без sgRNA, но это взаимодействие неспецифично – константа диссоциации составляет около 25 нМ в то время как с sgRNA константа диссоциации уменьшается до 0.5 нМ [15].

При специфическом взаимодействии сначала Cas9 узнает PAM. Затем Cas9 связывается с ДНК, взаимодействие с PAM активирует геликазную активность, и начинается плавление цепи ДНК. Вероятнее всего, PAM нужен именно для активации/обеспечения геликазной активности Cas9. Показано, что если два первых нуклеотида протоспейсера постоянно расплавлены, то в PAM уже нет необходимости [15]. Область “seed sequence” играет в данном случае важную роль, стабилизируя взаимодействие Cas9 с начавшей плавиться ДНК.

Эффективность связывания Cas9 с ДНК зависит от доступности целевого сайта. Геном эукариот обладает разным уровнем компактизации, что влияет на доступность сайтов. По всей видимости, *in vivo* Cas9 способна инициировать декомпактизацию хроматина [16]. В *in vitro* экспериментах показано, что даже нуклеосомы могут в 6–8 раз снижать активность Cas9 [17]. О взаимодействии Cas9 с хроматином известно немного, но можно достаточно уверенно утверждать, что компактизация хроматина действительно влияет на эффективность редактирования генома. Это подтверждается также методом library-on-library, подробно рассмотренным в разделе “Выбор sgRNA”.

Получено достаточно много свидетельств того, что Cas9 очень прочно связывается с целевой ДНК [18, 19]. Время жизни комплекса с ДНК измеряется часами, Cas9 не диссоциирует из комплекса даже в присутствии 0.5 М NaCl [15]. Некоторые характеризуют Cas9 как однооборотный фермент (single turn over enzyme). В пользу этого говорит еще одно наблюдение: клетке млекопитающего требуется примерно 1 ч, чтобы исправить ущерб от воздействия УФ-излучения, и око-

до 15 ч для репарации DSB, созданных Cas9 [20]. Это может объясняться тем, что после расщепления ДНК Cas9 еще долго находится на месте и защищает концы разорванной ДНК от опознавания системой репарации.

К сожалению, специфичность связывания Cas9 с ДНК не является абсолютной. Нуклеаза может связываться не только с целевым сайтом, но и с нецелевыми сайтами, которые незначительно отличаются от целевого, но при этом содержат PAM. Вероятность расщепления нецелевого сайта зависит, в первую очередь, от количества замен, их характера и положения. Собрана статистическая информация о влиянии этих и некоторых других факторов на специфичность взаимодействия с ДНК и ее расщепления комплексом Cas9-sgRNA. Это будет рассмотрено в разделе, посвященном дизайну sgRNA.

МЕТОДЫ ОЦЕНКИ СПЕЦИФИЧНОСТИ CRISPR/Cas9

Как уже сказано, специфичность CRISPR/Cas9 не является абсолютной, что недопустимо для некоторых важных приложений, например генной терапии, поэтому одна из первых задач — увеличение специфичности. Это дало толчок к разработке методов картирования DSB на уровне полного генома. Методы картирования DSB незаменимы при оценке специфичности нуклеаз, поскольку позволяют объективно оценивать нецелевую активность. Предложено несколько методов картирования DSB, каждый из которых был применен в небольшом количестве работ. Два метода картирования DSB, GUIDE-seq и BLESS, можно назвать наиболее перспективными.

GUIDE-seq (Genome-wide Unbiased Identification of DSBs Enabled by Sequencing — полногеномный непредвзятый поиск DSB методом секвенирования)

Для проведения GUIDE-seq [21] клетки выращивают в среде с короткими двухцепочечными олигодезоксирибонуклеотидами (dsODN), защищенными фосфотиоатом, которые встраиваются в DSB, образованные Cas9, при репарации ДНК методом NHEJ. При этом метка (dsODN) может содержать сайт рестрикции, чтобы включение метки можно было оценить с помощью анализа полиморфизма длин рестрикционных фрагментов. Эти dsODN используют для секвенирования и картирования DSB. Для приготовления библиотеки ДНК для NGS-секвенирования выделенную геномную ДНК разрушают до среднего размера 500 п.н., затем к одному концу пришивают адаптер. Амплификацию проводят с использованием праймеров на адаптер и метку, поэтому в библиотеку попадают только участки вблизи DSB. Представленность последовательности в библиотеке зависит от эффективности включе-

ния метки. Если проанализировать результаты GUIDE-seq [22], то корреляция между числом прочтений и эффективностью интеграции метки во все сайты в пяти генах составляет всего 0.43. При анализе сайтов только гена *FANCF* корреляция была намного выше, около 0.9. По-видимому, на представленность последовательности в библиотеке влияет не только интеграция метки, но и другие факторы. Наиболее вероятной причиной является ПЦР. Зависимость эффективности ПЦР от содержания GC-нуклеотидов — давно обсуждаемая проблема подготовки библиотек для NGS.

Безусловно, GUIDE-seq — это чувствительный метод, поскольку интеграция метки может произойти в любое время существования DSB. Однако количество прочтений в результатах GUIDE-seq не отражает прямо вероятность образования DSB на данном месте. Результаты GUIDE-seq по одному сайту (с использованием одной sgRNA) можно использовать для количественного сравнения активности разных нуклеаз, например разных мутантов Cas9. Однако нельзя проводить прямое количественное сравнение двух разных сайтов.

BLESS (direct in situ breaks labeling, enrichment on streptavidin, and next-generation sequencing — прямое мечение разрывов in situ, обогащение стрептавидином и NGS)

В случае метода BLESS прикрепление метки к месту DSB происходит после экстракции хроматина [23]. Использование биотинилированной метки позволяет обогатить препарат последовательностями с DSB. Затем ДНК разрушают и лигируют вторую метку. После этого проводят ПЦР с праймерами на использованные метки.

Метод BLESS значительно менее чувствителен, чем GUIDE-seq, так как позволяет картировать только DSB, присутствующие в момент выделения ДНК. Редко возникающие DSB, вероятнее всего, будут репарированы, поэтому не появятся в результатах BLESS. На данный момент есть возможность сравнить чувствительность этих методов для двух sgRNA. В обоих случаях GUIDE-seq определяет больше нецелевых сайтов — 8 и 4 для одной sgRNA и 32 и 10 для второй [22, 24].

Другие методы

Еще один высокопроизводительный метод оценки DSB — высокопроизводительное полногеномное секвенирование транслокаций (HTGTS, high throughput genome wide translocation sequencing) [25], и его вариант LAM-HTGTS (linear amplification mediated HTGTS — HTGTS, опосредованное линейной амплификацией). Сначала в геном клеток вводят так называемый константный сайт рестрикции, который *in vivo* будет постоянно расщеплен. При репарации DSB в любом

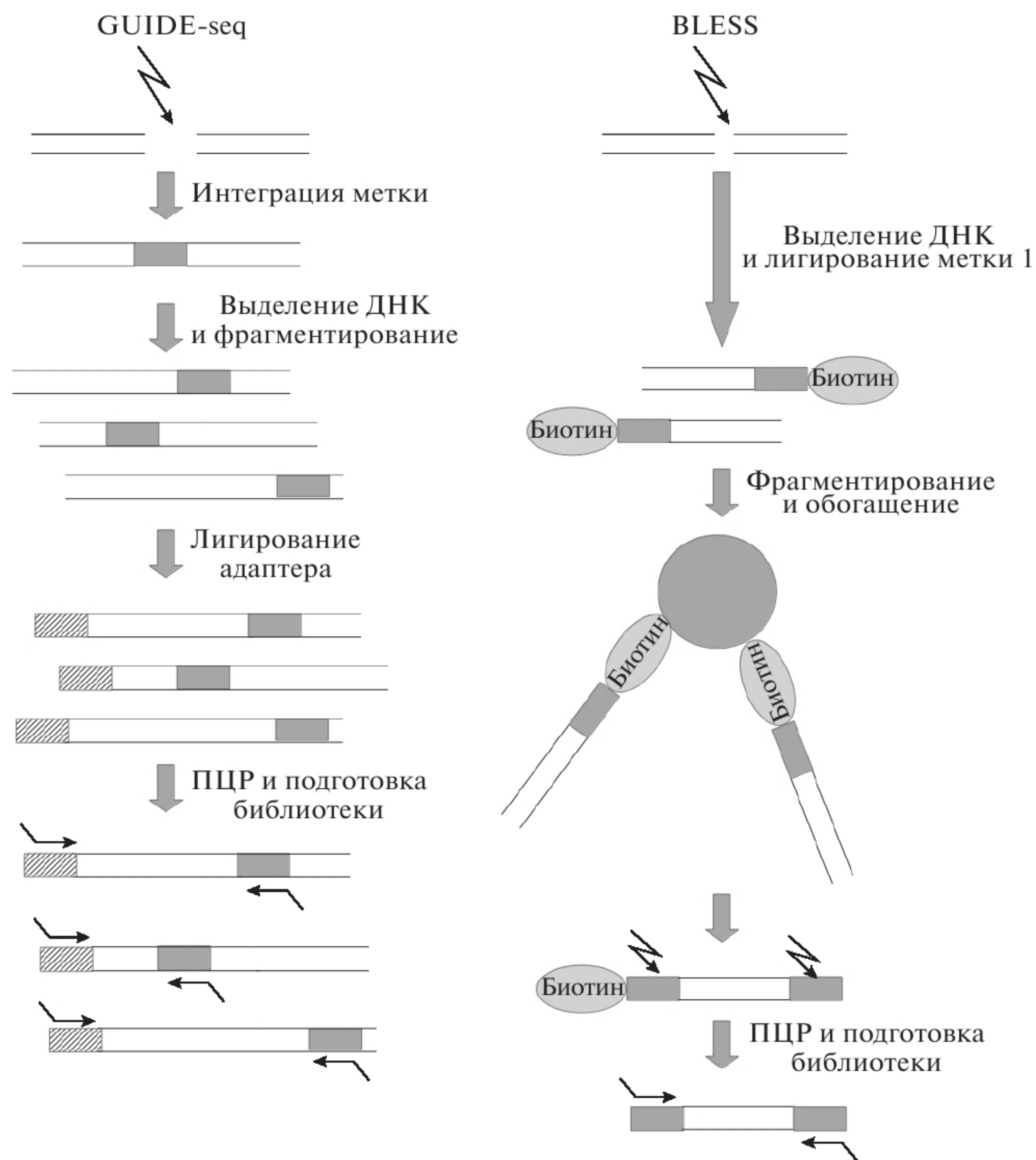


Рис. 2. Схема проведения GUIDE-seq и BLESS.

месте генома существует вероятность его лигирования к константному DSB, т.е. транслокации. Зная последовательность локуса константного DSB, можно с помощью NGS картировать любые DSB в геноме. Транслокация — это событие, примерно в 200–1000 раз более редкое, чем делеции и вставки при NHEJ, поэтому для (LAM)-HTGTS необходимо анализировать большее количество ДНК. Кроме того, изначально HTGTS был разработан для анализа образования DSB в генах иммуноглобулинов, где можно разместить константный сайт вблизи изучаемых DSB, что достаточно важно, поскольку вероятность внутривромосомных транслокаций выше, чем межвромосомных. Из-за низкой чувствительности методы определения активности нуклеаз, основанные на транслокации, вытесняются такими конкурентами, как GUIDE-seq.

Среди методов оценки активности нуклеаз невозможно не упомянуть T7 и Surveyor [26], расщепляющих ДНК в месте отсутствия гибридизации, на чем и основан принцип их действия. Из клеток, трансфицированных плазмидой с Cas9 и соответствующей sgRNA, выделяют ДНК. Область редактирования затем амплифицируют с помощью обычной ПЦР. При этом часть локусов содержит мутации — результат NHEJ, часть — нет. Если затем регибризовать этот ПЦР-продукт, то мутантные цепи будут иметь примерно равную вероятность соединиться с цепями дикого типа. В месте мутации такие дубликаты содержат негибридизованные основания и, следовательно, могут расщепляться T7 или Surveyor. Затем проводят анализ фрагментов. Максимальное число расщепленных дубликатов получено при 50%-ной эффективности редактирования. Широкое применение T7 и Surveyor обусловлено их относительно низкой стоимостью, но возможности этих

методов ограничены тем, что DSB можно оценить только на уже известных сайтах, и нижний порог чувствительности составляет несколько процентов, значительно уступая NGS.

ИНЖЕНЕРИЯ Cas9

Инженерия Cas9 направлена в первую очередь на уменьшение нецелевой активности Cas9. Генно-инженерные варианты Cas9 можно разделить на две группы:

- (1) мутантные белки – все варианты, полученные заменой одной или нескольких аминокислот;
- (2) разделенные Cas9 – нуклеаза экспрессируется в виде двух пептидов, которые затем собираются в единую нуклеазу.

Мутантные белки

К настоящему времени для уменьшения нецелевой активности получены и охарактеризованы мутантные варианты Cas9 [22, 24].

С этой целью была модифицирована геликазная активность Cas9 [24]. В процессе плавления ДНК как sgRNA, так и геликазный домен Cas9 могут стабилизировать одноцепочечное состояние ДНК. С уменьшением геликазной активности относительный вклад sgRNA в стабилизацию оцДНК увеличивается, что предположительно должно снижать чувствительность Cas9 к неполной комплементарности между sgRNA и целевой ДНК. Всего было мутировано 32 аминокислотных остатка в различных сочетаниях, из них у двух вариантов, K810A/K1003A/R1060A и K848A/K1003A/R1060A, увеличилась специфичность [24]. Активность Cas9 в этой работе оценивали с помощью BLESS.

Устранение неспецифических контактов должно уменьшать силу взаимодействия с нецелевыми сайтами существенно сильнее, чем с целевыми. Поэтому цель работы [22] состояла в уменьшении неспецифических электростатических взаимодействий Cas9 с ДНК. Всего получено и протестировано четыре варианта Cas9: HF1 (N497A/R661A/Q695A/Q926A), HF2 (D1135E/N497A/R661A/Q695A/Q926A), HF3 (L169A/N497A/R661A/Q695A/Q926A), HF4 (Y450A/N497A/R661A/Q695A/Q926A) [22]. По результатам анализа кристаллической структуры Cas9 были выбраны аминокислотные остатки, мутированные в HF1. Остальные варианты несли по одной охарактеризованной ранее дополнительной мутации. Наиболее вероятно, что с PAM взаимодействует аминокислотный остаток D1135 [27]; уменьшение нецелевой активности происходит за счет снижения аффинности к PAM, при этом “seed sequence” начинает играть большую роль в связывании с ДНК. L169 участвует в Ван-дер-Ваальсовых взаимодействиях с ДНК [10] и, предположительно, позво-

ляет различать ДНК и РНК, Y450 смещается на 120 градусов при связывании с ДНК [27] и, вероятнее всего, участвует в гидрофобном взаимодействии. Активность Cas9 анализировали с помощью GUIDE-seq.

Безусловно, исчерпаны не все возможности мутагенеза Cas9. Как минимум необходимо проверить комбинированный эффект мутаций, полученных в двух работах. Во-вторых, следует отметить, что оценка специфичности новой мутантной нуклеазы зависит от методов, которые нельзя назвать устоявшимися. Описанные работы хорошо иллюстрируют это – в первой был использован BLESS, во второй – GUIDE-seq. Оба метода позволяют уверенно утверждать, что мутантная нуклеаза лучше фермента дикого типа, но абсолютная специфичность нуклеазы в силу большей чувствительности точнее отражается GUIDE-seq. Выбирая нуклеазу для генной терапии, скорее следует доверять результатам GUIDE-seq, но и в этом случае не обходится без проблем.

Cas9 можно превратить в никазу, инактивировав один из нуклеазных доменов [28]. Только использование двух никаз, в одной из которых инактивирован HNH, а в другой RuvC, может привести к образованию DSB. Использование никаза уменьшает нецелевую активность, так как для образования DSB необходимо одновременное присоединение двух белков. Но дизайн всей системы при этом усложняется, что ограничивает применение никаза.

Полезными могут быть также варианты Cas9, использующие другие PAM. Это позволило бы свободнее выбирать целевые сайты. Так, с помощью направленной эволюции созданы три варианта Cas9: D1135V/R1335Q/T1337R для PAM NGAN; D1135E/R1335Q/T1337R для PAM NGAG; и D1135V/G1218R/R1335E/T1337R для PAM NGCG [29]. Использование таких ферментов может быть единственным решением при необходимости внесения разреза в места, недоступные другим Cas9.

Разделенные Cas9

Wright A.V. и соавт. [30] разделили Cas9 на нуклеазный и связывающий домены, которые экспрессировали по отдельности и соединили за счет нековалентных взаимодействий. В *in vitro* экспериментах показано, что разделенная версия Cas9 работает примерно в 10 раз медленнее, однако максимальный уровень активности у нее такой же, как у фермента дикого типа. *In vivo* разделенная Cas9 оказалась намного менее эффективной – 0.6 и 2% против 22 и 34% в двух экспериментах, что объяснялось возможной неравной экспрессией доменов и относительно высокой экспрессией sgRNA с секвестрированием доменов отдельно

друг от друга. В пользу этой гипотезы говорит тот факт, что увеличение концентрации sgRNA ведет к снижению эффективности Cas9.

Улучшить связывание двух разделенных доменов Cas9 можно с помощью лиганд-зависимой димеризации. С этой целью к одному домену Cas9 присоединили FKBP (FK506 binding protein 12), к другому — домен, связывающий FKBP [31]. Димеризация этих доменов возможна в присутствии рапамицина. Таким образом была получена индуцируемая разделенная Cas9. Однако активность неиндуцированной системы оставалась высокой — примерно 1/3 индуцированной активности, поэтому к С-домену Cas9 добавили два NLS, а к N-домену сигнал экспорта из ядра (NES, nuclear exporting signal). После трансляции домены попадают в разные клеточные компартменты. Если домены Cas9 димеризуются в цитозоле, то два NLS обеспечивают транспорт в ядро, несмотря на присутствие NES. Использование такой конструкции приводило к отсутствию неиндуцированной активности. Активность димеризованной Cas9 оказалась примерно в 2 раза ниже, чем у фермента дикого типа — 95 и 43% на целевом сайте, нецелевая активность при этом отсутствовала. Интересно, что активация рапамицином была устойчивой: эффект однократной активации в течение 2 ч был таким же, как при непрерывной активации.

Для сшивания доменов Cas9 использовали также интеины — короткие аминокислотные последовательности, обладающие автокаталитической пептидазной активностью [32]. Интеин состоит из двух частей, С- и N-концевых, которые могут принадлежать одной (*цис*-интеины) или разным (*транс*-интеины) белковым цепям. *Цис*-интеины могут без участия каких-либо кофакторов вырезаться из белка. Половины *транс*-интеинов сначала ковалентно связываются друг с другом, а затем вырезаются подобно *цис*-интеинам.

Эффективность Cas9, собранной с помощью интеинов, была такой же, как у фермента дикого типа [33]. Интеиновую Cas9 удалось доставить в клетки с помощью аденоассоциированного вируса (adeno-associated virus — AAV). В настоящее время AAV считается наиболее безопасным вирусным вектором, но его вместимость ограничена примерно 4.7 т.п.н. Размер полной spCas9-системы чуть превышает этот лимит, что затрудняет ее использование, например ограничивает выбор регуляторных последовательностей. Интеиновая Cas9 решает проблему упаковки — система доставляется в двух вирусных частицах.

Новые Cas9

Важной задачей является доставка Cas9 в клетку. Cas9 имеет размер, неудобный для переноса в эффективном и безопасном AAV, что заставляет

искать более короткие варианты фермента. Наиболее перспективной можно назвать saCas9 из *S. aureus*. Длина кДНК saCas9 всего 3252 п.н., по некоторым данным фермент обладает большей точностью, чем spCas9. Согласно результатам определения нецелевой эффективности методом BLESS saCas9 является более точным ферментом, чем spCas9 дикого типа [34], но менее точным, чем высокоспецифичная мутантная spCas9(1.1) [24]. Методом GUIDE-seq оценили специфичность редактирования spCas9 (мутантной и дикого типа) и saCas9 целевого сайта в гене *VEGFA* [22, 35]. Выявили восемь нецелевых сайтов у saCas9 и более 30 у spCas9, т.е. показали, что saCas9 обладает более высокой специфичностью.

Таким образом, saCas9 может иметь и другие преимущества, кроме меньшей длины. Однако опубликованных данных недостаточно, чтобы это утверждать.

ВЫБОР sgRNA

Известно, что нуклеотидная последовательность sgRNA влияет как на эффективность расщепления целевого сайта ДНК, так и на нецелевую активность. Собраны статистические данные, которые помогают предсказывать эффективность sgRNA по ее последовательности, а также данные о влиянии количества, характера и положения некомплементарных нуклеотидов на эффективность sgRNA, что позволяет предсказывать нецелевую активность.

Во всех работах эффективность sgRNA приравнивают к эффективности нокаута гена, против которого направлена sgRNA. Нокаут возможен, когда DSB репарируется путем NHEJ, при этом часто наблюдается потеря нескольких нуклеотидов и значительно реже — вставка. Смещение рамки считывания на 1, 2, 3 нуклеотида имеет равную вероятность, т.е. примерно в 2/3 случаев происходит сдвиг рамки, приводящий к нокауту гена [36]. При проведении такого эксперимента клетки обрабатывают библиотекой ДНК, состоящей из тысяч и десятков тысяч sgRNA, направленных против генов, нокаут которых достаточно легко проанализировать. Это могут быть гены поверхностных рецепторов, экспрессию которых можно оценить с помощью проточной цитофлуориметрии, или жизненно важные гены, эффективность нокаута которых обратно пропорциональна выживаемости клеток. Клетки обрабатывают библиотекой ДНК таким образом, чтобы в среднем одна клетка содержала только одну sgRNA. Например, с целью выявления генов, участвующих в репарации некомплементарных оснований (MMR — mismatch repair), проанализированы 73300 sgRNA, направленных против 7330 генов — по 10 sgRNA на ген [37]. Клетки, трансфицированные системой CRISPR/Cas9, культивировали в присутствии

6-тиогуанина в концентрации, вызывающей их гибель, поскольку белки репарации распознают 6-тиогуанин и вызывают остановку клеточного цикла. sgRNA к генам пути MMR нарушают контроль клеточного цикла, и клетки продолжают делиться в присутствии 6-тиогуанина. С помощью NGS-анализа выживших клеток идентифицированы sgRNA, а следовательно и гены, нокаут которых приводит к выключению пути MMR. Анализ эффективности разных sgRNA, направленных против одного гена, позволил установить связь между специфическими мотивами в sgRNA и их эффективностью. Все эти данные используют для изучения зависимости эффективности sgRNA от их последовательности.

Эффективность sgRNA всегда определяется несколькими факторами – внутриклеточной устойчивостью, сродством к Cas9 и, наконец, последовательностью самой sgRNA. Детальному анализу и разграничению роли этих факторов посвящены несколько работ. В одной из них мРНК Cas9 и sgRNA инъецировали в эмбрионы полосатого данио [36]. При инъекции только sgRNA оценивали ее внутриклеточную стабильность, а при совместной инъекции с Cas9 – аффинность к нуклеазе (учитывая ранее измеренную устойчивость). Обнаружили, что sgRNA тем устойчивее, чем больше в ней остатков G. Это обусловлено экспериментально подтвержденным формированием квадруплексных структур. А аффинность к Cas9, напротив, выше у sgRNA с низким содержанием G. Несмотря на это нуклеотидный состав эффективных sgRNA был сходен с составом стабильных sgRNA, т.е. в данном случае существует антагонизм между стабильностью и аффинностью, высоким и низким содержанием G, но стабильность оказывается более важным фактором. Аналогичные результаты получены и другими исследователями [38]. У всех активных sgRNA содержание GC было больше 50%. Следует отметить, что метод доставки компонентов CRISPR/Cas9 должен отражаться на получаемых результатах. Например, в случае инъекции устойчивость sgRNA намного важнее, чем при ее непрерывной экспрессии. Это приводит к доминированию соответствующих характеристик “эффективных” sgRNA.

Еще один важный момент – доступность целевого сайта, уровень его компактизации. Фактор организации хроматина учли [39], применив так называемый метод библиотека на библиотеке (library-on-library). Суть этого метода проста – при обычном скрининге используется лишь библиотека sgRNA, целевые сайты расположены в геноме. В методе library-on-library используется вторая библиотека – библиотека целевых сайтов, т.е. клетки трансфицируют плазмидами не только с компонентами системы Cas9, но и с целевыми сайтами. Обе библиотеки доставляют в клетки в составе лентивирусных векторов, которые встраиваются в области с высокой доступностью хро-

матина. Оценивая активность на интегрированном целевом сайте, можно исключить влияние эпигенетических факторов, а сравнивая эффект на интегрированном и геномном сайтах, оценить их вклад в активность sgRNA. Полученные данные говорят о роли эпигенетических факторов в снижении активности Cas9, что подтверждается корреляцией между активностью sgRNA и чувствительностью к ДНКазе I. Однако связь между хроматинизацией и активностью sgRNA наблюдали не во всех работах [38].

Еще одна сложность заключается в том, что допущение равной вероятности смещения рамки считывания на 1, 2, 3 н. как минимум не всегда справедливо [40]. Отмечена выраженная предрасположенность DSB к тому или иному исходу репарации в зависимости от положения в геноме [36, 41]. Например, если сайт репарируется в основном с делецией 3 н., то эффективность нокаута гена, а, следовательно, и кажущаяся эффективность sgRNA будут низкими. Насколько это влияет на измеряемую активность sgRNA неизвестно.

Программа sgRNA designer

Существуют специализированные программы, позволяющие подбирать sgRNA с учетом различных параметров, влияющих как на целевую, так и на нецелевую активность sgRNA. Все доступные на данный момент программы дизайна sgRNA описаны в обзоре [42]. В качестве примера детально рассмотрим сбор данных для одного из широко используемых инструментов – программы sgRNA designer, разработанной в Broad Institute (США) и доступной по ссылке <http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design>.

В первой работе по сбору экспериментальных данных для создания sgRNA designer анализировали влияние последовательности 1841 sgRNA, направленных на гены поверхностных рецепторов, на целевую активность [43]. Экспрессию целевых генов определяли методом проточной цитофлуориметрии. Используя собранные данные, сформулировали так называемый набор правил № 1 для подбора sgRNA. Затем эти правила были расширены, в них включили дополнительные факторы [44]. Был приведен также индекс Gini – мера важности данного фактора для определения активности sgRNA: характер нуклеотидов и динуклеотидов в зависимости от позиции (58%), общее число нуклеотидов и динуклеотидов (16%), положение sgRNA в гене белка (13%), температура плавления (11%). Кроме того, с целью определения нецелевой активности проанализировали влияние некомплементарных оснований на активность sgRNA, которое суммируется в индексе CFD (cutting frequency determination – фактор частоты расщепления). Статистические данные для CFD собирали с использованием 65 sgRNA, направленных на ген *CD33*, активность которых под-

тверждена экспериментально. Получены sgRNA со всеми возможными модификациями — однонуклеотидными заменами, делециями и вставками — всего 9914 sgRNA. Влияние замены в данной позиции определяли как среднее уменьшение активности по всем sgRNA, содержащим данную замену. Например, если нужно определить влияние замены G>T в позиции 5, то усредняют изменение активности всех sgRNA с такой заменой, даже если они содержат и другие замены. Для оценки эффективности индекса CFD провели кросс-валидацию на 89 sgRNA, т.е. вычислили коэффициент корреляции между индексом CFD для sgRNA и ее экспериментально определенной активностью. У sgRNA с одной заменой корреляция составила 0.572, с двумя — 0.512, с тремя и более — 0.164.

Результаты, полученные с использованием sgRNA designer, подтверждены и в других работах. Были проанализированы три набора sgRNA: 2073 sgRNA против 58 рибосомных генов человека, 1667 sgRNA против 163 нерибосомных генов человека и 1443 sgRNA против нерибосомных генов мыши [45]. Хорошее соответствие для всех трех наборов получено только в области “seed sequence”, где в позициях 1 и 2 предпочтителен G, в позиции 3 — C, T был нежелателен в первых четырех нуклеотидах спейсера.

Три упомянутые работы [43–45] содержат наиболее значительный набор статистических данных об активности sgRNA. Многие доступные в интернете программы дизайна sgRNA используют именно эти результаты, например CHOP-CHOP [46] и WU-Crispr [47].

Собранные данные не позволяют выработать четкие правила выбора sgRNA. Использование доступных программ, по-видимому, оправдано при дизайне большой библиотеки, но чаще необходимо выбирать всего из нескольких sgRNA. В этом случае лучшим решением будет экспериментальная проверка.

NHEJ или HDR?

Известно, что Cas9 вызывает DSB, который может репарироваться в клетке с использованием двух основных механизмов — NHEJ и HDR. NHEJ часто сопровождается делециями и вставками, которые, если находятся в кодирующей области гена, могут вести к сдвигу рамки считывания и, следовательно, к нокауту гена. Создание модифицированных клеточных линий с помощью CRISPR/Cas прекрасно описано в протоколах из лаборатории F. Zhang [48]. HDR протекает с участием шаблона, который копируется во время репарации. В качестве шаблона при проведении генной терапии может использоваться плазмидная ДНК с соответствующими участками гомологии, около 1000 п.н. с каждой стороны от DSB или более короткая оцДНК, общая длина которой не

превышает 120 н. Таким образом, с помощью HDR можно вносить вставку и изменять геномную последовательность. Использование HDR принципиально сложнее, чем NHEJ. Гомологичная репарация DSB *in vivo* значительно менее вероятна. Проблема “усиления” HDR изучена недостаточно, поэтому эффективность редактирования с помощью HDR чаще всего исчисляется единицами процентов.

Эффективность Cas9 изучают с использованием разных методов оценки NHEJ и HDR. Наиболее распространен анализ флуоресценции, когда HDR и/или NHEJ восстанавливают работу флуоресцентного белка, например traffic light reporter (TLR), включающего два белка — mCherry и GFP. Ген *mCherry* содержит мутацию со сдвигом рамки, поэтому может репарироваться NHEJ, таким образом содержание (%) клеток mCherry+ составляет примерно 1/3 клеток, репарированных NHEJ. Ген *GFP* содержит делецию, поэтому репарировается только HDR.

Предложен метод оценки эффективности NHEJ и HDR, основанный на цифровой ПЦР [49]. В этом методе используют два зонда: один для детекции NHEJ (этот зонд несет флуорофор 1, он покрывает область разреза и полностью комплементарен ей), другой зонд — для HDR (находится в области репарации, он комплементарен репарированной последовательности и несет флуорофор 2). Третий зонд находится вне области редактирования и служит для подсчета сайтов (несет флуорофор 1). NHEJ-зонд не присоединяется к последовательности с делециями/вставками, таким образом, уменьшение флуоресценции 1 является мерой NHEJ, а увеличение флуоресценции 2 — мерой HDR.

Следует также обращать внимание на селективный отбор трансфицированных клеток. В большинстве работ Cas9 трансфицируют одновременно с селективным маркером устойчивости к антибиотикам, в дальнейшем используют только клетки, экспрессирующие Cas9 [50].

Низкая эффективность HDR препятствует получению стабильно трансфицированных клеточных линий, поэтому для вставки больших генов предложено использовать NHEJ [51]. В этом случае донором служит линейаризованная плаزمид с тупыми концами, которая вставляется в DSB по механизму NHEJ. Эффективность вставки гена *GFP* с помощью HDR в восьми клеточных линиях колебалась от 1.5 до 6%. Использование NHEJ повысило эффективность до 15–20%. Эффективность встраивания больших фрагментов размером 12 и 34 т.п.н. составила 7.49 и 1.18% соответственно.

К настоящему времени в попытках увеличить эффективность HDR использовали два подхода — выключение генов белков, участвующих в NHEJ, и синхронизацию экспрессии Cas9 с клеточным

циклом. Главным белком, инициирующим NHEJ, считается Ku70, который присоединяется к концам разрыва и блокирует доступ факторов, необходимых для HDR. Гены *Ku70* и ДНК-лигазы IV (*LIG4*), сшивающей концы разрыва при NHEJ, “выключали” с помощью РНК-интерференции, низкомолекулярного ингибитора SCR7 [50] и белковых ингибиторов. Одновременное подавление Ku70 и LIG4 привело к значительному, в 5–7 раз, увеличению HDR: с 5 до 25–36%. Эффект SCR7 не был подтвержден [52], однако использование SCR7 при инъекции компонентов Cas в зиготу крысы позволило повысить эффективность HDR примерно в 2 раза [53].

Второй подход основан на зависимости режима репарации от фазы клеточного цикла. HDR происходит в S- и G2-фазах, тогда как активность NHEJ в клеточном цикле не меняется. Ограничение экспрессии Cas9 S- и G2-фазами может увеличить шансы HDR по сравнению с NHEJ. Многие контролируемые клеточным циклом белки содержат сигналы деградации (убиквитинирования), которые позволяют гидролизовать белок при наступлении определенной фазы. Один из наиболее хорошо изученных белков – геминин. Первые 110 аминокислотных остатков геминина содержат сигнал убиквитинирования в фазе G1. Присоединение этого фрагмента к Cas9 должно вызывать деградацию белка в G1. Действительно, использование химерного белка Cas9-Gem1 с фрагментом геминина позволило увеличить HDR в 1.28–1.72 раза в зависимости от количества трансфицированной плазмиды. Это, вероятно, связано с тем, что Cas9-Gem1 экспрессируется менее интенсивно, поэтому количество плазмиды, несущей его ген, может быть лимитирующим фактором [52].

Изменяя уровень Cas9, можно влиять на эффективность HDR. Показано, что как целевая, так и нецелевая активности Cas9 линейно зависят от количества плазмиды, трансфицированной в клетку [54]. Таким образом, увеличить специфичность можно за счет снижения общего уровня Cas9, жертвуя при этом целевой активностью.

Важно отметить, что, во-первых, режим репарации (соотношение NHEJ/HDR, длина делеций и вставок) зависит от локуса и типа клеток. Целевые сайты можно охарактеризовать средней длиной делеций и вставок, средней эффективностью HDR. Эти значения различаются, даже если сайты расположены на расстоянии всего нескольких нуклеотидов друг от друга. Во-вторых, эффективность работы Cas9 зависит от типа клеток, например, плюрипотентные стволовые клетки имеют наименьшую эффективность HDR [55, 56]. В-третьих, отсутствует корреляция между NHEJ и HDR [49, 51], что стоит учитывать при оценке активности Cas9.

ВЫБОР ДОНОРНОЙ ДНК

На эффективность HDR влияет выбор донорной ДНК, которой может быть плазида или олигонуклеотиды. Известно, что наибольшей эффективностью обладают одноцепочечные олигодезоксирибонуклеотиды (ssODN). При изучении механизма действия spCas9 установлено, что spCas9 диссоциирует симметрично с двух цепей ДНК, но при этом при нахождении нуклеазы на ДНК нецелевая цепь (содержащая NGG) остается достаточно подвижной, способной “дышать” [19]. На этом основании предположили, что использование ssODN, комплементарных этой цепи, позволит повысить эффективность HDR. Затем это предположение подтвердили экспериментально – при использовании “правильных” ssODN частота HDR возросла до 2.6 раз.

Другой важный параметр ssODN – его длина, которая не должна превышать 100–120 н., более длинные ssODN цитотоксичны [56]. Внутриклеточная устойчивость ssODN также играет некоторую роль. В некоторых случаях ssODN, защищенные фосфотиоатом (PS-ssODN), обеспечивают более высокую эффективность редактирования, чем незащищенные (PO-ssODN) [57]. Например, при редактировании зиготического генома мыши наблюдали огромную разницу в эффективности PS-ssODN и PO-ssODN – 62 и 9.5% соответственно. Но одновременно с этим не выявлено различий в эффективности редактирования другого гена в зиготе крысы с использованием двух типов ssODN. Это связали с длиной вставки – вставка в геном мыши имела длину 40 н., а в геном крысы – 1 н. Анализ проверенных ssODN позволяет сделать вывод, что укорачивание прилежащего к PAM плеча ssODN сильнее сказывается на эффективности ssODN, но, если плечи ssODN не слишком короткие (не менее 20 н.), то более важна общая длина ssODN. Это подтверждают результаты, согласно которым наиболее эффективны ssODN с плечами из 91 и 36 н. (прилежащее к PAM плечо) от сайта расщепления [19].

Использование dsODN незначительно уменьшает вероятность репарации, но в большинстве случаев сопровождается дополнительными мутациями [57].

РЕДАКТИРОВАНИЕ НЕКОДИРУЮЩИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Особую проблему представляет редактирование некодирующих последовательностей, которые можно разделить на транскрибируемые (miRNA – микроРНК и lncRNA – длинная некодирующая РНК) и нетранскрибируемые (регуляторные). В первом случае интересно осуществить нокаут генов РНК, во втором – отдельных частей регуляторных последовательностей, чтобы установить их функциональную значимость.

Нокаут микроРНК сложен из-за их малой длины и, следовательно, малого количества подходящих сайтов. Например, при попытке нокаута miR-21 на РНК направили одновременно три sgRNA, но получить клеточные линии с нокаутом miR-21 без отбора не удалось [58]. Во второй попытке нокаут проводили в триплоидных клетках и использовали HDR — сначала в ген вставили кассету с генами устойчивости к пурамицину и GFP. После отбора получили пять клонов с полным нокаутом miR-21. Оказалось, что кассета попала в одну хромосому, остальные две копии miR-21 были инактивированы NHEJ.

Нокаут lncRNA сложно осуществить по противоположным причинам — молекулярная масса lncRNA зачастую не уступает массе белков, но, как правило, делеции и вставки не приводят к инактивации гена, так как здесь нет рамки считывания. Проблему нокаута lncRNA решили путем делеции большого фрагмента гена с помощью пары sgRNA [58].

Легкость редактирования с помощью Cas9 позволяет на новом уровне проводить функциональный анализ некодирующих регуляторных последовательностей. Основная сложность заключается в большой протяженности таких последовательностей. Описан функциональный скрининг интронного энхансера гена *BCL11A* [59]. С этой целью к трем гиперчувствительным к ДНКазе сайтам длиной 1284, 1264 и 1370 п.н., содержащим достаточное количество NGG-участков, были подобраны 212, 174 и 147 sgRNA соответственно. Лишь несколько из этих sgRNA серьезно влияли на активность энхансера. Оценить точность и полноту такого анализа затруднительно.

ЗАКЛЮЧЕНИЕ

Можно выделить две большие области применения CRISPR/Cas9 — научно-исследовательскую и клиническую (генная терапия). CRISPR/Cas9 используется в разнообразных научных приложениях — от нокаута генов до инженерии антител и картирования хроматина. Однако основные надежды связаны с использованием Cas9 в качестве надежного инструмента генной терапии. На пути к этому стоят два препятствия — нецелевая активность и низкая эффективность HDR. Первая проблема преодолевается достаточно успешно. Новые мутантные Cas9 имеют значительно более низкую нецелевую активность. Вторая проблема более фундаментальна, ее решение может затянуться на длительный срок. Станет ли Cas9 золотым стандартом генной терапии остается под вопросом.

СПИСОК ЛИТЕРАТУРЫ

- Ishino Y., Shinagawa H., Makino K., Amemura M., Nakata A. (1987) Nucleotide sequence of the *iop* gene,

responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433.

- 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.
- Lander E.S. (2016) The Heroes of CRISPR. *Cell.* **164**, 18–28.
- Mojica F.J., Díez-Villaseñor C., García-Martínez J., Soria E. (2005) Intervening sequences of regularly interspaced prokaryotic repeats derive from foreign generic elements. *J. Mol. Evol.* **60**, 174–182.
- Васильева Е.А., Мелино Д., Барлев Н.А. (2015) Применение системы направленного геномного редактирования CRISPR/CAS к плюрипотентным стволовым клеткам. *Цитология.* **57**, 19–30.
- Tsui T.K., Li H. (2015) Structure principles of CRISPR-Cas Surveillance and effector complexes. *Annu. Rev. Biophys.* **44**, 229–255.
- Nishimasu H., Ran F.A., Hsu P.D., Konermann S., Shehata S.I., Dohmae N., Ishitani R., Zhang F., Nureki O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell.* **156**, 935–949.
- Koonin E.V., Krupovic M. (2015) Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat. Rev. Genet.* **16**, 184–192.
- Fagerlund R.D., Staals R.H., Fineran P.C. (2015) The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biol.* **16**, 251.
- Shmakov S., Abudayyeh O.O., Makarova K.S., Wolf Y.I., Gootenberg J.S., Semenova E., Minakhin L., Joung J., Konermann S., Severinov K., Zhang F., Koonin E.V. (2015) Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell.* **60**, 385–397.
- Пугач К.С., Лопатина А.В., Северинов К.В. (2012) CRISPR-системы адаптивного иммунитета прокариот *Молекуляр. биология.* **46**, 195–203.
- Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**, 816–821.
- Kouranova E., Forbes K., Zhao G., Warren J., Bartels A., Wu Y., Cui X. (2016) CRISPRs for optimal targeting: Delivery of CRISPR components as DNA, RNA and protein into cultured cells and single-cell embryos. *Hum. Gene Ther.* **27**, 464–475.
- Oakes B.L., Nadler D.C., Flamholz A., Fellmann C., Staahl B.T., Doudna J.A., Savage D.F. (2016) Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* **34**, 646–651.
- Sternberg S.H., Redding S., Jinek M., Greene E.C., Doudna J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature.* **507**, 62–67.
- Barkal A.A., Srinivasan S., Hashimoto T., Gifford D.K., Sherwood R. (2016) Cas9 functionally opens chromatin. *PLoS One.* **11**.
- Hinz J.M., Laughery M.F., Wyrick J.J. (2015) Nucleosomes inhibit Cas9 endonuclease activity *in vitro*. *Biochemistry.* **54**, 7063–7066.
- Knight S.C., Xie L., Deng W., Guglielmi B., Witkowsky L.B., Bosanac L., Zhang E.T., El Beheiry M., Mas-

- son J.B., Dahan M., Liu Z., Doudna J.A., Tjian R. (2015) Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*. **350**, 823–826.
19. Richardson C.D., Ray G.J., DeWitt M.A., Curie G.L., Corn J.E. (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344.
 20. Kim S., Kim D., Cho S.W., Kim J., Kim J.S. (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019.
 21. Tsai S.Q., Zheng Z., Nguyen N.T., Liebers M., Topkar V.V., Thapar V., Wyvekens N., Khayter C., Iafrate A.J., Le L.P., Aryee M.J., Joung J.K. (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197.
 22. Kleinstiver B.P., Pattanayak V., Prew M.S., Tsai S.Q., Nguyen N.T., Zheng Z., Joung J.K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. **529**, 490–495.
 23. Hu J., Meyers R.M., Dong J., Panchakshari R.A., Alt F.W., Frock R.L. (2016) Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* **11**, 853–871.
 24. Slaymaker I.M., Gao L., Zetsche B., Scott D.A., Yan W.X., Zhang F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science*. **351**, 84–88.
 25. Hu J., Meyers R.M., Dong J., Panchakshari R.A., Alt F.W., Frock R.L. (2016) Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* **11**, 853–871.
 26. Vouillot L., Th  lie A., Pollet N. (2015) Comparison of T7E1 and Surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *Genes, Genomes, Genetics*. **5**, 407–415.
 27. Jiang F., Zhou K., Ma L., Gressel S., Doudna J.A. (2015) A Cas9–guide RNA complex reorganized for target DNA recognition. *Science*. **348**, 1477–1481.
 28. Ran F.A., Hsu P.D., Lin C.Y., Gootenberg J.S., Konermann S., Trevino A.E., Scott D.A., Inoue A., Matoba S., Zhang Y., Zhang F. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. **154**, 1380–1389.
 29. Kleinstiver B.P., Prew M.S., Tsai S.Q., Topkar V.V., Nguyen N.T., Zheng Z., Gonzales A.P., Li Z., Peterson R.T., Yeh J.R., Aryee M.J., Joung J.K. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. **523**, 481–485.
 30. Wright A.V., Sternberg S.H., Taylor D.W., Staahl B.T., Bardales J.A., Kornfeld J.E., Doudna J.A. (2015) Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci. USA*. **112**, 2984–2989.
 31. Zetsche B., Volz S.E., Zhang F. (2015) A split Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142.
 32. Elleuche S., P  ggeler S. (2010) Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl. Microbiol. Biotechnol.* **87**, 479–489.
 33. Truong D.J., K  hner K., K  hn R., Werfel S., Engelhardt S., Wurst W., Ortiz O. (2015) Development of an intein-mediated split-Cas9 system for gene therapy. *Nucl. Acids Res.* **43**, 6450–6458.
 34. 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.
 35. Friedland A.E., Baral R., Singhal P., Loveluck K., Shen S., Sanchez M., Marco E., Gotta G.M., Maeder M.L., Kennedy E.M., Kornepati A.V., Sousa A., Collins M.A., Jayaram H., Cullen B.R., Bumcrot D. (2015) Characterization of *Staphylococcus aureus* Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol.* **16**, 257.
 36. Moreno-Mateos M.A., Vejnar C.E., Beaudoin J.D., Fernandez J.P., Mis E.K., Khokha M.K., Giraldez A.J. (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo*. *Nat. Methods*. **12**, 982–988.
 37. Wang T., Wei J.J., Sabatini D.M., Lander E.S. (2014) Genetic screens in human cells using the CRISPR/Cas9 system. *Science*. **343**, 80–84.
 38. Gagnon J.A., Valen E., Thyme S.B., Huang P., Akhmetova L., Pauli A., Montague T.G., Zimmerman S., Richter C., Schier A.F. (2014) Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One*. **9**, e98186.
 39. Chari R., Mali P., Moosburner M., Church G.M. (2015) Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods*. **12**, 823–826.
 40. Koike-Yusa H., Li Y., Tan E.P., Velasco-Herrera Mdel C., Yusa K. (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273.
 41. Wang X., Raghavan A., Chen T., Qiao L., Zhang Y., Ding Q., Musunuru K. (2016) CRISPR-Cas9 targeting of PCSK9 in human hepatocytes *in vivo*. *Arterioscler. Thromb. Vasc. Biol.* **36**, 783–786.
 42. Periwal V. (2016) A comprehensive overview of computational resources to aid in precision genome editing with engineered nucleases. *Brief Bioinform.* doi 10.1093/bib/bbw052
 43. Doench J.G., Hartenian E., Graham D.B., Tothova Z., Hegde M., Smith I., Sullender M., Ebert B.L., Xavier R.J., Root D.E. (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267.
 44. Doench J.G., Fusi N., Sullender M., Hegde M., Vaimberg E.W., Donovan K.F., Smith I., Tothova Z., Wilen C., Orchard R., Virgin H.W., Listgarten J., Root D.E. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191.
 45. Xu H., Xiao T., Chen C.H., Li W., Meyer C.A., Wu Q., Wu D., Cong L., Zhang F., Liu J.S., Brown M., Liu X.S. (2015) Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157.
 46. Montague T.G., Cruz J.M., Gagnon J.A., Church G.M., Valen E. (2014) CHOPCHOP: a CRISPR/Cas9 and

- TALEN web tool for genome editing. *Nucl. Acid Res.* **42**, W401–407.
47. Wong N., Liu W., Wang X. (2015) WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.* **16**, 218.
48. Ran F.A., Hsu P.D., Wright J., Agarwala V., Scott D.A., Zhang F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308.
49. Miyaoka Y., Berman J.R., Cooper S.B., Mayerl S.J., Chan A.H., Zhang B., Karlin-Neumann G.A., Conklin B.R. (2016) Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep.* **6**, 235–249.
50. Chu V.T., Weber T., Wefers B., Wurst W., Sander S., Rajewsky K., Kühn R. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548.
51. He X., Tan C., Wang F., Wang Y., Zhou R., Cui D., You W., Zhao H., Ren J., Feng B. (2016) Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucl. Acids Res.* **44**, e85
52. Gutschner T., Haemmerle M., Genovese G., Draetta G.F., Chin L. (2016) Post-translational regulation of Cas9 during G1 enhances homology-directed repair. *Cell.* **14**, 1555–1566.
53. Ma Y., Chen W., Zhang X., Yu L., Dong W., Pan S., Gao S., Huang X., Zhang L. (2016) Increasing the efficiency of CRISPR/Cas9-mediated precise genome editing in rats by inhibiting NHEJ and using Cas9 protein. *RNA Biol.* **13**, 605–612.
54. Davis K.M., Pattanayak V., Thompson D.B., Zuris J.A., Liu D.R. (2015) Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318.
55. 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.
56. Yang L., Guell M., Byrne S., Yang J.L., De Los Angeles A., Mali P., Aach J., Kim-Kiselak C., Briggs A.W., Rios X., Huang P.Y., Daley G., Church G. (2013) Optimization of scarless human stem cell genome editing. *Nucl. Acids Res.* **41**, 9049–9061.
57. Renaud J.B., Boix C., Charpentier M., De Cian A., Cochennec J., Duvernois-Berthet E., Perrouault L., Tesson L., Edouard J., Thinarth R., Cherifi Y., Menoret S., Fontanière S., de Crozé N., Fraichard A., Sohm F., Anegon I., Concordet J.P., Giovannangeli C. (2016) Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 nucleases. *Cell Rep.* **14**, 2263–2272.
58. Ho T.T., Zhou N., Huang J., Koirala P., Xu M., Fung R., Wu F., Mo Y.Y. (2015) Targeting non-coding RNAs with the CRISPR/Cas9 system in human cell lines. *Nucl. Acids Res.* **43**, e17.
59. Canver M.C., Smith E.C., Sher F., Pinello L., Sanjana N.E., Shalem O., Chen D.D., Schupp P.G., Vinjamur D.S., Garcia S.P., Luc S., Kurita R., Nakamura Y., Fujiwara Y., Maeda T., Yuan G.C., Zhang F., Orkin S.H., Bauer D.E. (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature.* **527**, 192–197.

CRISPR/Cas9, the King of the Tools for Genome Editing

A. V. Bannikov^{1,*}, A. V. Lavrov^{1,2}

¹Research Centre for Medical Genetics, Moscow, 115478 Russia

²Pirogov Russian National Research Medical University, Moscow, 117997 Russia

*e-mail: broyler3@mail.ru

Discovery of CRISPR/Cas9 brought a hope for acquiring an efficient, reliable and accessible method of genome editing which has been so much looked forward. No doubt that CRISPR/Cas9 is easy to use, while its efficiency and reliability remains a subject of studies. In this review we describe the general principles of organization and function of Cas nucleases and a number of important aspects to be considered before planning genome editing experiments using CRISPR/Cas9: estimation of efficacy and specificity of Cas9, sgRNA selection, artificially designed variants of Cas9 and usage of homologous recombination and non-homologous end joining in DNA editing.

Keywords: CRISPR/Cas9, genome editing, sgRNA