

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА
БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи

Васильев Руслан Алексеевич

**Направленная модификация геномов с помощью новых
эндонуклеаз CRISPR/Cas V типа**

1.5.3 «Молекулярная биология»

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата биологических наук

Научный руководитель:
доктор биологических наук
Каменский Петр Андреевич

Москва – 2022

ОГЛАВЛЕНИЕ

СПИСОК СОКРАЩЕНИЙ	5
ВВЕДЕНИЕ	7
Глава 1. ОБЗОР ЛИТЕРАТУРЫ	14
1.1 Системы борьбы с фагами	14
1.2.1 Характеристика систем CRISPR/Cas	17
1.2.2 Адаптация	20
1.2.3 Экспрессия и процессинг	25
1.2.4 Интерференция	27
1.2.5 CRISPR/Cas нуклеазы V типа	29
1.2.6 CRISPR диагностика	33
Глава 2. МАТЕРИАЛЫ И МЕТОДЫ	35
2.1.1 Поиск нуклеазы	35
2.1.2 Анализ CRISPR локусов	35
2.1.3 Молекулярное клонирование RbCas12a для получения рекомбинантного препарата	36
2.1.4 Молекулярное клонирование crPHK	38
2.1.5 Получение рекомбинантного RbCas12a	39
2.1.6 Получение библиотеки РАМ	40

2.2 Характеристика и определение оптимальных условий работы фермента RbCas12a	43
2.2.1 Получение crРНК для реакций разрезания <i>in vitro</i>	43
2.2.2 Подбор буфера для работы RbCas12a <i>in vitro</i>	44
2.2.3 Определение температурного оптимума RbCas12a	44
2.2.4 Определение кофактора RbCas12a	46
2.2.5 Определение последовательности PAM	46
2.2.6 Изучение коллатеральной активности RbCas12a	48
2.2.7 Определение природы выступающих концов двуцепочечной ДНК после направленного гидролиза матрицы с помощью RbCas12a	49
2.2.8 Разделение направляющей РНК на две молекулы	50
2.3.1 Трансфекция культуры клеток человека	52
2.3.2 Детекция мРНК RbCas12a в культуре клеток НЕК293Т	52
2.3.3 Формирование делеций в геноме культуры клеток НЕК293Т с помощью RbCas12a	53
Глава 3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ	61
3.1 Получение плазмидных конструкций	61
3.2 Получение рекомбинантного препарата RbCas12a	61
3.3 Активность RbCas12a в различных буферах	65

3.4	Определение температурного оптимума работы RbCas12a	67
3.5	Определение кофактора RbCas12a	70
3.6	Определение последовательности PAM	72
3.7	RbCas12a проявляет коллатеральную активность	75
3.8	RbCas12a формирует выступающие 5'-концы	77
3.9	Сплит направляющей РНК	78
3.10	Формирование делеций в геноме клеток линии НЕК293Т	87
	ЗАКЛЮЧЕНИЕ	91
	ВЫВОДЫ	94
	СПИСОК ЛИТЕРАТУРЫ	95
	ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ	108

СПИСОК СОКРАЩЕНИЙ

АТФ – аденозинтрифосфат

ДНК – дезоксирибонуклеиновая кислота

дцДНК — двуцепочечная ДНК

ОТ-ПЦР – разновидность ПЦР, с наличием этапа в виде обратной транскрипции (ОТ) РНК в ДНК

оцДНК — одноцепочечная ДНК

п.о. - пары оснований

ПЦР – полимеразная цепная реакция

ПЦР-РВ – полимеразная цепная реакция в реальном времени

РНК – рибонуклеиновая кислота

Abi (от англ. Abortive infection) — abortивная инфекция

BREX (аббр. от англ. bacteriophage exclusion) — бактериофаговая эксклюзия

Cas (от англ. CRISPR-associated) — CRISPR-ассоциированный

Cascade (аббр. от англ. CRISPR-associated complex for antiviral defense) — CRISPR-ассоциированный комплекс для антивирусной защиты

CrРНК – направляющая (CRISPR) РНК

CRISPR (аббр. от англ. Clustered regularly interspaced short palindromic repeats) — регулярно расположенные группами короткие палиндромные повторы

DI (аббр. от англ. defense islands) — защитные островки

DISARM (аббр. от англ. defence islands system associated with R-M) — система защитных островков, ассоциированная с рестрикцией-модификацией

DND (аббр. от англ. DNA degradation) — система защиты с фосфоротиоатной модификацией ДНК хозяина

DSB (аббр. англ. double-stranded break) — двуцепочечный разрыв

HDR (аббр. от англ. homology-directed repair) — гомологичная репарация

NGS (аббр. от англ. next generation sequencing) – технология секвенирования следующего поколения

NHEJ (аббр. от англ. non-homologous end joining) — негомологичное соединение концов

ORF (аббр. от англ. open reading frame) — открытая рамка считывания

PAM (аббр. от англ. Protospacer adjacent motif) — мотив, примыкающий к протоспейсеру

PBS (аббр. от англ. Phosphate-buffered saline) – натрий-фосфатный буфер

SNP (аббр. от англ. single nucleotide polymorphism) – однонуклеотидный полиморфизм

R-M (аббр. от англ. Restriction-modification) — рестрикция-модификация

ВВЕДЕНИЕ

Актуальность темы исследования и степень ее разработанности

За время, прошедшее с момента открытия коротких палиндромных повторов, регулярно расположенных группами (CRISPR), в 1987 году группой японских исследователей [1] до первых попыток применения Cas-нуклеаз в качестве редакторов генома в 2013 году, системы CRISPR/Cas представляли ограниченный интерес для мирового научного сообщества. Была выявлена биологическая роль CRISPR-повторов в геномах многих прокариот [2], осуществлены попытки классифицировать эти защитные системы [3], а также исследованы структурные и функциональные особенности систем [4], [5]. Однако попыток к использованию систем CRISPR/Cas в качестве редакторов генома клеток млекопитающих не предпринималось вплоть до 2013 года, когда появилось большое количество работ, исследовавших различные аспекты использования новой технологии для редактирования ядерных геномов различных организмов [6]–[9]. Подобный интерес был вызван потенциальными возможностями применения исследуемых ферментов, а также сравнительной легкостью проводимых экспериментов, ставших доступными большинству современных молекулярно-биологических лабораторий. Однако с расширением фундаментальных знаний о структуре и функциях различных систем CRISPR/Cas стало очевидно, что возможности каждой отдельно взятой системы ограничены теми или иными факторами. На сегодняшний день не известно систем CRISPR/Cas, универсально пригодных для решения любых задач практической геномной инженерии. В этой связи поиск новых инструментов редактирования генома и изучение их свойств является актуальной научной задачей, решения которой постепенно снимают барьеры, стоящие на пути исследований в данной области. Одним из подобных барьеров является РАМ-специфичность Cas-нуклеаз, не позволяющая направлять эффекторные комплексы к любому желаемому локусу генома. В сфере геномного

редактирования растений известен другой фактор, затрудняющий работы по получению мутантов: многие растения, в том числе сельскохозяйственного пользования, плохо переносят температуры выше комнатной в лабораторных условиях, в то время как комплексы CRISPR/Cas часто имеют температурный оптимум работы в районе 37 °С. Возможным путем преодоления этих и других барьеров является изучение нуклеаз, обладающих свойствами, выгодно отличающих их от ортологов, таких как менее строгие требования к PAM или активность в широком диапазоне температур. Исходя из этого, поиск новых направляемых нуклеаз и определение их биохимических свойств является важной задачей современной молекулярной биологии.

Цель работы

Целью данной работы является получение и характеристика новой эндонуклеазы семейства Cas12a, способной вносить направленные модификации в геном клеток человека.

Задачи работы

1. Биоинформатический поиск нуклеотидной и аминокислотной последовательностей новой нуклеазы Cas подтипа V-A.
2. Получение рекомбинантного белка и определение его активности в составе комплексов с различными направляющими РНК в условиях *in vitro*.
3. Исследование требований к PAM-последовательности новой нуклеазы Cas подтипа V-A.
4. Определение свойств новой нуклеазы Cas подтипа V-A в условиях *in vitro*: кофактор-зависимость, температурный оптимум работы, определение структуры выступающих концов, наличие коллатеральной активности.

5. Исследование возможности осуществления разрезания *in vitro* комплексами различных Cas-нуклеаз подтипа V-A, заряженными двумя молекулами РНК: отдельно спейсерной и скаффолдной частями направляющей РНК.

6. Исследование возможности получения направленной делеции в гене *DNMT1* в геноме культуры клеток НЕК293Т при помощи новой нуклеазы Cas подтипа V-A.

Научная новизна

В работе впервые была показана и изучена активность эффекторной нуклеазы системы CRISPR/Cas подтипа V-A из бактерии *Ruminococcus bromii* sp. (RbCas12a). Продемонстрирована способность нуклеазы RbCas12a в комплексах с направляющей РНК (crРНК) осуществлять гидролиз дцДНК *in vitro*. Показана активность RbCas12a *in vitro* в отношении ДНК, содержащих PAM-последовательности с консенсусом 5'-YYN-3'. Выявлена независимость эффективности разрезания от последнего нуклеотида PAM в отношении одного и того же протоспейсера; в частности, не наблюдалось дискриминации в отношении тимидинового нуклеотида в последнем положении PAM, свойственной некоторым ортологам Cas12a [10]. Проведен анализ свойств нуклеазы RbCas12a, подтверждающий наибольшую активность фермента при использовании в качестве кофакторов ионы магния или марганца. Проведен сравнительный анализ зависимости эффективности разрезания матрицы от времени инкубации с комплексами RbCas12a/crРНК и AsCas12a/crРНК, продемонстрировавший высокую эффективность разрезания RbCas12a спустя несколько секунд инкубации. Проведен сравнительный анализ активности комплекса RbCas12a/crРНК при различных температурах, по результатам которого обнаружено, что комплекс способен эффективно функционировать при диапазоне температур от 20 до 42 °С. Впервые показано, что различные

ортологи Cas12a, включая RbCas12a, способны формировать эффекторные комплексы с двумя отдельными молекулами РНК, функционирующими соответственно как спейсерная и скаффолдная части crРНК, практически не теряя при этом эффективности в условиях *in vitro*. Продемонстрирована возможность разрезания *in vitro* матрицы комплексами Cas12a только со спейсерной частью crРНК при отсутствии скаффолда. Показано, что RbCas12a проявляет коллатеральную активность в отношении оцДНК. Определена природа выступающих концов при гидролизе ДНК матрицы комплексом RbCas12a/crРНК: формирование 5'-выступающих концов на 1 нуклеотид длиннее, чем при гидролизе ортологом Cas12a из *Francisella novicida* (FnCas12a). Показана возможность внесения направленных делеций в геном культуры клеток человека HEK293T с использованием кодон-оптимизированного RbCas12a.

Научная и практическая значимость

На текущий момент интерес к использованию новых нуклеаз существует в таких отраслях, как биотехнология, медицина, сельское хозяйство, но для их использования в этих сферах необходимо четко и всесторонне охарактеризовать ферменты, выявить оптимальные условия их работы и возможные положительные и отрицательные параметры в приложении к конкретным задачам. Поиск и характеристика новых представителей систем CRISPR/Cas расширяет возможности их применения в различных областях. Обнаружение нуклеаз с различной специфичностью по отношению к РАМ-последовательности позволит редактировать практически любой желаемый локус в геноме. Новые ферменты с уникальными свойствами могут найти свое применение в решении большого количества текущих задач промышленности, медицины и сельского хозяйства. Как показывает международная практика, системы геномного редактирования уже сейчас используются в отдельно взятых государствах. Предложены и внедряются подходы к терапии генетических

заболеваний, таких как миодистрофия Дюшенна [11], гемофилия [12], талассемия [13], генетически-обусловленная глухота [14]. Рассматриваются варианты редактирования генома с целью снижения рисков инфекционных, в частности, вирусных заболеваний, на примере ВИЧ-1 [15] и других социально-значимых патологий [16], [17]. Также применение культур с отредактированным геномом в сельском хозяйстве за последние десятилетия выросло во всем мире и продолжает расти, что в условиях повышения плотности населения и возможного глобального потепления будет играть ведущую роль в поддержании темпов урожайности сельскохозяйственных растений [18]. Получение отечественных ферментов для редактирования генома, в частности CRISPR/Cas-нуклеаз, является стратегически-важным направлением развития биотехнологической отрасли.

Личный вклад автора

Личный вклад автора заключается в работе с литературными источниками, планировании и проведении экспериментов, анализе полученных результатов, подготовке материалов к печати и написании диссертационной работы. Основные результаты диссертационной работы получены лично автором или при его непосредственном участии. Имена всех соавторов указаны в опубликованных работах или отражены в тексте работы. Планирование экспериментов было выполнено совместно с д.б.н. П.А. Каменским, к.б.н. И.О. Мазуниным, к.б.н. М.В. Патрушевым. Выделение белков было выполнено совместно с к.х.н. А.В. Власкиной. Эксперименты по разрезанию *in vitro* комплексами Cas нуклеаз с двумя молекулами РНК были проведены совместно с Р.М. Шебановой, Н.А. Никитчиной, к.б.н. И.О. Мазуниным, д.б.н. К.В. Севериновым. Анализ данных высокопроизводительного секвенирования проводился совместно с к.б.н. С.В. Тошаковым, А.А. Корженковым.

Методология и методы исследования

Рекомбинантные белки выделены с использованием хроматографических методов с подтвержденной чистотой, пригодной для выполнения ферментативных реакций. Исследования проводились на клеточной культуре НЕК293Т со статусом отсутствия инфицирования микоплазмой, подтвержденным тестами с использованием коммерческого набора MycoReport (Евроген). Анализ данных высокопроизводительного секвенирования был выполнен с использованием различных биоинформатических программ. Визуализация полученных результатов высокопроизводительного секвенирования была выполнена с использованием различных пакетов языка Python3.

Положения, выносимые на защиту

1. Идентифицирована новая нуклеаза Cas12a из *Ruminococcus bromii* sp. (RbCas12a).
2. Рекомбинантный препарат RbCas12a обладает активностью в отношении двуцепочечной ДНК в условиях *in vitro*.
3. Комплексы RbCas12a/crРНК способны направленно гидролизовать ДНК, содержащие последовательность 5'-YYN-3' в качестве PAM.
4. Нуклеаза RbCas12a способна использовать в качестве кофакторов ионы магния и марганца, а также, с меньшей эффективностью, ионы кобальта, цинка и никеля. Температурный оптимум работы RbCas12a находится в диапазоне от 20 до 42 °С. В качестве продуктов гидролиза образуется дцДНК с 5'-выступающими концами. RbCas12a проявляет выраженную коллатеральную активность.
5. Cas-нуклеазы V-A подтипа, в том числе RbCas12a, способны формировать комплексы с двумя молекулами РНК, представляющими собой соответственно спейсерную и скаффолдную части crРНК, и осуществлять в их

составе эффективный направленный гидролиз двуцепочечной ДНК в условиях *in vitro*.

6. Нуклеаза RbCas12a активна в клетках человека, что показано путем внесения направленной делеции в ген *DNMT1* культуры клеток HEK293T.

Степень достоверности и апробация результатов

Результаты были получены с использованием современных методов молекулярной биологии и биоинформатики. По теме диссертационной работы опубликовано 3 статьи в рецензируемых научных изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности 1.5.3 – Молекулярная биология.

Результаты диссертационной работы были представлены на Всероссийской конференции «Синтетическая биология и биофармацевтика» (Новосибирск, 24-28 июля 2022 г.).

Структура и объем диссертации

Диссертационная работа изложена на 116 страницах и состоит из разделов: Введение, Обзор литературы, Материалы и методы, Результаты и обсуждение, Заключение, Выводы, Список использованной литературы, Дополнительные материалы. Рукопись включает 2 таблицы и 23 рисунка. Список литературы содержит 116 источников. Дополнительные материалы включают 3 приложения.

Глава 1. ОБЗОР ЛИТЕРАТУРЫ

1.1 Системы борьбы с фагами

Гонка вооружений между прокариотическими микроорганизмами и их вирусами (бактериофагами) привела к появлению множества ферментативных активностей в процессе эволюции, которые опосредуют противофаговый «иммунный ответ» у бактерий и архей [19]. Результаты исследований методами сравнительной геномики показали, что «антифаговые» гены расположены в основном группами в локусах генома, называемых защитными островками (DI) [20]. DI одного микроорганизма часто содержат гены нескольких разных защитных систем; известны примеры наличия до 100 защитных генов в одном островке [21]. Для этих локусов характерна высокая скорость рекомбинации, что дает находящимся в DI системам защиты возможность быстро эволюционировать путем обмена доменами и целыми генами [22]. Число охарактеризованных защитных систем против фагов постоянно растет, однако на данный момент принято условно выделять системы «врожденного иммунитета» и системы «адаптивного иммунитета» прокариот [23].

В системах «врожденного иммунитета» представлены различные механизмы противофаговой активности: от специфических модификаций ДНК хозяина либо фага (последнее относится к рестрикции-модификации (R-M) IV типа) с дальнейшей деградацией фаговой ДНК до abortивной инфекции, химической защиты и сигнальных систем. R-M и токсин-антитоксиновые системы (ТА) известны несколько десятилетий, их конкретные примеры широко известны и хорошо охарактеризованы [24]. Системы R-M функционируют, основываясь на активностях метилтрансферазных доменов и эндонуклеаз рестрикции: в то время как первые вносят модификации в ДНК, вторые специфически гидролизуют фаговую ДНК, распознавая по метильным группам «свою» и «чужую» генетическую информацию. Системы ТА представлены

токсичным белком и его ингибитором белковой или рибонуклеиновой природы, находящихся в норме в сбалансированных концентрациях в клетке. В случае стресса и нарушения равновесия системы, токсичный белок приводит к клеточной гибели. Системы R-M и TA являются наиболее распространенными защитными системами среди как эубактерий, так и архей с расшифрованным на данный момент геномом [25]. Ряд защитных систем, таких как эксклюзия бактериофагов BREX и ассоциированные с R-M защитные островки DISARM, также распознают метильные группы в ДНК, хотя подробные механизмы работы этих систем остаются не изученными. Метилирование ДНК является не единственной химической модификацией, используемой для дискриминации «своего» и «чужого» в клетках прокариот: системы DND вносят фосфоротиоатные модификации в ДНК хозяина [26], а системы DPD модифицируют остатки гуанозина хозяйской ДНК, превращая его в 7-дезагуанозин [27]. Другой распространенной стратегией противофагового «иммунитета» является abortивная инфекция Abi. Данный термин, отражающий непродуктивную фаговую репликацию, появился в научной литературе уже в 1950 годах, однако начать раскрывать принцип работы системы удалось лишь в 1980-х [28]. Работа систем Abi нацелена на гибель зараженной фагом клетки либо остановку клеточного метаболизма во избежание завершения цикла размножения фага. Системы синтеза доксорубицина и даунорубицина у *Streptomyces* spp. являются примером химической защиты хозяина. Данные вещества являются интеркаляторами ДНК и способны ингибировать инфекцию, не оказывая существенного влияния на репликацию генома хозяина [29]. Защита на основе сигнальных систем с циклическими олигонуклеотидами CBASS устроена по принципу сенсорного обнаружения неизвестного компонента фага, после которого запускается работа cGAS/DncV-подобной нуклеотидилтрансферазы, формирующей вторичные мессенжеры в виде малых циклических молекул, которые передают сигнал эффекторам (нуклеазам, фосфолипазам или компонентам мембранных ионных

каналов в зависимости от типа системы). Это приводит к клеточной гибели хозяина, предотвращая, соответственно, размножение бактериофага [28]. Множество других систем защиты против фагов было в последнее время предсказано биоинформатически. Многие из них удалось в дальнейшем проверить на функциональность, однако без детального описания ферментативных активностей входящих в их состав белков, некоторые из которых имеют неаннотированные домены, либо домены, неассоциированные на данный момент с защитными системами (такие как HAD — галоацид дегидрогеназа-подобная гидролаза), либо домены с неизвестной функцией [30]. Примерами доказанной активности защитных систем без четкого понимания механизмов сопротивления фагам являются также некоторые системы на основе ретронов [19]. Ретроны представляют собой генетические элементы, состоящие из ревертазы, некодирующей РНК и ассоциированной с ревертазой открытой рамкой считывания (в некоторых случаях дополнительного домена ревертазы). На данный момент обнаружено 10 различных групп (клад) ретронов, внутри которых найдено не менее 19 типов ассоциированных с ревертазой ORF, каждый из которых обладает своей ферментативной активностью [31]. Еще одним интересным примером защитных систем являются белки-аргонавты, найденные во всех трех доменах жизни [32]. В эукариотических организмах они осуществляют РНК-зависимую РНК-интерференцию [33], а в прокариотах по крайней мере некоторые из аргонавтов участвуют в защите против мобильных генетических элементов. Принцип действия прокариотических аргонавтов схож с таковым для систем CRISPR/Cas, детально охарактеризованным ниже: аргонавт связывается с направляющей молекулой оцДНК и гидролизует комплементарную молекулу другой оцДНК, осуществляя тем самым ДНК-зависимую ДНК-интерференцию [34], хотя известны примеры аргонавт-опосредованной РНК-зависимой ДНК/РНК-интерференции у бактерий [35].

В то время как открытые ранее защитные системы называются аббревиатурами и сокращениями, многие из описанных в последние годы противофаговых систем получили свои названия в честь мифологических богинь и богов из эпосов разных культур и художественных произведений [36]. К таковым, например, относятся, широко распространенные Zorya, Nachiman, Druantia, Gabija и менее распространенные Lamassu, Thoeris, Septu, Shedu, Wadjet, Kiwa.

Второй группой защитных систем являются системы «адаптивного иммунитета». На данный момент к ней относят только CRISPR/Cas (от англ. clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins — регулярно расположенные группами короткие палидромные повторы/CRISPR-ассоциированные белки), поскольку именно эти системы обладают способностью «запоминать» предыдущие фаговые инфекции, после которых микроорганизмам удавалось выжить. В случае повторного заражения тем же или родственным фагом, системы CRISPR/Cas быстро реагируют на это, гидролизуя «чужую» нуклеиновую кислоту. Ранние работы были в основном сфокусированы на изучении систем R-M и систем abortивной инфекции, в то время как в последнее десятилетие фокус мировых исследований сместился на всестороннюю характеристику систем CRISPR/Cas и создание молекулярных инструментов на их основе [36].

1.2.1 Характеристика систем CRISPR/Cas

В бактериях и археях системы CRISPR/Cas осуществляют адаптивную защиту, используя некодирующие РНК в качестве направляющих молекул к чужеродной нуклеиновой кислоте. Связываясь с одной или несколькими РНК, эффекторные белки Cas или их комплексы осуществляют гидролиз молекул, полностью или частично комплементарных так называемой спейсерной части crisprРНК (crРНК). Гены crРНК находятся в локусах CRISPR в так называемых

CRISPR-массивах (или кассетах) и транскрибируются полицистронно, продуцируя относительно длинную молекулу пре-crPHK, которая затем процессируется, давая ряд коротких зрелых crPHK [5]. Локусы CRISPR также содержат гены белков Cas, количество и набор которых варьируют между различными типами систем CRISPR/Cas. Ключевой особенностью CRISPR/Cas является «запоминание» коротких фрагментов фаговых нуклеиновых кислот, называемых спейсерами, благодаря встраиванию одного или нескольких таких спейсеров в CRISPR-массив хозяина [2]. Таким образом при следующей попытке фаговой инфекции тем же или близкородственным фагом, система CRISPR/Cas потомков микроорганизма способна быстро отреагировать благодаря экспрессии crPHK, содержащей новоинтегрированный спейсер. С этой crPHK связываются эффекторные белки Cas, комплекс с которыми осуществляет направленный гидролиз комплементарной спейсеру части чужеродной нуклеиновой кислоты. Комплементарная спейсеру часть целевой молекулы ДНК называется протоспейсером, и в большинстве случаев для его эффективного узнавания белкам Cas требуется относительно короткая последовательность, примыкающая к протоспейсеру – PAM (англ. protospacer adjacent motif). Благодаря последовательности PAM белки Cas не разрезают собственный геном хозяина, поскольку PAM отсутствует в CRISPR-массивах. Спейсеры в CRISPR-массивах разделены повторами, которые в зрелых транскриптах становятся скаффолдной частью crPHK, формирующей шпильку, с которой и связываются белки Cas. Следовательно, в механизме действия систем CRISPR/Cas можно выделить 3 стадии: 1) адаптация, во время которой происходит «запоминание» путем интеграции новых спейсеров в локус CRISPR прокариотического генома; 2) процессинг, во время которого CRISPR-массив транскрибируется в пре-crPHK и происходит ее специфичный гидролиз с образованием зрелых crPHK; 3) интерференция, во время которой происходит распознавание целевой нуклеиновой кислоты и ее разрушение [37].

Рибонуклеопротеиновый комплекс crРНК и белков Cas, осуществляющих интерференцию, называется эффекторным комплексом.

Системы CRISPR/Cas известны в широких кругах в первую очередь как инструменты редактирования генома. Однако стоит иметь в виду, что лишь малая часть изученных систем может быть использована в качестве редакторов генома. Эти системы принято классифицировать в соответствии с несколькими признаками одновременно, поскольку стандартные филогенетические методы исследования сталкиваются со сложностями ввиду слишком мощного разнообразия систем CRISPR/Cas, выражающемся в отсутствии белковой последовательности, которая была бы гомологична внутри всех найденных систем [38]. Тем не менее подход к классификации был предложен и, претерпевая небольшие изменения ежегодно, успешно используется научным сообществом (Таблица 1). Принято делить системы CRISPR/Cas на 2 класса, 6 типов, обозначаемых римскими цифрами, и множество подтипов. 2-й класс длительное время представлял наибольший интерес для исследователей, поскольку в нем представлены системы, в которых эффекторный комплекс может быть сформирован с участием лишь одного фермента - нуклеазы, направляемой на целевой участок нуклеиновой кислоты с помощью одной или нескольких коротких РНК. 1-й класс в свою очередь представляют системы, в которых эффекторный комплекс состоит из нескольких ферментов, зачастую в совокупности более массивных, чем однобелковые комплексы. Интерес к системам 2-го класса обусловлен легкостью генно-инженерных манипуляций с одним геном, а также сравнительно небольшим размером этого гена, что позволяет использовать различные агенты для доставки таких систем внутрь клеток. Однако в последнее время стали появляться работы, в которых успешно продемонстрировано, что некоторые представители CRISPR/Cas 1-го класса могут быть использованы в качестве инструментов для манипуляций с нуклеиновыми кислотами в клетках млекопитающих. В частности, работа по

изучению однобелкового эффекторного комплекса Cas7-11 оставляет за системами 1-го класса право рассматриваться в качестве генетических редакторов будущего [39]. Механизмы адаптации, процессинга и интерференции различаются между двумя классами и требуют более подробного рассмотрения.

Таблица 1. Подтипы Cas-нуклеаз V типа, их таргеты и требование к наличию дополнительной РНК

Подтип	Таргет	Наличие tracrРНК
V-A	ДНК	Нет
V-B	ДНК	Да
V-C	ДНК	scoutРНК
V-D	ДНК	scoutРНК
V-E	ДНК	Да
V-F	ДНК	Да
V-G	РНК	Да
V-H	ДНК	Нет
V-I	ДНК	Нет
V-J	ДНК	Нет
V-K	ДНК	Да

1.2.2 Адаптация

Наиболее консервативными генами во всех известных системах CRISPR/Cas являются *cas1* и *cas2*, наличие продуктов которых необходимо для интеграции новых спейсеров [3]. Структурные и биохимические исследования показали, что белки Cas1 и Cas2 формируют гетерогексамерный комплекс, который вводит новые спейсеры в сайт между АТ-богатым линкером и первым повтором CRISPR-массива (Рисунок 1). В случае если в CRISPR-массиве отсутствует предыдущий добавленный спейсер, клетка хозяина наивна по отношению к инфекции фага и ей требуется формирование преспейсера *de novo*. Преспейсером называют короткую последовательность ДНК, которая будет в дальнейшем интегрирована в качестве спейсера в CRISPR-массив. Подобный процесс называют наивной адаптацией. Стоит уточнить, что наивная адаптация может протекать как с фаговой нуклеиновой кислотой, так с другими мобильными генетическими элементами и фрагментами собственного генома хозяина[40].

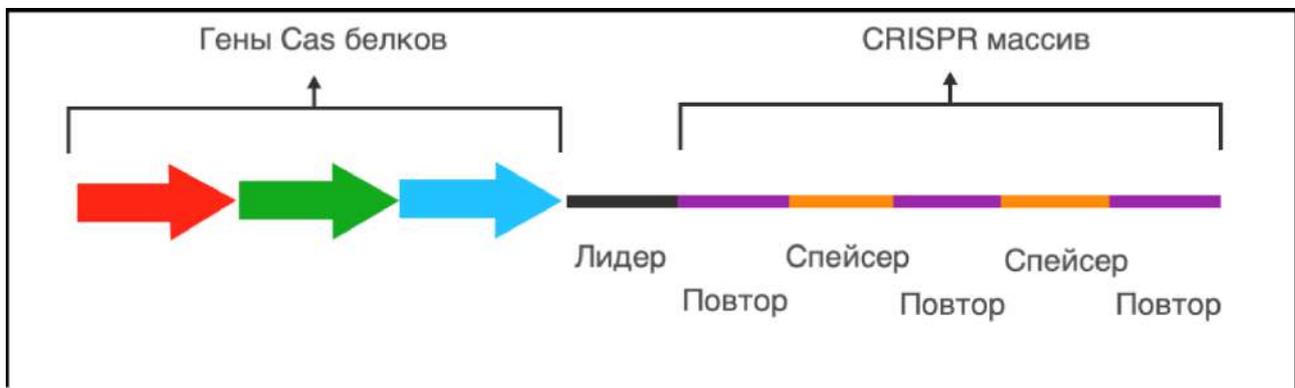


Рисунок 1. Общая схема организации локусов CRISPR. Локусы CRISPR состоят из генов, кодирующих белки Cas, CRISPR-массива, внутри которого находятся повторы, перемежающиеся спейсерами, и 5'-концевой лидерной последовательности, примыкающей к первому повтору. Ориентация *cas* генов, их количество, а также относительное расположение CRISPR-массива и расстояние до него от ближайшего *cas* гена варьирует между разными системами CRISPR/Cas.

Последний случай либо приводит к гибели клетки при наличии РАРМ и активных белков Cas, либо требует от хозяина активных действий по противодействию собственной системе защиты. К ним можно отнести репарацию ДНК, в том числе путем гомологичной рекомбинации (HDR — англ. homology-directed repair) с дополнительной копией бактериальной хромосомы или путем негомологичного соединения концов (NHEJ — англ. non-homologous end joining), в результате которого формируются инделы (вставки и/или делеции) [41]. После внесения двуцепочечного разрыва (DSB — англ. double-stranded break) свободные концы ДНК связываются экзонуклеазным комплексом RecBCD грамотрицательных штаммов или AddAB грамположительных штаммов [40], которые расплетают и деградируют концы вплоть до достижения Хи-сайта (Chi, аббр. от англ. crossover hotspot instigator — зачинщик кроссинговера), высвобождая в качестве продуктов деградации короткие одноцепочечные фрагменты ДНК (оцДНК), служащие в качестве преспейсеров [42]. Геномы прокариотических организмов содержат довольно много Хи-сайтов, которые предотвращают чрезмерную активность RecBCD в отношении собственного генома хозяина. Например, 8-нуклеотидный 5'-GCTGGTGG-3' Хи-сайт у *Escherichia coli* повторяется 1008 раз на геном размером 4.64 Мбр (аббр. от англ. megabase pair — миллион пар оснований), однако не стоит считать, что все фаговые геномы и плазмиды бедны данными последовательностями: фаги T3, T5, T7 имеют плотность Хи-сайтов, близкую или большую, чем у *E. coli*, а плотность консенсуса в геномах фагов P1, P7, D6 и NGUA22 в 2-3 раза выше, чем у *E. coli*. Тем не менее, приведенные фаги довольно быстро способны уничтожить клетки *E. coli*, что говорит о ложности предположения о дискриминации «своего» и «чужого» с помощью Хи-сайтов [43]. В то же время множество мобильных генетических элементов, не содержащих в своем геноме Хи-сайтов, деградируют по указанному пути.

Еще одним способом борьбы прокариот с собственными системами CRISPR/Cas является мутагенез. Различные мутации могут появиться в PAM, в генах *cas*, в спейсере или в таргетном регионе генома хозяина, предотвращая гидролиз собственной ДНК. Эффекторные комплексы некоторых систем CRISPR/Cas гидролизуют только РНК, таким образом наличие спейсера, аналогичного фрагменту собственного генома микроорганизма, не приводит к его гибели, и в то же время система активна против фагов и профагов, входящих в литический цикл, требующий транскрипции «чужих» генов, которые могли быть репрессированы в лизогенную фазу. Более активным инструментом ингибирования эффекторного комплекса пользуются бактерии и археи, экспрессирующие белки «анти-CRISPR» (Acrs — англ. Anti-CRISPRs) [41]. Большинство изученных на данный момент Acrs связываются с функционально важными сайтами эффекторных нуклеаз Cas, а некоторые из них мимикрируют структуру двуцепочечной ДНК (дцДНК), снижая или полностью подавляя возможность гидролиза таргетной последовательности [44].

Более эффективным способом встраивания новых спейсеров в CRISPR-массив является праймированная адаптация, или прайминг (англ. primed adaptation), которая запускается в случае наличия в кассете заранее адаптированного спейсера того же или близкородственного фага, и позволяет клетке быстро реагировать на возможные возникшие в протоспейсере или PAM мутации, резко снижающие эффективность интерференции [45]. Как было указано выше, активность комплексов RecBCD и AddAB приводит к возникновению набора возможных преспейсеров для наивной адаптации. Альтернативными источниками формирования преспейсеров при праймированной адаптации являются комплексы ферментов Cas. В системах I типа формируется комплекс Cascade (англ. CRISPR-associated complex for antiviral defense — CRISPR-ассоциированный комплекс для антивирусной защиты) из белков Cas5, Cas6, Cas7, Cas8 (а также иногда Cas11) и crРНК[46].

Стоит отметить, что непосредственно в комплексах Cascade разных подтипов I типа белок Cas6 может отсутствовать, действуя независимо от Cascade [47]. Комплекс Cascade при связывании с crРНК распознает целевую ДНК, после чего связывает Cas3 с хеликазной и нуклеазной активностями, который формирует пул одноцепочечных преспейсеров и необходим как в прайминге, так и в интерференции [48]. Помимо белка Cas3, комплекс Cascade связывает Cas1-Cas2, позволяя продуктам реакций Cas3 переходить непосредственно в Cas1-Cas2. В системах II типа роль Cascade-Cas3 выполняет нуклеаза Cas9, продуцирующая двуцепочечные фрагменты. Комплексы Cascade-Cas3 и Cas9 могут являться источниками преспейсеров как для наивной, так и для праймированной адаптации. Высвобожденные после RecBCD или Cascade-Cas3 фрагменты оцДНК, содержащие PAM, связываются комплексом Cas1-Cas2 (подтип I-E) или Cas4-Cas1-Cas2 (подтип I-G), который способствует отжигу комплементарной цепи и защищает PAM-содержащий конец от экзонуклеаз. Последние процессируют торчащие концы той цепи ДНК, которая не содержит PAM (нетаргетная цепь), после чего комплекс осуществляет интеграцию спейсера в два этапа. На первом шаге процессированный конец нетаргетной цепи встраивается в одну из цепей между лидерной последовательностью CRISPR-массива и первым повтором (так называемый L-сайт). На втором шаге конец таргетной цепи высвобождается из Cas1-Cas2, процессируется экзонуклеазой DnaQ и встраивается в S-сайт второй цепи между первым повтором и спейсером. Интеграция нового спейсера завершается диссоциацией белкового комплекса и репарационной активностью ферментов микроорганизма [40], [49]. Про механизмы адаптации других типов систем CRISPR/Cas известно гораздо меньше. Системы II типа требуют присутствия РНК trans-CRISPR (tracrРНК) и белка Csn2, который входит в комплекс с Cas1-Cas2, однако в процессе интеграции спейсера непосредственного участия не принимает; роль этого белка в адаптационном комплексе до конца не изучена [50]. Системы III типа нуждаются в ревертазной активности ферментов для синтеза преспейсеров

из РНК-протоспейсеров [51], [52]. У систем IV типа адаптационный модуль отсутствует, и скорее всего для интеграции спейсеров используется комплекс Cas1-Cas2 других систем CRISPR/Cas организма [53]. Такой же метод заимствования адаптационного модуля доказан и для некоторых систем VI типа [54]. Локусы CRISPR некоторых систем V типа содержат гены *cas1*, *cas2* и *cas4*, продукты которых формируют адаптационный модуль [55]. Стоит отметить, что разнообразие подтипов систем CRISPR/Cas V типа наиболее велико среди всех остальных типов, поэтому наборы необходимых компонентов для адаптации конкретных систем могут различаться или вообще отсутствовать [18]. Например, для функциональности нуклеаз большинства подтипов V типа необходимо наличие дополнительной РНК (tracrРНК или scoutРНК (англ. short-complementarity untranslated RNA — нетранслируемая РНК с короткой комплементарностью)) [56], а для адаптации некоторых систем с рибонуклеазным интерференционным модулем, вероятно, требуется ревертазная активность. Стоит отметить, что для удобства проведения лабораторных экспериментов молекулы tracrРНК и crРНК часто объединяют в одну молекулу sgРНК (от англ. single guide — единственная направляющая).

1.2.3 Экспрессия и процессинг

После интеграции нового спейсера CRISPR-массив часто транскрибируется с единственного промотора, в результате чего возникает длинная некодирующая пре-crРНК, представляющая собой набор спейсерных и скаффолдных участков. Ферментативная машинерия, осуществляющая процессинг пре-crРНК в зрелые crРНК различается для разных типов и некоторых подтипов систем CRISPR/Cas. Для большинства систем I и III типов была показана эндорибонуклеазная активность белков Cas6, вносящих разрыв на расстоянии 8 нуклеотидов в сторону 5'-конца от точки спейсер-скаффолдного сочленения. Однако подтип I-C использует для этого белки Cas5, которые осуществляют металл-независимый гидролиз на расстоянии 11 нуклеотидов в

сторону 5'-конца от точки спейсер-скаффолдного сочленения [47]. Лocus II типа систем CRISPR/Cas кодирует дополнительную РНК, называемую tracrРНК, с которой связывается эффекторная нуклеаза Cas9. Молекула tracrРНК имеет полностью или частично комплементарный CRISPR-повторам фрагмент, благодаря чему Cas9 связывается с дуплексом tracrРНК-пре-crРНК [57]. Процессинг в зрелые crРНК осуществляется в 2 этапа с помощью фермента РНКазы III. Роль Cas9 в этом процессе не совсем ясна, однако есть предположение, что связывая tracrРНК и пре-crРНК, Cas9 экранирует их и тем самым препятствует воздействию других РНКаз, а также связывание позволяет до окончания процессинга собрать эффекторный комплекс tracrРНК-Cas9-crРНК. Во время первого этапа процессинга РНКазы III продуцирует скаффолд-спейсер-скаффолд-содержащие crРНК, после чего неизвестный фермент осуществляет либо эндо-, либо экзонуклеазную активность в отношении 5'-концевого скаффолда, оставляя зрелые crРНК [47]. Стоит отметить, что в системах II-C подтипа на примере системы *Neisseria meningitidis* был показан механизм процессинг-независимого созревания crРНК благодаря наличию дополнительных промоторных областей для каждой пары спейсер-скаффолд [58]. В 2019 году было показано, что по крайней мере некоторые системы CRISPR/Cas IV типа аналогично I и III типам используют Cas6-индуцированный процессинг пре-crРНК [59]. Однако про этот тип систем до сих пор известно довольно мало. Нуклеазы V-A подтипа, известные как Cas12a, или Cpf1 (аббр. от англ. CRISPR-associated from *Prevotella* and *Francisella* — CRISPR-ассоциированный из *Prevotella* и *Francisella*), осуществляют процессинг пре-crРНК самостоятельно, внося разрывы на расстоянии 4 нуклеотидов в сторону 5'-конца от шпилечного псевдоузла, благодаря наличию РНКазного домена [60]. Интересно, что РНКазная активность Cas12a оказалась металл-независимой, причем 5'-концевой продукт разрезания пре-crРНК содержит 2'-3'-циклическую фосфатную группу [61]. Для некоторых других подтипов V типа также была показана способность нуклеаз Cas12

процессировать пре-crРНК (Cas12i1, Cas12c) [62], [63]. Причем механизмы процессинга отличаются от таковых для Cas12a. Например, Cas12c процессирует пре-crРНК в зрелые молекулы только в присутствии scoutРНК, но без помощи РНКазы III. Нуклеазы подтипов V-C и V-D процессируют crРНК с 3'-конца самостоятельно в присутствии scoutРНК, в отличие от Cas12a, осуществляющих процессинг с 5'-конца crРНК [64]. Cas12i процессирует пре-crРНК с 5'-конца, но разрезание осуществляет в месте петли, предшествующей стебель-петлевой структуре пре-crРНК, в отличие от Cas12a, который режет пре-crРНК рядом с 5'-псевдоузлом [63]. Интересный пример одной из наиболее компактных из известных нуклеаз Cas (~750 п.о.), найденной в геноме бактериофага клады Biggerphage и классифицированный как Cas12j (CasΦ), также способен к процессингу пре-crРНК, однако осуществляет его металл-зависимо с помощью нуклеазного домена RuvC [65]. Системы VI типа содержат в качестве эффекторного модуля моносубъединичные эндорибонуклеазы Cas13, которые имеют два различных активных сайта для металл-независимого процессинга пре-crРНК и гидролиза таргетной РНК соответственно [66]–[68].

1.2.4 Интерференция

Прошедшие процессинг зрелые crРНК в составе эффекторного комплекса направляют последний к целевому участку нуклеиновой кислоты посредством водородных связей между спейсерной частью crРНК и таргетом. Как уже указывалось выше, в большинстве случаев интерференционный модуль осуществляет направленный гидролиз только в случае наличия последовательности PAM, расположенной в непосредственной близости к протоспейсеру. PAM чаще всего отсутствует в собственном геноме хозяина, что позволяет микроорганизмам сохранять собственный генетический материал интактным. В системах CRISPR/Cas 1-го класса за узнавание PAM ответственны отдельные компоненты эффекторных комплексов, в то время как моносубъединичные эффекторы 2-го класса II и V типов содержат PAM-

узнающие P1 домены (P1 — аббр. от англ. PAM-interacting) в составе белковой глобулы. Требования к PAM могут различаться среди различных гомологов нуклеаз и требуют отдельной идентификации для каждого отдельного представителя ряда. Эндорибонуклеазы Cas13a не требуют наличия последовательности PAM, вместо нее должен присутствовать конкретный нуклеотид на некотором расстоянии от протоспейсера — PFS (аббр. от англ. protospacer flanking site — сайт, фланкирующий протоспейсер), тем не менее другие подтипы CRISPR/Cas13 систем могут не нуждаться в PFS для направленного гидролиза РНК [18]. Нуклеазы Cas13a систем VI типа при формировании дуплекса целевой оцРНК с сгРНК меняют свою конформацию, активируя каталитический РНКазный домен. Подобная активация приводит к неспецифическому разрезанию РНК, как таргетной, так и фаговых и хозяйских РНК, с которыми взаимодействует активированный фермент [68].

Мультисубъединичные эффекторные комплексы систем III типа способны разрезать и РНК, и ДНК фагов. Связываясь с транскрибируемой матричной РНК фагов благодаря сгРНК, комплекс деградирует выпетлившуюся оцДНК в транскрипционном пузыре, предотвращая тем самым дальнейший синтез транскриптов, а также синтезируя вторичный мессенджер — циклический олигоаденилат (сОА) [69]. Последний индуцирует активность вспомогательных РНКаз и ДНКаз, которые неспецифично деградируют доступные нуклеиновые кислоты. В целях предотвращения нацеливания нуклеаз на собственные транскрипты, вместо PAM данные системы используют 8-нуклеотидную последовательность скаффолда, комплементарную собственной РНК хозяина. В случае комплементарности эффекторный комплекс не активируется и интерференция не происходит [70].

В случае эффекторов, осуществляющих разрезание дцДНК, содержащая PAM цепь называется нетаргетной, в то время как вторая цепь носит название таргетной цепи. Связывание интерференционного модуля, заряженного малыми

РНК, необходимыми для конкретного подтипа CRISPR/Cas, с целевой последовательностью дцДНК вызывает локальное плавление ДНК и формирование R-петли, представляющей из себя дуплекс crРНК с таргетной цепью ДНК и выпетливающуюся нетаргетную цепь [71]. Эксперименты по изучению структуры эффекторного комплекса Cas9 показали, что при распознавании PAM, аминокислотная петля Lys1107-Glu1108-Ser1109 нуклеазы (петля «фосфатный замок»), взаимодействуя с фосфатом таргетной цепи в положении +1, поворачивает его, вызывая тем самым искажение в конформации таргетной цепи, которое делает возможным формирование водородной связи между нуклеотидом таргетной цепи в положении +1 и 20-м нуклеотидом спейсерной части crРНК и инициирует инвазию crРНК в дцДНК [72]. Таким образом, механизм инициации образования R-петли в случае систем CRISPR/Cas II типа зависит от узнавания PAM, поскольку в случае отсутствия индуцированного поворота фосфата в положении +1 вероятность формирования Уотсон-Криковской пары между нуклеотидом таргетной цепи в положении +1 и crРНК крайне низкая [73]. Формирование R-петли во время стадии интерференции, видимо, является общим свойством Cas-нуклеаз, гидролизующих дцДНК. Cascade-индуцированное образование R-петель было показано для систем I типа [74], а Cas12-индуцированное - для различных систем V типа [75], [76]. Подробно ферментативная активность систем V типа рассмотрена ниже.

1.2.5 CRISPR/Cas нуклеазы V типа

Как было указано выше, в 2013 году появились работы, в которых была показана принципиальная возможность использования систем CRISPR/Cas для редактирования клеток млекопитающих, в том числе клеток человека. В приведенных работах в качестве эффектора использовался единственный на тот момент используемый представитель систем CRISPR 2-го класса - Cas9, причем большинство работ было посвящено изучению работы ортолога SpCas9, как

исторически первой охарактеризованной Cas-нуклеазы из *Streptococcus pyogenes*. Двумя годами позже группа под руководством Zhang открыла другого представителя нуклеаз 2-го класса - белок Cas12a (он же Cpf1) [10]. Последний имел несколько важных отличительных особенностей относительно Cas9: в первую очередь АТ-богатую последовательность PAM, позволяющую направлять редактор на другие области генома, недоступные Cas9; ассиметричный характер разрезания целевой ДНК, позволяющий оставлять «висячие» концы; а также требование для своей работы лишь одной РНК (этим Cas12a выгодно отличается от Cas9). Помимо этого, при дальнейших исследованиях представителей большинства подтипов систем V типа выяснилось, что эти ферменты способны проявлять так называемую коллатеральную активность, то есть неспецифично гидролизовать одноцепочечные фрагменты ДНК (а в случае с некоторыми подтипами - РНК) в случае присутствия в образце целевой матрицы. Это свойство положило начало CRISPR-диагностике, основанной на различных модификациях тест-систем с использованием различных белков Cas12. Самые компактные из известных по длине аминокислотной последовательности направляемые нуклеазы тоже являются представителями CRISPR/Cas систем V типа. Всего на данный момент выделяют 11 подтипов систем V типа (Таблица 1). В отличие от Cas9, осуществляющих гидролиз таргетной молекулы дцДНК двумя нуклеазными доменами — по одному на каждую цепь ДНК, нуклеазы V типа имеют лишь один нуклеазный домен — RuvC. Механизм внесения DSB, таким образом, должен отличаться от такового у Cas9. Детально изучить работу Cas12 удалось благодаря структурным исследованиям, в большинстве своем использующих подтип V-A в качестве объекта.

Нуклеазы подтипа V-A структурно представляют из себя двудольные глобулы — с распознающей долей REC (от англ. Recognition — распознавание) и нуклеазной долей NUC (от англ. Nuclease — нуклеаза), соединенных

перемежающимся WED-доменом (от англ. Wedge - клин). Доля REC расположена на N-конце фермента и состоит из двух альфа-спиральных доменов REC1 и REC2, в то время как в составе C-концевой NUC доли представлены домены RuvC, Nuc, BH (аббр. от англ. Bridge helix - мостиковая спираль) и PI [77]. Доля REC связывает пре-crРНК и координирует гетеродуплекс crРНК с таргетной цепью ДНК. Нуклеазная активность осуществляется в пространстве между RuvC- и Nuc-доменами, причем обе цепи гидролизуются исключительно RuvC-доменом [61]. PI-домен, как указывалось выше, отвечает за связывание PAM, причем, в отличие от Cas9 с 3'-концевой последовательностью PAM, PI-домен Cas12a распознает PAM, находящийся с 5'-стороны от протоспейсера. Как было показано для FnCas12a, связывание PAM необходимо для локального расплетания дцДНК с дальнейшим формированием R-петли, то есть дуплекса crРНК с таргетной цепью ДНК. После формирования R-петли Cas12a вносит разрыв в выпетлившуюся нетаргетную цепь, причем последующее разрезание таргетной цепи происходит только в случае предшествующего гидролиза первой цепи [78] и гораздо медленнее его [79]. Возможность захвата таргетной цепи RuvC-доменом обусловлена асимметрией R-петли, проявляющейся в конформационном искажении 3'-концевой ее границы (в сайте связывания 3'-конца crРНК) после произошедшего гидролиза нетаргетной цепи [75]. С помощью метода магнитного молекулярного пинцета было показано влияние ароматического аминокислотного остатка в REC2-доме на скорость расплетания 3'-концевой границы R-петли, происходящего постепенно в Cas12 нуклеазах дикого типа. На этом основании была предложена модель, согласно которой полное расплетание R-петли приводит к сближению REC2 и Nuc. Последнее сокращает борозду между таргетной цепью и RuvC, но возможно только после внесения разрыва в нетаргетную цепь. Мутации в указанном ароматическом остатке приводят к практически идентичным скоростям разрезания обеих цепей, поскольку он является «зажимом», не позволяющим расплести дуплекс на

полную длину crРНК. Однако после внесения одноцепочечного разрыва в нетаргетную цепь «зажим» через некоторое время смещается, дуплекс crРНК-таргетная цепь становится протяженнее. Это индуцирует 3'-концевые искажения в R-петле и взаимные изменения конформации REC2 и Nuc. Более того, после разрезания таргетной цепи RuvC-домен остается активированным и осуществляет тримминг таргетной цепи до длины 4, 5 или 6 нуклеотидов выступающего 5'-конца. Механизм коллатеральной активности ДНКаз Cas12 по сути представляет собой тот же тримминг, поскольку активированный RuvC-домен остается доступен также для процессинга любой оцДНК *in trans* [80].

Нуклеазы остальных подтипов V типа систем CRISPR/Cas, для которых известны структуры комплексов, функционируют схожим с Cas12a образом и имеют единственный RuvC-домен. Однако некоторые детали взаимодействия с таргетом могут отличаться. Cas12b использует tracrРНК вместе с crРНК, и, будучи также двудольной глобулой, распознает tracrРНК с помощью доли REC, хотя и взаимодействует с ней в то же время доменами доли NUC [81]. Нуклеазы подтипов V-B и V-E не имеют PI домена, в их контактах с PAM задействованы разные домены долей REC и NUC. Cas12c использует RuvC-домен как для последовательного гидролиза дцДНК, так и для 3'-процессинга пре-crРНК [82]. Для Cas12d (CasY) и Cas12h структуры на данный момент не получены. Белки Cas12e (CasX) имеют небольшой процент идентичности с другими представителями нуклеаз V типа и обладают уникальной доменной организацией. Эти ферменты также используют единственный RuvC-домен для последовательного гидролиза дцДНК, однако практически не проявляют коллатеральной активности [83]. Нуклеазы Cas12f (Cas14) являются самыми компактными эффекторами из известных Cas-нуклеаз (~400-700 п.о. по сравнению с ~950-1400 п.о. для остальных Cas-нуклеаз 2 класса) [84]. Было показано, что они способны осуществлять направленный гидролиз дцДНК с помощью единственного RuvC-домена только в виде димеров, связывающих по

одной молекуле crРНК и tracrРНК на димер, и располагая гетеродимер crРНК с таргетной цепью ДНК в центральном канале между долями REC и NUC [85], [86]. Нуклеазы подтипа V-I являются РНК-направляемыми никазами, разрезающими нетаргетные цепи дцДНК, хотя они и способны крайне низкоэффективно вносить DSB; для успешного формирования гетеродуплекса crРНК с таргетной цепью ДНК длина спейсерной части crРНК Cas12i должна составлять 28 нуклеотидов в отличие от 20 нуклеотидов для многих других нуклеаз Cas12 [63]. Ферменты CasФ (Cas12j), как уже указывалось выше, помимо компактного размера уникальны тем, что используют RuvC-домен и для последовательного гидролиза дцДНК, и для процессинга пре-crРНК [65]. Нуклеазы Cas12k обладают нативно инактивированным RuvC-доменом и не способны гидролизовать ДНК, однако участвуют в узнавании ДНК для инсерции Tn7-подобных транспозонов [87]. Рибонуклеазы Cas12g нуждаются в tracrРНК и не имеют PI-домена, однако их RuvC-домен способен после узнавания таргетной молекулы РНК осуществлять коллатеральную активность как по отношению к РНК, так и оцДНК [88].

1.2.6 CRISPR диагностика

Использование различных свойств биомолекул в настоящее время позволяет быстро выявлять вирусные и бактериальные заболевания человека и животных. В клинической практике на данный момент используются классические подходы, использующие антитела против поверхностных эпитопов целевых агентов, и методы детекции нуклеиновых кислот, основанные на ПЦР. В ряде современных подходов используется детекция изменения вторичной структуры РНК; к таким подходам, например, относятся диагностические инструменты, основанные на toehold-подобных переключателях [89], [90]. Благодаря интенсивному развитию инструментов геномного редактирования в последнее время, экспресс-диагностика на основе систем CRISPR/Cas потенциально может служить альтернативой

«классическим» методам. Транс-репортер CRISPR с направляемой на ДНК эндонуклеазой (DETECTR — аббр. от англ. DNA endonuclease-targeted CRISPR trans-reporter) [91], специфическое высокочувствительное ферментативное разблокирование репортера (SHERLOCK — аббр. от англ. specific high sensitivity enzymatic reporter unlocking) [92], недорогие многоцелевые высокоэффективные системы за один час (HOLMES и HOLMESv2 — аббр. от англ. one-hour low-cost multipurpose highly efficient system) [93], [94], Cas12b-опосредованное обнаружение ДНК (CDetection — аббр. от англ. Cas12b-mediated DNA detection) используют свойства эндонуклеаз Cas для детекции целевых нуклеиновых кислот *in vitro* [95]. Чаще всего в предлагаемых подходах детектирующей реакцией является неспецифическое расщепление оцРНК или оцДНК, то есть коллатеральная активность ферментов V и VI типов. Если в образце присутствует интересующий фрагмент нуклеиновой кислоты, то Cas-эффектор, заряженный сгРНК или sgРНК, распознает его, гидролизует и, оставляя нуклеазный домен активным, способен деградировать любой олигонуклеотид в образце коллатерально. В реакции разрезания *in vitro* добавляют короткий олигонуклеотид, модифицированный флуорофором и гасителем - так называемый FQ-репортер (англ. fluorophore, quencher), после чего детектируют флуоресцентный сигнал в случае ранее произошедшего направленного гидролиза матрицы. Стоит отметить, что на фоне пандемии вируса SARS-CoV-2 было предложено множество различных модификаций приведенных методов диагностики, в том числе с колориметрической детекцией [96], [97]. Все эти системы являются довольно быстрыми - от минут до 1 часа, чувствительными - вплоть до субатомолярных концентраций целевой нуклеиновой кислоты и относительно несложными в исполнении в лабораторных условиях.

Глава 2. Материалы и методы

2.1.1 Поиск нуклеазы

Аминокислотная последовательность нуклеазы Cas12a из *Ruminococcus bromii* sp. была найдена в базе данных NCBI non-redundant (nr) protein database наряду с другими ортологами. Далее последовательности выгружались в алгоритм blastp (protein-protein BLAST) с настройками поиска в базе патентованных последовательностей белков NCBI Patented protein sequences (pataa). Результаты сначала ранжировались по возрастанию значения e-value и в дальнейшем анализе использовались только последовательности со значением идентичности менее 50 %. Для полученного набора последовательностей были найдены гены, кодирующие предполагаемые белки-кандидаты в Cas12a. Нуклеотидные последовательности найденных генов далее выгружались в алгоритм blastn (megablast) с настройками поиска в локальной базе секвенированных нуклеотидных последовательностей Курчатовского геномного центра. Контиги, содержащие результаты поиска, далее анализировались на предмет целостности локусов CRISPR и поиска последовательностей повторов CRISPR-массива.

2.1.2 Анализ CRISPR локусов

Контиги, содержащие результаты поиска, в формате fasta выгружались в онлайн-утилиту CRISPRCasTyper со стандартными настройками [98]. Среди полученных результатов отбирались только полноразмерные локусы CRISPR. Спейсеры из CRISPR-массивов затем анализировались с помощью CRISPRTarget со стандартными настройками для поиска нуклеотидных последовательностей возможных мобильных генетических элементов, чьи протоспейсеры были интегрированы в процессе CRISPR-адаптации [99].

2.1.3 Молекулярное клонирование RbCas12a для получения рекомбинантного препарата

Образцы ДНК, в которых на предыдущих этапах были найдены целевые локусы CRISPR, сначала анализировались для подтверждения наличия в них гена *cas12a*. Для этого была проведена полимеразная цепная реакция (ПЦР) со праймерами Fwd_test и Rev_test (последовательности олигонуклеотидов, используемых в рамках работы, приведены в Таблице 2). 50 мкл реакционной смеси смешивали с использованием Q5® High-Fidelity 2X Master Mix (New England Biolabs) в соответствии с рекомендациями производителя. Три из семи образцов, которые, как было обнаружено, содержат целевые фрагменты (~590 п.н.), в дальнейшем использовали в реакциях ПЦР для амплификации целого гена с 5'-фосфорилированными праймерами RbCpf_Fwd_Nde и RbCpf_Rev_Xho.

Фосфорилирование проводили при 10 мкМ концентрации каждого праймера с использованием T4 полинуклеотидкиназы (New England Biolabs) в соответствии с рекомендациями производителя. Результаты ПЦР визуализировали электрофоретически с окраской этидий бромидом (5 мкг/мл) в 1 % агарозном геле в трис-ацетатном буфере при напряженности электрического поля 3.8 В/см. Целевые фрагменты далее вырезали из агарозного геля и выделяли ДНК с помощью набора QIAGEN Gel Extraction Kit (Qiagen) в соответствии с рекомендациями производителя. Далее фрагменты, содержащие целевой ген, клонировали в составе линейаризованного плазмидного вектора pUC118 с использованием T4 ДНК-лигазы (New England Biolabs) в течение 16 часов при 16 °С. Плазмидный вектор pUC118 предварительно линейаризовали и дефосфорилировали с использованием эндонуклеазы рестрикции HincII и фосфатазы QuickCIP (New England Biolabs) в соответствии с рекомендациями производителя.

Полученной лигазной смесью трансформировали штамм химически-компетентных клеток *E. coli* TOP10 (Invitrogen), приготовленных по протоколу Inoue [100]. Трансформацию тепловым шоком проводили в соответствии с рекомендациями авторов протокола приготовления компетентных клеток. Бактериальную суспензию размазывали по чашкам Петри с агаризованной средой LB-Lennox, содержащей 100 мкг/мл ампициллина. На следующий день (через ориентировочно 14 часов) выросшие колонии подвергали скринингу с помощью метода ПЦР с использованием праймеров M13, после чего прошедшие скрининг клоны инокулировали в среду LB-Lennox и инкубировали в течение ночи при 37°C при 180 оборотах в минуту. Плазмидную ДНК выделяли из бактериальных суспензий с помощью набора Monarch® Plasmid Miniprep Kit (New England Biolabs) и секвенировали по Сэнгеру на капиллярном генетическом анализаторе AB3730xl (AppliedBiosystems). Для этого 230 нг плазмидной ДНК использовали в секвенирующих реакциях с использованием набора BrilliantDye (NimaGen) в соответствии с рекомендациями производителя. Реакционные смеси далее очищались методом переосаждения. Для этого готовили раствор из 3 мл 96 % этанола, 110 мкл 125 мМ ЭДТА, 110 мкл 3 М ацетата натрия (рН 4.6) и добавляли по 29 мкл этого раствора к 10 мкл реакционной смеси после секвенирующей реакции. Полученные смеси инкубировали в течение 20 минут при комнатной температуре, после чего центрифугировали в течение 20 минут при 4000 оборотов в минуту. Осадок далее однократно промывали 75 % этанолом, повторно центрифугировали при тех же параметрах, после чего 75 % этанол удаляли, и осадок высушивали в течение 5 минут при комнатной температуре. Осадок далее растворяли в 12 мкл формамида (AppliedBiosystems), денатурировали при 95 °С в течение 1 минуты и сразу использовали для анализа с использованием стандартного протокола электрофореза в 50 см капиллярах с полимером POP7 (AppliedBiosystems), предусмотренного производителем AB3730xl LongSeq_50_POP7 при

спектральной калибровке Z-BigDye_3.1. Хроматограммы анализировали с использованием программного обеспечения SeqA6 и SnapGene.

Плазмидные векторы из клонов с подтвержденной нуклеотидной последовательностью, а также плазмидный вектор pET30a+, далее гидролизовали с использованием рестриктаз NdeI и XhoI (New England Biolabs) в соответствии с рекомендациями производителя. Продукты гидролиза разделяли в 1 % TAE агарозном геле электрофоретически. Целевые фрагменты, содержащие ген *cas12a* и линейаризованный вектор pET30a+, выделяли из геля аналогично указанному выше методу, после чего лигировали аналогично предыдущему протоколу и трансформировали в штамм *E. coli* TOP10 как описано ранее, за исключением использования 100 мкг/мл канамицина вместо ампициллина. Для подтверждения успешной сборки проводили ПЦР скрининг, после чего прошедшие клоны инокулировали в LB-Lennox, инкубировали и выделяли плазмидную ДНК аналогично предыдущим экспериментам, проводили контрольную рестрикцию плазмидной ДНК и анализировали в 1 % TAE агарозном геле электрофоретически. Полученный плазмидный вектор RbCas12a_pET30a+ далее использовали для получения рекомбинантного препарата белка. Нуклеотидная и аминокислотная последовательности RbCas12a приведены в Приложении 3.

2.1.4 Молекулярное клонирование crPHK

Все плазмидные векторы, кодирующие crPHK, для трансфекции культуры клеток человека конструировали следующим образом. Сначала ген RFP под контролем CMV промотора был клонирован в составе вектора pSQT1313, приобретенного в Addgene (#53370). Затем с помощью сайт-направленного мутагенеза с использованием набора Q5® Site-Directed Mutagenesis Kit (New England Biolabs) клонировали скаффолд crPHK RbCas12a с двумя сайтами рестриктазы IIS Esp3I для дальнейшего клонирования спейсеров, получая

pSQT1313-CMV_RFP-U6_scaffold вектор. Все манипуляции по клонированию, трансформации, скринингам, выделению плазмидной ДНК и секвенированию по Сэнгеру проводили аналогично указанному выше протоколу. Все спейсеры клонировали в полученный вектор с использованием рестриктазы Esp3I (New England Biolabs) путем отжига двух комплементарных олигонуклеотидов с выступающими концами следующим образом. Пары олигонуклеотидов, содержащих спейсеры и выступающие липкие концы для клонирования по сайту рестрикции Esp3I, фосфорилировали в одной пробирке аналогично выше указанному в 40 мкл реакциях. Затем реакционные смеси нагревали до 95 °C в течение 5 минут и медленно снижали температуру в течение 14 одноминутных циклов при скорости 5 °C/цикл до 25 °C с использованием термоциклера BioRad CFX96. Полученные дуплексы олигонуклеотидов разводили в 200 раз водой *mq* и использовали по 2 мкл в качестве вставки для лигирования в 50 нг pSQT1313-CMV_RFP-U6_scaffold вектора.

2.1.5 Получение рекомбинантного RbCas12a

Для получения рекомбинантного препарата RbCas12a штамм *E. coli Rosetta (DE3)* трансформировали плазмидным вектором RbCas12a_pET30a+. Компетентные клетки *E. coli Rosetta (DE3)* готовили и трансформировали аналогично TOP10. После инкубации при 37 °C в течение ночи бактериальную суспензию высевали в свежую среду LB-Lennox и культивировали в течение 3 ч при 37 °C с последующим охлаждением до температуры 18 °C. После охлаждения к культуре добавляли 1 mM IPTG и инкубировали клетки при 18 °C в течение ночи. Осадок клеток, содержащих рекомбинантный белок (21.7 г из 3.5 литров среды), суспендировали в 100 мл буфера (40 mM Tris-HCl pH 7.8; 400 mM NaCl; 5 mM имидазола; 0.2% Triton-X100, 5 Мм β-меркаптоэтанол, 1 mM PMSF) и обрабатывали ультразвуком (дезинтегратор Branson Ultra, импульсы 2 с с интервалом 6 с, суммарная продолжительность 360 с) при 40 °C. Бактериальный лизат далее центрифугировали при 28000 g в течение 30 мин

при +4 °С и наносили супернатант на 5 мл металлохелатную колонку Ni-NTA Superflow (Qiagen), предварительно уравновешенную буфером Б (40 мМ Tris-HCl pH 7.8; 400 мМ NaCl; 5 мМ имидазол; 0.1% Triton-X100). Затем колонку промывали последовательно 40 мл стартового буфера Б, 40 мл буфера W (40 мМ Tris-HCl pH 7.8; 400 мМ NaCl; 40 мМ имидазол) и 15 мл буфера E (40 мМ Tris-HCl pH 7.8; 400 мМ NaCl; 300 мМ имидазол). Элюцию белка контролировали УФ-детектором при длине волны 280 нм. Чистоту белка во фракциях оценивали с помощью электрофореза в денатурирующем полиакриламидном геле по методу Лэммли и визуализацией окрашиванием Coomassie-250. Объединенные фракции концентрировали и переводили в 50 мМ натрий-фосфатный буфер с pH 7.4 с помощью хроматографии на колонке PD-10 Fast Desalting (GE LifeScience). Полученный раствор наносили на колонку MonoS 10/100 (GE LifeScience) и проводили элюцию линейным градиентом (0-0.6 М, 20 объемов колонки) концентрации NaCl в 50 мМ Na-фосфатном буфере (pH 7.4). Чистоту белка во фракциях оценивали, как описано выше. Фракции, содержащие целевой белок, объединяли, переводили в буфер S (50 мМ Tris-HCl pH 7.5, 500 мМ NaCl, 2 мМ ДТТ, 5% (об/об) глицерин), определяли концентрацию белка в растворе спектрофотометрическим методом при длине волны 280 нм и концентрировали с помощью центрифужного концентратора Corning с размером пор 30 кДа, после чего быстро замораживали жидким азотом небольшими аликвотами, которые хранили при -80 °С до использования. Для оценки степени олигомеризации итогового препарата его аликвоту подвергли гель-фильтрационной хроматографии на колонке Superdex 200GL 10/300, уравновешенной буфером S.

2.1.6 Получение библиотеки РАМ

Определение последовательности РАМ на предыдущем этапе ограничивалось биоинформатическим предсказанием, что хоть и сильно ускоряет работы по первичной характеристике фермента, все же является

неполным и грубым приближением. Для биохимического подтверждения ферментативной активности RbCas12a в отношении всех возможных РАМ проводили 2 реакции ПЦР, в которых в качестве матрицы использовался либо прямой, либо обратный 100-нуклеотидный олигонуклеотид, фланкированный участками, комплементарными праймерам M13 и содержащий 8 вырожденных оснований с 5'-конца от заданного протоспейсера, по факту являющихся потенциальной последовательностью РАМ. Реакции проводили с использованием Q5 ДНК-полимеразы в соответствии с рекомендациями производителя. Далее ампликоны, полученные после обеих реакций ПЦР, анализировали электрофоретически в 2 % ТАЕ-агарозном геле, окрашенном этидием бромидом, с последующей очисткой целевых фрагментов с использованием набора Qiagen Gel Extraction Kit в соответствии с инструкцией производителя. Концентрацию полученных фрагментов измеряли с помощью Qubit dsDNA HS Assay Kit (ThermoFisher Scientific).

Далее очищенную ДНК гидролизовали при помощи EaeI рестриктазы (New England Biolabs), переосаждали с помощью 96 % этанола и использовали в качестве вставки в реакции лигирования в линейизованный вектор pBR322 по сайтам EagI и PvuII (New England Biolabs) с использованием T4 ДНК-лигазы в течение 16 часов при 16 °С, после чего инактивировали фермент при 65 °С в течение 10 мин и переосаждали лигазную смесь 96 % этанолом. После промывки 75 % этанолом и растворения осадка в 12 мкл *mq*, полученной смесью трансформировали ультраэлектрокомпетентные клетки *E. coli* штамма Top10 с помощью электропоратора Bio-Rad Gene Pulser Xcell при 2500 вольт в кюветах с расстоянием 1 мм между электродами. Бактериальную суспензию далее ресуспензировали средой LB, нагретой до 37 °С и инкубировали при 37 °С и 200 оборотах в минуту в течение часа, после чего наносили на агаризованную LB-среду на чашках Петри в присутствии 100 мкг/мл ампициллина и инкубировали при 37 °С в течение 14 часов.

Ультразлектрокомпетентные клетки готовили по следующему протоколу. Стартовую культуру бактерий растили при 37 °С в течение ночи при 180 оборотах в минуту. С утра 2 мл суспензии инокулировали в 98 мл свежей среды и растили в тех же условиях до достижения значения оптической плотности $A_{260} = 0.45$, после чего инкубировали на льду в течение 15 минут. Далее клетки центрифугировали при 4 °С и 3000 g в течение 15 минут. Супернатант выливали, клетки ресуспендировали в автоклавированном холодном PBS буфере, затем снова центрифугировали при тех же условиях. После чего осадок ресуспендировали в 100 мл стерильного холодного 10 % глицерина, центрифугировали аналогично предыдущим шагам. Шаги с 10 % глицерином повторяли 4 раза, после чего осадок ресуспендировали в 400 мкл 10 % глицерина, делали 4 аликвоты по 100 мкл и каждую сразу использовали для электропорации. 1/40 часть каждой электропорированной аликвоты (~10 мкл) использовали для нанесения суспензии на агаризованные чашки Петри, остальную часть морозили в жидком азоте и хранили при -80 °С до использования.

Наутро считали количество выросших колоний, исходя из чего делали вывод об эффективности трансформации в единицах отношения количества колоний к объему электропорированной смеси (N/V). Поскольку клонированные олигонуклеотиды содержали 8 вырожденных нуклеотидов, абсолютное количество всех возможных перестановок составляло $4^8 = 65536$. Исходя из предположения о возможных потерях значимых вариантов на всех этапах пробоподготовки, количество анализируемых колоний брали минимум в 2,7 раза больше количества возможных вариантов. Для этого размораживали суспензию электропорированных клеток на льду до полной разморозки, отбирали необходимый объем и размазывали по чашкам Петри с агаризованной средой. На следующий день проводили смыв более чем 180 тысяч выросших колоний средой LB на шейкере-ротаторе, после чего сразу выделяли

плазмидную ДНК с использованием набора GenElute HP Plasmid Maxiprep Kit (Sigma Aldrich). Выделенную плазмидную ДНК далее использовали в экспериментах по *in vitro* разрезанию в качестве РАМ библиотеки.

2.2 Характеристика и определение оптимальных условий работы фермента RbCas12a

2.2.1 Получение crРНК для реакций разрезания *in vitro*

Все crРНК, включая направляющую РНК на РАМ-библиотеку, использованные в экспериментах по разрезанию *in vitro*, получали в ходе транскрипции *in vitro* следующим образом. Для каждой crРНК осуществляли дизайн уникального олигонуклеотида, содержащего обратнo-комплементарную последовательность уникального спейсера и одинакового для всех скаффолда, а на 3'-конце обратнo-комплементарную последовательность T7 промотора. В качестве дуплекс-формирующего олигонуклеотида, необходимого для работы T7 РНК-полимеразы, во всех реакциях использовали прямую последовательность T7 промотора (Fwd_T7_guide). Далее готовили смеси из уникальных олигонуклеотидов для каждой crРНК и Fwd_T7_guide. Для этого смешивали по 5.3 мкл 10 мкМ олигонуклеотидов и доводили с помощью *mq* объем реакционных смесей до 12.6 мкл. Далее смеси инкубировали при 95 °С в течение 5 минут, после чего оставляли при комнатной температуре на полчаса для осуществления отжига олигонуклеотидов и формирования дуплексов. По истечении указанного времени 12 мкл смеси использовали в качестве матрицы в реакциях *in vitro* транскрипции с использованием набора TranscriptAid T7 High Yield Transcription Kit (ThermoFisher Scientific). Реакции проводили в объеме 40 мкл при 10 мМ концентрации рибонуклеозидтрифосфатов (rNTP) при 37 °С в течение ночи. Утром к реакциям добавляли 4 ед. акт. ДНКазы I (New England Biolabs) и доводили объем до 100 мкл с помощью *mq*, после чего инкубировали при 37 °С в течение 15 минут. Далее РНК чистили с помощью Trizol Reagent

(ThermoFisher Scientific) в соответствии с рекомендациями производителя и растворяли осадок в 50 мкл DEPC воды (Sigma). Концентрацию РНК оценивали с помощью Qubit™ RNA High Sensitivity (HS) Assay Kit, после чего аликвотировали и хранили при -80 °С.

2.2.2 Подбор буфера для работы RbCas12a *in vitro*

Определение физико-химических параметров для проведения реакций разрезания *in vitro* начали с подбора оптимального буфера для работы RbCas12a. Для этого смешивали рекомбинантный препарат RbCas12a, соответствующую crРНК и линейризованную по сайту ScaI рестриктазы (New England Biolabs) ДНК-матрицу в виде РАМ-библиотеки в соотношении 3:30:1 соответственно, с конечной концентрацией РНК 500 нМ, смесь далее инкубировали при 37 °С в течение 30 минут, после чего инактивировали нуклеазу добавлением 1 ед. акт. протеиназы К (New England Biolabs). В качестве буферов тестировали коммерчески-доступные буферы, упомянутые в работах по ортологичным ферментам: Nebbuffer1.1, Nebbuffer2.1, Nebbuffer3.1, Nebbuffer4, CutSmart, а также эти буферы с добавлением 5 мМ дитиотрейтола (DTT) и дополнительно буфер для разрезания (100 мМ HEPES pH 7.5; 500 мМ KCl; 25 мМ MgCl₂; 5 мМ DTT). Эффективности разрезов определяли с помощью программного обеспечения Bio-Rad ImageLab при подсчете интенсивностей полос на электрофореграммах.

2.2.3 Определение температурного оптимума RbCas12a

После сравнения буферных условий определяли температурный оптимум работы фермента. В первую очередь была определена точка плавления фермента RbCas12a с использованием метода термофлуоресцентного анализа. Для этого использовали флуоресцентный краситель ProteOrange Protein Gel Stain 5000× (Lumiprobe RUS Ltd) и проводили реакции в следующих условиях: 9 мкг RbCas12a смешивали в однократном буфере Nebbuffer 2 (последний в

отличие от 2.1 версии не содержит бычьего сывороточного альбумина BSA) с однократной концентрацией красителя в трех повторах в 25 мкл реакциях. Затем получали кривые плавления при постепенном повышении температуры от 4 °С до 95 °С с шагом 0.5 °С в секунду и регистрацией сигнала каждые 10 секунд на приборе Bio-Rad CFX96.

Сравнение эффективностей разрезания RbCas12a *in vitro* дцДНК на комнатной температуре проводили с гомологом из *Acidaminococcus sp. BV3L6* (AsCas12a). Для этого получали рекомбинантный препарат AsCas12a с использованием плазмидного вектора 6His-MBP-TEV-huAsCpf1 (#90095) из депозитария Addgene. Выделение и очистку белка проводили в соответствии с методикой, описанной в Zetsche et al. 2017 . Реакции разрезания *in vitro* для RbCas12a и AsCas12a проводили в присутствии 500 нМ crРНК и соотношении компонентов реакций 1:3:30 (ДНК:Cas12a:crРНК).

Далее определяли ферментативную активность RbCas12a на разных температурах. В качестве стандартного буфера для разрезания *in vitro* брали Nebbuffer 2.1. Для чистоты эксперимента было решено преднагревать по отдельности все компоненты реакции разрезания *in vitro* до соответствующих температур, преинкубировать рибонуклеопротеиновый комплекс на данной температуре, а затем добавлять в него ДНК матрицу, также с соответствующей температурой. Одновременное выполнение этих условий производилось с использованием роботизированной станции Tecan с последующей инактивацией реакции добавлением 33 мМ ЭДТА, после чего образцы анализировались электрофоретически. Для определения температурного оптимума использовали точки в 15 °С, 20 °С, 25 °С, 30 °С, 37 °С, 42 °С, а также 48 °С, которая выше полученного значения 45.5 °С при термофлуоресцентном анализе. Каждый из компонентов реакции в однократном буфере Nebbuffer 2.1: направляющая РНК, ДНК и RbCas12a преднагревали на соответствующих температурах в течение 5 минут, затем смешивали РНК и белок и инкубировали

далее 15 минут для успешного формирования рибонуклеопротеинового комплекса, после чего добавляли преднагретую ДНК, смешивали и инкубировали в течение 30 минут, завершая добавлением 33 мМ ЭДТА для инактивации комплекса. Соотношения ДНК:РНК:белок использовали 1:30:3. В качестве матрицы использовали ампликоны, полученные в реакциях ПЦР с геномной ДНК, выделенной из клеточной линии T-Rex (T-REx™-293 Cell Line, Invitrogen) с помощью набора GeneJET Genomic DNA Purification Kit (Thermo Scientific) в соответствии с рекомендациями производителя, с помощью праймеров Fwd_gDNA_site1 и Rev_gDNA_site1 и содержащие протоспейсер и прилегающий к нему 5'-TTC-3' PAM, и crРНК crRNA_site1. Фрагменты анализировали электрофоретически в 1x TAE 1 % агарозном геле с добавлением этидия бромида.

2.2.4 Определение кофактора RbCas12a

Чтобы более полно охарактеризовать полученную нуклеазу, были определены двухвалентные катионы, которые могут способствовать катализу разрезания ДНК эффекторным комплексом. Семь ионов, включая Ca²⁺, Ni²⁺, Mn²⁺, Mg²⁺, Zn²⁺, Cu²⁺ и Co²⁺ были изучены на предмет возможности играть роль кофактора. Реакции ставили в присутствии 10 мМ растворимых солей указанных катионов в буфере Nebbuffer 2 без магния при соотношении ДНК:РНК:белок 1:30:3 при 500 нМ РНК. Реакции проводили при 37 °С в течение получаса, после чего их инактивировали добавлением 33 мМ ЭДТА и анализировали электрофоретически в 1x TAE 1% агарозном геле с окраской этидием бромидом.

2.2.5 Определение последовательности PAM

Полученную на предыдущих этапах PAM-библиотеку гидролизовали рибонуклеопротеиновым комплексом RbCas12a и набором рестриктаз для получения фрагментов целевой длины: сначала матрицу разрезали с помощью

HindIII и NdeI, после чего целевые фрагменты анализировались электрофоретически и очищались из геля аналогично методам выше. Далее проводили 3 независимых реакции разрезания для получения фрагментов необходимой длины для приготовления из них NGS-библиотек: 1) рестрикция исходной PAM-библиотеки по сайтам BstZ17I и ScaI; 2) рестрикция и *in vitro* разрезание PAM-библиотеки по сайтам BbsI, DrrI; 3) рестрикция и *in vitro* разрезание PAM-библиотеки по сайтам BbsI, HinfI, Esp3I, PflFI. Реакция 1) позволила получить фрагмент, содержащий все имеющиеся последовательности PAM в полученной интактной PAM-библиотеке. Реакция 2) позволила получить 5'-фрагмент после разрезания. Реакция 3) позволила получить 3'-фрагмент после разрезания. Гидролизованная ДНК далее анализировались электрофоретически в 1x TAE 2.5 % агарозном геле, из которого выделяли порезанные и непорезанные фрагменты набором QIAEX II Gel Extraction Kit (Qiagen), после чего готовили NGS-библиотеки с помощью Ultra II Library Prep (New England Biolabs) в соответствии с рекомендациями производителя. Качество и длины полученных библиотек оценивали с помощью Bioanalyzer 2100 (Agilent) и набора High Sensitivity DNA Kit (Agilent) в соответствии с рекомендациями производителя, после чего секвенировали на платформе MiSeq (Illumina). Первичный анализ отобранных прочтений приемлемого качества, тримминг и выравнивание проводили в программном обеспечении CLC Genomics, после чего полученные выравнивания загружали в онлайн-утилиту WebLogo3 [101], [102].

Полученные статистические результаты далее отдельно проверяли в экспериментах по разрезанию *in vitro* в 2 этапа. Для этого на первом этапе сначала получили синтетические конструкции, содержащие один протоспейсер, но разные 3'-концевые нуклеотиды в последовательности PAM 5'-TTN-3', путем проведения сайт-направленного мутагенеза вектора, кодирующего известный протоспейсер с 5'-TTC-3' PAM. Используя 3 пары праймеров, вносящих

мутации в третий нуклеотид PAM, и набор Q5® Site-Directed Mutagenesis Kit, получили 4 генетические конструкции, содержащие один протоспейсер и один из четырех PAM: 5'-TTC-3', 5'-TTG-3', 5'-TTA-3', 5'-TTT-3'. Для этого лигазные смеси трансформировали в компетентные клетки *E. coli* штамма *Top10*, бактерии высаживали на агаризованную среду LB с добавлением 100 мкг/мл ампициллина и далее анализировали плазмидную ДНК аналогично описанным выше методам. Отсутствие мутаций проверяли с помощью секвенирования по Сэнгеру аналогично предыдущим этапам. Полученные векторы гидролизовали в трех повторах в условиях, аналогичным выше указанным, с использованием RbCas12a с направляющей РНК crRNA_site1 и ScaI рестриктазы (New England Biolabs). Результаты гидролиза анализировали электрофоретически в 1х TAE 1 % агарозе с окраской этидием бромидом.

Чтобы подтвердить данные, полученные после NGS и отображенные на WebLogo графике, на втором этапе к плазмидному вектору pUC119 были подобраны 13 crРНК на 5'-YYN-3' PAM: 12 РНК на 5'-CTN-3' и 5'-TCN-3' PAM, и одна дополнительная на 5'-TTC-3' PAM в качестве контроля. Используя полученные crРНК, ставили реакции разрезания *in vitro* с RbCas12a и HindIII рестриктазой (New England Biolabs) в трех повторах в условиях, аналогичным выше. Результаты гидролиза анализировали электрофоретически в 1х TAE 1 % агарозе с окраской этидием бромидом.

2.2.6 Изучение коллатеральной активности RbCas12a

Описанная в Главе 1 способность к неспецифическому гидролизу Cas-нуклеаз, называемая коллатеральной активностью, на данный момент показана только для представителей ортологичных рядов Cas12 и Cas13 [92], [94]. Поскольку RbCas12a принадлежит к одному из этих рядов, было решено выяснить, способен ли данный фермент проявлять похожую активность, то есть может ли потенциально использоваться в разработке тест-систем, основанных

на принципе коллатеральной активности. Для этого в реакции разрезания *in vitro* добавляли одноцепочечный ДНК-зонд ssDNA_FQ_rep, несущий на 5'-конце флуорофор ROX, а на 3'-конце тушитель флуоресценции BHQ-2. Матрицу для разрезания амплифицировали с помощью коммерческой RPA системы набором TwistAmp Basic (TwistDx) в соответствии с рекомендациями производителя в 10 мкл смеси. Затем 5 мкл брали в качестве матрицы для разрезания *in vitro*, которое проводили в соотношении РНК:белок 10:6 при 500 нМ направляющей РНК и 250 нМ зонда в течение 30 мин при 37 °С. Реакции проводили в 3 повторах. Детекцию флуоресценции проводили с помощью планшетного флуориметрического ридера Varioscan Lux (ThermoFisher Scientific).

2.2.7 Определение природы выступающих концов двуцепочечной ДНК после направленного гидролиза матрицы с помощью RbCas12a

Известно, что Cas12a формируют выступающие липкие концы после разрезания двуцепочечной матрицы, причем концы обязательно выступают с 5'-конца [10], [75]. Такой паттерн характерен благодаря структуре этих нуклеаз, а также специфической последовательности одноцепочечных разрывов, вносимых в двуцепочечную матрицу для формирования разрыва [78], [80]. Для определения природы выступающих концов, формируемых RbCas12a, был проведен ряд экспериментов. Плазмидный вектор pUC119 был гидролизован с помощью RbCas12a и рестриктазы AlwNI (New England Biolabs) таким образом, что фрагмент, кодирующий ген бета-лактамазы с промотором, обеспечивающий устойчивость бактерий к ампициллину, электрофоретически разделили от остального каркаса в 1x TAE 1 % агарозе и выделили с помощью набора Gel Extraction Kit (Qiagen) в соответствии с инструкцией производителя. Концентрацию полученных фрагментов измеряли с помощью Qubit dsDNA HS Assay Kit (ThermoFisher Scientific). Дополнительно в качестве контроля гидролиз проводили в присутствии FnCas12a вместо RbCas12a. Для этого

сначала получили рекомбинантный препарат FnCas12a, используя плазмидный вектор pET28_FnCas12a_TEV, приобретенный в Addgene (item #174658). Белок получали в штамме *E. coli Rosetta(DE3)* аналогично методам для RbCas12a. Очистку белка проводили в соответствии с методами статьи Enghiad et al [103]. Выделенные из геля фрагменты обрабатывали с помощью NEBNext® Ultra™ End Repair/dA-Tailing Module (New England Biolabs) в соответствии с рекомендациями производителя для «затупления» выступающих концов и их аденилирования. Затем реакции очищали с помощью NEBNext Sample Purification Beads (New England Biolabs), измеряли концентрацию с помощью Qubit dsDNA HS Assay Kit и клонировали в составе pKan-T вектора (Евроген) с использованием T4 ДНК-лигазы. Далее проводили трансформацию *E. coli Top10* с высевом на агаризованную среду LB-Lennox с 50 мкг/мл ампициллина и 50 мкг/мл канамицина, ПЦР-скрининг, контрольную рестрикцию и секвенирование по Сэнгеру аналогично описанным выше экспериментам.

2.2.8 Разделение направляющей РНК на две молекулы

Еще одним важным аспектом проведенных экспериментов является изучение обнаруженной способности различных ортологов Cas12a использовать в качестве направляющей РНК не одну цельную молекулу, а две молекулы, представляющие собой отдельные спейсер и скаффолд РНК (сплит-версию). В ходе работы было изучено влияние укорочения направляющей РНК и расщепление ее на две молекулы на *in vitro* разрезание матрицы различными нуклеазами Cas12a — AsCas12a, FnCas12a, LbCas12a, RbCas12a. Для оценки специфической и неспецифической ДНКазной активности AsCas12a, LbCas12a и FnCas12a использовали матрицы как одноцепочечной ДНК, так и двуцепочечной ДНК. Одноцепочечная ДНК вириона бактериофага φX174 была приобретена в New England Biolabs. Дуплексы олигонуклеотидов ДНК были сформированы путем смешивания меченой цепи и немеченой комплементарной цепи при молярном соотношении 1:1.5 в безнуклеазном дуплексном буфере

(IDT), дальнейшего нагревания в течение 2 минут при 95 °С и медленном охлаждении до 20 °С. Комплексы получали в реакционном буфере путем смешивания одной из нуклеаз Cas12a и соответствующей полноразмерной направляющей РНК или двух сплит-РНК в молярном соотношении 1:2 с последующей 20 минутной инкубацией при комнатной температуре. Все реакции разрезания *in vitro* проводили в 10 мкл буфера, содержащего 40 мМ Трис-НСl (рН 7.5), 50 мМ NaCl, 5 мМ MgCl₂, 0,5 мМ ТСЕР, 0,1 мг/мл BSA, при 37 °С в течение 30 мин. Неспецифическое расщепление 10 нМ ДНК фХ174 проводили комплексом 25 нМ Cas12a-направляющая РНК в присутствии 60 нМ целевой одноцепочечной ДНК или двуцепочечной ДНК. Реакции останавливали с помощью PurpleDye (New England Biolabs), содержащим 0,4 мг/мл гепарина, затем продукты разрезания разделяли электрофоретически в 1х TAE 1 % агарозном геле, предварительно окрашенном этидием бромидом. Чтобы сравнить связывание полноразмерной и сплит-версии направляющей РНК с AsCas12a, был проведен анализ расщепления ДНК *in vitro* в присутствии возрастающих количеств другой направляющей РНК (далее именуемой «конкурентная РНК»), которая нацелена на последовательность, отсутствующую в матрице ДНК. Конкурентная РНК была добавлена к 30 нМ AsCas12a вместе с полноразмерной, усеченной до 15 нуклеотидов или сплит-версии направляющей РНК с последующим добавлением целевой матрицы. Чтобы проверить, может ли коллатеральная активность AsCas12a поддерживаться сплит-версией направляющей РНК, был предварительно сформирован комплекс AsCas12a со сплит-версией направляющей РНК, который инкубировали с одноцепочечной ДНК или двуцепочечной ДНК в присутствии одноцепочечной ДНК фХ174, которая не имела последовательностей, соответствующих направляющей РНК. Для проверки специфической активности разрезания одноцепочечной ДНК с помощью AsCas12a, заряженной сплит-версией направляющей РНК, использовали 3'-

концевые флуоресцеин-меченые субстраты ДНК. Соответствующие субстраты двуцепочечной ДНК использовали в качестве контролей.

Дополнительно была протестирована возможность использования сплит-версии направляющей РНК в разрезания *in vitro* дцДНК с помощью RbCas12a. Для этого получали ампликоны с использованием Q5 ДНК-полимеразы (New England Biolabs) и очисткой их из геля с помощью набора Qiagen Gel Extraction Kit в соответствии с инструкциями производителя с использованием праймеров Fwd_gDNA и Rev_gDNA. Направляющие РНК crRNA_site1, crRNA_site2, crRNA_site3 получали *in vitro* транскрипцией, аналогично описанным выше методам. Спейсерная РНК crRNA_spacer1 была заказана в Lumiprobe Rus Ltd.

2.3.1 Трансфекция культуры клеток человека

Для трансфекции клеток человека была выполнена оптимизация кодонов нуклеотидной последовательности RbCas12a компанией DAPCEL, Inc. путем выбора синонимичных кодонов для обеспечения правильной котрансляционной укладки белков. Синтез плазмидного вектора pcDNA3.1+, кодирующего кодон-оптимизированный hRbCas12a с тремя копиями SV40 сигнала ядерной локализации (NLS) на N-конце и одной копией NLS нуклеоплазмина *Xenopus laevis* sp. и HA-метки на C-конце, был проведен компанией GenScript LCC. Итоговая конструкция 3xSV40_NLS-hRbCas12a-NLS-HA использовалась далее для трансфекции линии клеток HEK293T. Для трансфекции иммортализованной культуры клеток человека HEK293T, клетки выращивали в среде DMEM (Gibco) с добавлением 10% фетальной бычьей сыворотки FBS (Gibco) при 37 °C и 5 % концентрации CO₂. Трансфекцию проводили с использованием Lipofectamine3000 (Invitrogen) в соответствии с рекомендациями производителя.

2.3.2 Детекция мРНК RbCas12a в культуре клеток HEK293T

Экспрессию мРНК RbCas12a проверяли методом ПЦР «в реальном времени» с помощью EvaGreen Supermix (Bio-Rad) через 48 часов после трансфекции либо 500 нг, либо 2500 нг плазмидной ДНК в 6-луночной планшете при 90% конфлюэнтности. Для этого клетки снимали, трижды промывали PBS, затем выделяли тотальную РНК с помощью Trizol Reagent по инструкции производителя. После этого в реакциях обратной транскрипции получали кДНК с использованием набора ProtoScript® II First Strand (New England Biolabs). ПЦР-амплификацию кДНК в режиме реального времени проводили в 20 мкл реакционной смеси, содержащей 10 мкл EvaGreen Supermix, 500 нМ каждого праймера Fwd_RbCpf_RT и Rev_RbCpf_RT, 100 нг кДНК и свободную от РНКаз воду (Евроген). Условия термоциклирования кДНК были следующими: 95 °С в течение 3 мин, 39 циклов 95 °С в течение 10 с, 56 °С в течение 5 с и 68 °С в течение 10 с. Амплификацию и детекцию проводили с помощью термоциклера Bio-Rad CFX96.

2.3.3 Формирование делеций в геноме культуры клеток HEK293T с помощью RbCas12a

После проверки экспрессии гена RbCas12a в клетках млекопитающих на уровне мРНК, нуклеазу тестировали на предмет возможности формирования делеций в геноме. Для этого получили две генетические конструкции, кодирующие направляющие РНК на ген *DNMT1*, способом, описанным в п.2.1.4. Подбор направляющих РНК осуществляли с помощью онлайн-сервиса Breaking-Cas [104]. Затем оба вектора вместе с конструкцией, кодирующей кодон-оптимизированную нуклеазу RbCas12a, смешивали и трансфецировали смесью культуру клеток HEK293T аналогично экспериментам, описанным в п.2.3.1. 1195 нг плазмидной ДНК 3xSV40_NLS-hRbCas12a-NLS-NA и по 488 нг каждого вектора, кодирующего одну из двух crРНК использовали для нацеливания на ген *DNMT1*. Через 48 и 96 часов клетки снимали, трижды промывали PBS и лизировали в буфере NP-40 (Invitrogen) в соответствии с

рекомендациями производителя. Далее полученные лизаты разводили стерильной водой, измеряли концентрацию ДНК с помощью набора Qubit dsDNA BR Assay Kit (ThermoFisher Scientific) и использовали в качестве матрицы для ПЦР с Q5 ДНК-полимеразой и праймерами Fwd_dnmt1_del_ngs и Rev_dnmt1_del_ngs, отжигающимися на участок предполагаемой делеции. Затем фрагменты анализировали в 1x TAE 1 % агарозном геле электрофоретически с окраской этидием бромидом. Появление в геле полос совпадающих с ожидаемой и дополнительной, означает успешное формирование делеции. Делетированные фрагменты были выделены из геля аналогично описанному выше, и далее клонированы в составе плазмидного вектора pUC119. Далее лигазные смеси трансформировали в *E. coli Top10* и анализировали аналогично другим экспериментам. Из ампликонов с делецией и без нее готовили NGS библиотеки с помощью NEBNext® Ultra™ II DNA Library Prep (New England Biolabs), анализировали на Bioanalyzer 2100 и секвенировали с помощью системы Illumina MiSeq. Данные NGS анализировали в программном обеспечении CLC Genomics и в геномном браузере UGENE [105].

Таблица 2. Названия и последовательности олигонуклеотидов, использованных в работе

Название	Последовательность 5'->3'
Fwd_test	GGATGCAACTGAATCCTGTAG
Rev_test	TACCGCATAAGCCGGAATAAG
Fwd_RbCpf_Nde	TGAACATATGATGCAAGAGCGTAA AAAAATATCGCATC

Rev_RbCpf_Xho	ATCTACTCGAGATTATTCGCCATATC ATTCTCCTGAACA
Fwd_NLS_Nhe_Hind	CTAGATGCCGAAGAAAAAGCGCAA GGTCA
Rev_NLS_Nhe_Hind	AGCTTGACCTTGCGCTTTTTCTTCG GCAT
Fwd_NLS_Hind	AGCTCCCGAAGAAAAAGCGCAAG GTCA
Rev_NLS_Hind	AGCTTGACCTTGCGCTTTTTCTTCG GG
Fwd_dnmt1_spacer1	TAGAAAGGAAGTCTTGGCTGGCCT TCC
Rev_dnmt1_spacer1	AAAAGGAAGGCCAGCCAAGACTTC CTT
Fwd_dnmt1_spacer2	TAGATCACGGGACTTCTGGCTGAG GTCA
Rev_dnmt1_spacer2	AAAATGACCTCAGCCAGAAGTCCC GTGA
Fwd_dnmt1_spacer3	TAGATCTGATGGTCCATGTCTGTTA CTC
Rev_dnmt1_spacer3	AAAAGAGTAACAGACATGGACCAT CAGA
Fwd_vegfa_spacer	TAGATCTAGGAATATTGAAGGGGGC AGG

Rev_vegfa_spacer	AAAACCTGCCCCCTTCAATATTCCT AGA
Fwd_emx1_spacer	TAGATTCCTCCGGTCTGGAACCAC ACC
Rev_emx1_spacer	AAAAGGTGTGGTTCAGAACCGGA GGAA
Fwd_dnmt1_del_ngs	GACAGACACACGGAGTGTCTAGCT
Rev_dnmt1_del_ngs	AATTTGGCTCAGCAGGCACC
Fwd_dnmt1_sp1_ngs	CAAGGGCAGCTCAGTGGTACTT
Rev_dnmt1_sp1_ngs	CTGAGTCACGTGAGTTGATCCCCAT
Fwd_dnmt1_sp2_ngs	GACAGACACACGGAGTGTCTAGCT
Rev_dnmt1_sp2_ngs	GAAGCTGTTGTGTGAGGTTTCGCTTA TC
Fwd_dnmt1_sp3_ngs	CAGAACTAGTCCTTAGCAGCT
Rev_dnmt1_sp3_ngs	CTCTGGGGACCGTTTGAG
Fwd_vegfa_ngs	GGGTCACTCCAGGATTCC
Rev_vegfa_ngs	CCAAGGTTACAGCCTGAAA
Fwd_emx1_ngs	GCCTCCTGAGTTTCTCATCTG
Rev_emx1_ngs	CTAGTCATTGGAGGTGACATCG
Fwd_SDM_T	TTTGGAATTTGTGCCACTTCTG
Fwd_SDM_A	TTTGGAATTTAGTGCCACTTCTG
Fwd_SDM_C	TTTGGAATTCGTGCCACTTCTG

Rev_SDM	GGTTTACCTTGACCCCTATAG
Fwd_PAM_library	GTAAAACGACGGCCAGTGCCGCAG TACTGATCATNNNNNNNNNCCCCTCT ATTGATCCCCACCTCCAAATATCTCA TCAACAACGTCATAGCTGTTTCCTG
Rev_PAM_library	CAGGAAACAGCTATGACGTTGTTG ATGAGATATTTGGAGGTGGGGATCA ATAGAGGGGNNNNNNNNNATGATCA GTACTGCGGCACTGGCCGTCGTTTT AC
Fwd_RbCpf_RT	AGTGACGACTTGAAGGCTGA
Rev_RbCpf_RT	CGTCACATGGCAGCTTAGAC
Fwd_T7_guide	TAATACGACTCACTATAGG
Rev_TCA_guide	CTAAGAAACCATTATTATCAATCTAC AAGAGTAGAAATTACCCTATAGTGA GTCGTATTA
Rev_TCG_guide	GGGTTCCGCGCACATTTCCCATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
Rev_TCC_guide	CGTCAGGTGGCACTTTTCGGATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
Rev_TCT_guide	AAAGTGCCACCTGACGTCTAATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
Rev_CTA_guide	TATTTAGAAAATAAACAAAATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA

Rev_CTG_guide	AATAATGGTTTCTTAGACGTATCTAC AAGAGTAGAAATTACCCTATAGTGA GTCGTATTA
Rev_CTC_guide	CCGTCATCACCGAAACGCGCATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
Rev_CTT_guide	AAAAGTGCCACCTGACGTCTATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
Rev_CCA_guide	ACATTTCCCCGAAAAGTGCCATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
Rev_CCG_guide	AGACGTCAGGTGGCACTTTTATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
Rev_CCC_guide	GACGTCAGGTGGCACTTTTCATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
Rev_CCT_guide	ATAATGGTTTCTTAGACGTATCTAC AAGAGTAGAAATTACCCTATAGTGA GTCGTATTA
Rev_TTC_guide	GTCAGGTGGCACTTTTCGGGATCTA CAAGAGTAGAAATTACCCTATAGTG AGTCGTATTA
crRNA_site1	rUrArArUrUrUrCrUrArCrUrArUrUrGrU rArGrArUrCrCrCrCrUrCrUrArUrUrGrA rUrCrCrCrCrArCrC
Fwd_gDNA_site1	TCTTGCACTCATGAGCTGTC
Rev_gDNA_site1	GTTGAGGGTTATGAGAGTAGC
Fwd_gDNA	GATTTGAGAAGCCTTCGCTTC

Rev_gDNA	AAGTGGAGTCCGTAAGAGG
crRNA_site2	rUrArArUrUrUrCrUrArCrUrArUrUrGrU rArGrArUrArCrUrGrGrGrUrCrUrCrUrA rUrUrUrUrArCrCrC
crRNA_site3	rUrArArUrUrUrCrUrArCrUrArUrUrGrU rArGrArUrGrCrUrGrUrUrCrCrCrCrArA rCrCrUrUrUrUrCrC
crRNA_site4	rUrArArUrUrUrCrUrArCrUrCrUrUrGrU rArGrArUrGrArGrArArGrUrCrArUrArU rArArUrArArGrGrUrA
crRNA_competitor	rUrArArUrUrUrCrUrArCrUrCrUrUrGrU rArGrArUrGrCrUrGrUrUrCrCrCrCrArA rCrCrUrUrUrUrCrC
crRNA_scaffold	rUrArArUrUrUrCrUrArCrUrCrUrUrGrU rArGrArU
crRNA_spacer1	rCrCrCrCrUrCrUrArUrUrGrArUrCrCrCr CrArCrC
crRNA_spacer4	rGrArGrArArGrUrCrArUrArUrArArUrA rArGrGrUrA
«-15» crRNA	rUrCrUrArCrUrCrUrUrGrUrArGrArUrC rCrCrCrUrCrUrArUrUrGrArUrCrCrCrCr ArCrC
«-17» crRNA	rUrUrUrCrUrArCrUrCrUrUrGrUrArGrA rUrCrCrCrCrUrCrUrArUrUrGrArUrCrCr CrCrArCrC
«-7» crRNA	rUrGrUrArGrArUrCrCrCrCrUrCrUrArU rUrGrArUrCrCrCrCrArCrC
«-2» crRNA	rArUrCrCrCrCrUrCrUrArUrUrGrArUrC rCrCrCrArCrC
«+2» crRNA	rUrArArUrUrUrCrUrArCrUrCrUrUrGrU

	rArGrArUrCrC
Lb crRNA	rUrArArUrUrUrCrUrArCrUrArArGrUrG rUrArGrArUrGrArGrArArGrUrCrArUrA rUrArArUrArArGrGrUrA
Fn crRNA	rUrArArUrUrUrCrUrArCrUrGrUrUrGrU rArGrArUrGrArGrArArGrUrCrArUrArU rArArUrArArGrGrUrA
NTS_site4	TACCCGGAGATCGTTTAGAGAAGT CATATAATAAGGTAAGTGT AGAAGCTTGG
TS_site4	CCAAGCTTCTAACAGTTACCTTATT ATATGACTTCTCTAAACGATCTCCG GGTA
NTS_site4_labeled	36FAM- TACCCGGAGATCGTTTAGAGAAGT CATATAATAAGGTAAGTGT AGAAGCTTGG
TS_site4_labeled	36FAM- CCAAGCTTCTAACAGTTACCTTATT ATATGACTTCTCTAAACGATCTCCG GGTA

Глава 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1 Получение плазмидных конструкций

Результаты выдачи CRISPRTarget позволили идентифицировать T-богатую последовательность PAM для нуклеазы Cas12a из *Ruminococcus bromii sp.* для большинства строк выдачи. T-богатые PAM являются общим признаком нуклеаз подтипа V-A, что соотносится с полученными данными.

Последовательности гена RbCas12a длиной 3735 п.о. и консенсуса повтора из CRISPR-массива (то есть скаффолда crPНК) далее использовались для молекулярного клонирования векторов. Вектор RbCas12a_pET30a+, кодирующий целевой белок с гексагистидиновой меткой (6xHis-tag) на С-конце, далее использовался для получения рекомбинантного RbCas12a. Вектор pSQT1313-CMV_RFP-U6_scaffold далее использовался в экспериментах по клонированию спейсеров crPНК для экспрессии в культуре клеток HEK293T. Нуклеотидная и аминокислотная последовательности RbCas12a приведены в Приложении 3.

3.2 Получение рекомбинантного препарата RbCas12a

Чистоту рекомбинантного препарата RbCas12a после элюции с Ni-NTA Superflow колонки проверяли электрофоретически в денатурирующем полиакриламидном геле по методу Лэммли с дальнейшим окрашиванием Coomassie-250 (Рисунок 2).

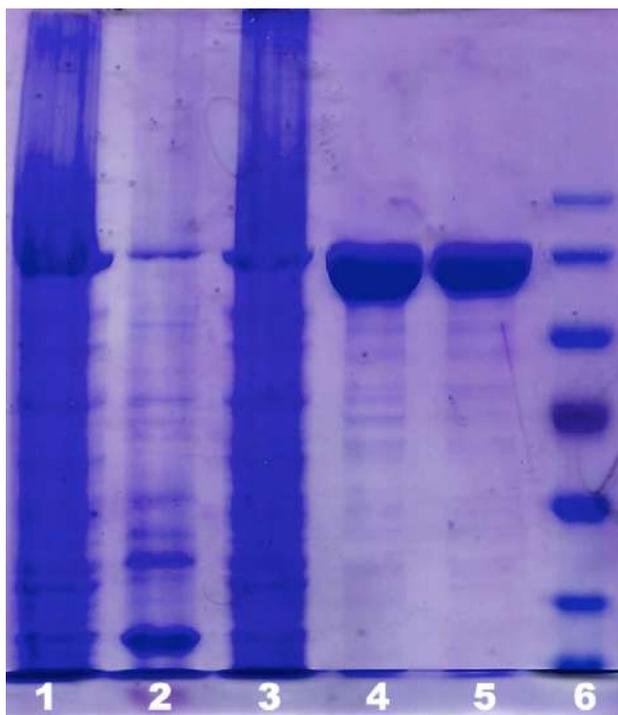


Рисунок 2 - Электрофорез в денатурирующем полиакриламидном геле фракций, полученных после металло-аффинной хроматографии. Дорожки слева направо: 1 — клеточный лизат, 2 — осадок, 3 — проток через колонку, 4 — фракция W с колонки, 5 — фракция E с колонки, 6 — маркер молекулярных масс PageRuler Plus (ThermoFisher Scientific)

Целевой белок оказался распределен во фракциях, элюировавшихся буферами W и E (Рисунок 3), при этом высокое оптическое поглощение фракции W при 260 нм указывало на содержание в ней примеси нуклеиновых кислот.

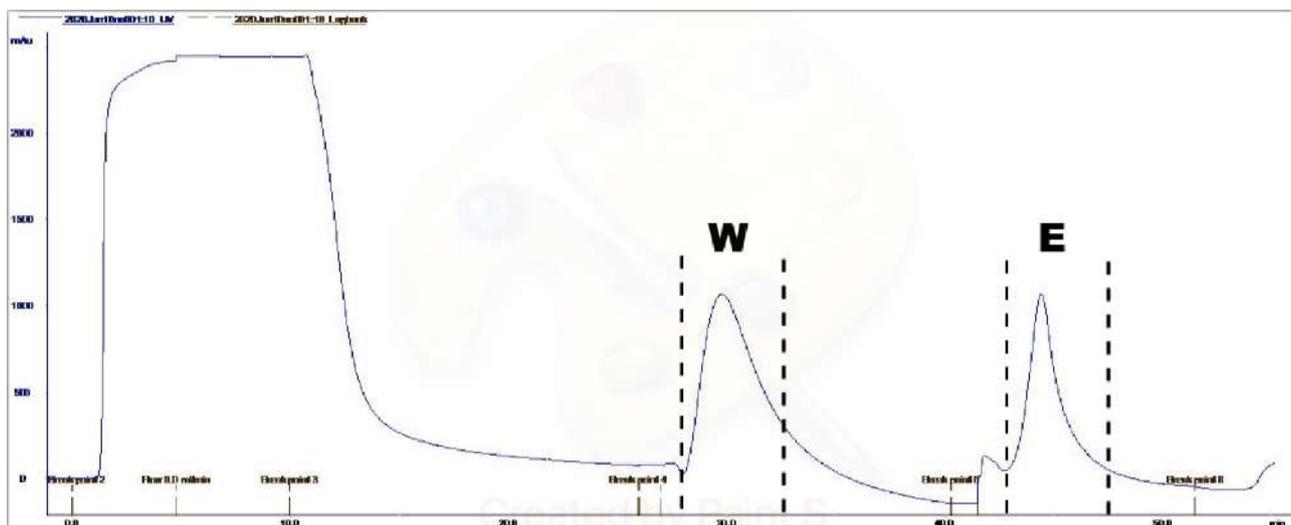


Рисунок 3 - Распределение целевого белка RbCas12a в промывочной W и элюирующей E фракциях. Широкий пик слева - проскок с колонки. По оси абсцисс показано время в минутах, по оси ординат - оптическая плотность в mAU.

Для удаления этой примеси, а также примесных белков, содержащихся в обеих фракциях, фракции объединяли и подвергали катионно-обменной хроматографии. Целевой белок элюировался в виде острого пика в диапазоне концентраций NaCl 170-220 мМ (Рисунок 4).

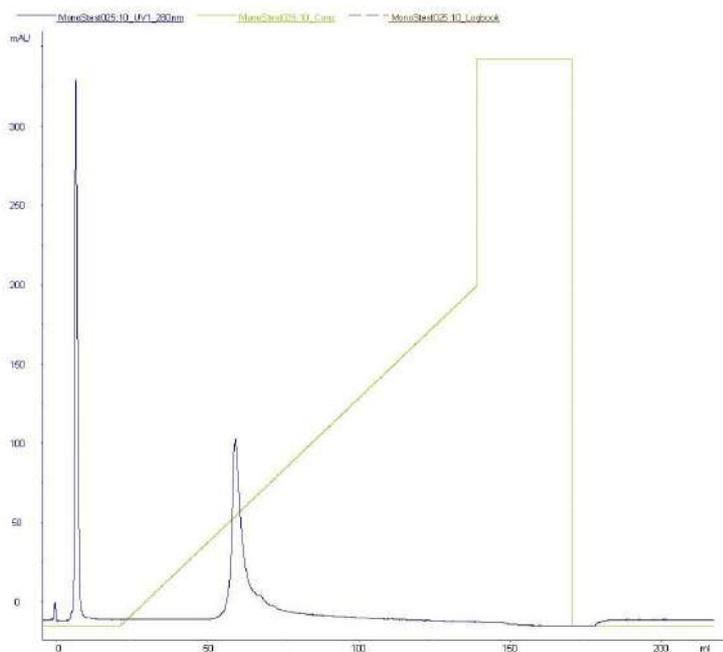


Рисунок 4 - Распределение элюата с целевым белком RbCas12a. По оси абсцисс указан объем в мл, по оси ординат - оптическая плотность в mAU.

Чистоту белка проверяли аналогично вышеописанному электрофоретически (Рисунок 5).

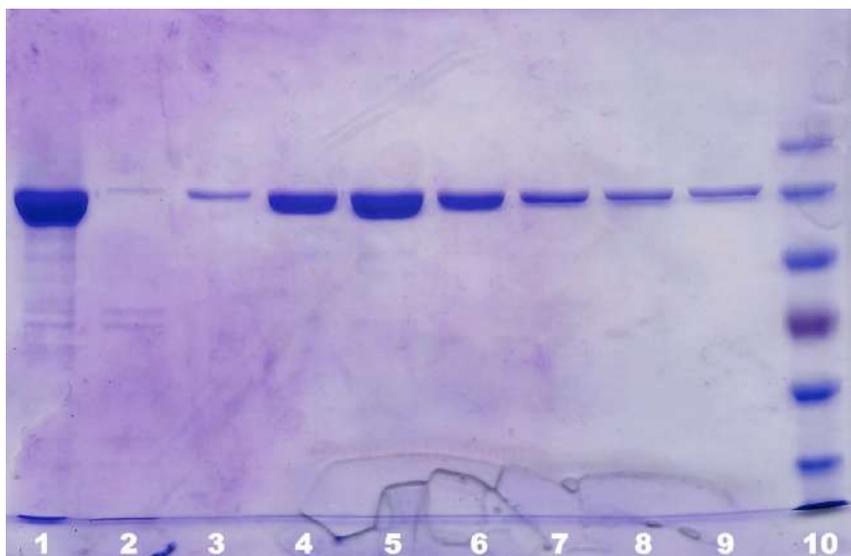


Рисунок 5 - Электрофорез в денатурирующем полиакриламидном геле фракций, полученных после катионно-обменной хроматографии. Дорожки слева направо:

1 — нанос на колонку, 2-9 — фракции, 10 — маркер молекулярных масс PageRuler Plus (ThermoFisher Scientific)

Для изучения степени олигомеризации полученного препарата, проводили гель-фильтрационную хроматографию. Выяснилось, что белок представляет собой мономер (Рисунок 6).

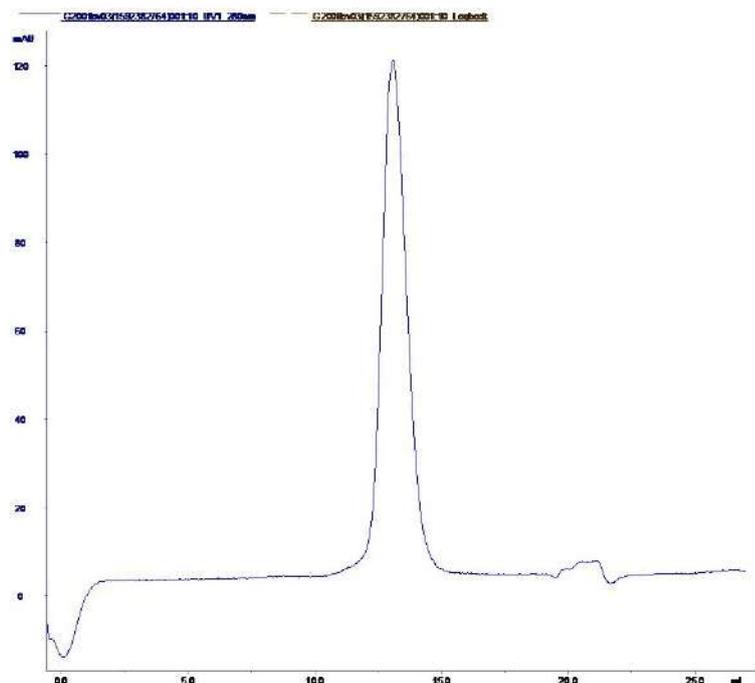


Рисунок 6 - Распределение целевого белка RbCas12a в ходе гель-фильтрации. По оси абсцисс указан объем в мл, по оси ординат - оптическая плотность в mAU.

3.3 Активность RbCas12a в различных буферах

Разрезание ДНК рекомбинантным RbCas12a *in vitro* проводили в ряде буферов: Nebbuffer1.1, Nebbuffer2.1, Nebbuffer3.1, Nebbuffer4, CutSmart (NEB), а также при добавлении в них 5 мМ дитиотрейтола (DTT) и буфера для разрезания (100 мМ HEPES pH 7.5; 500 мМ KCl; 25 мМ MgCl₂; 5 мМ DTT). Результаты представлены на электрофореграмме (Рисунок 7).

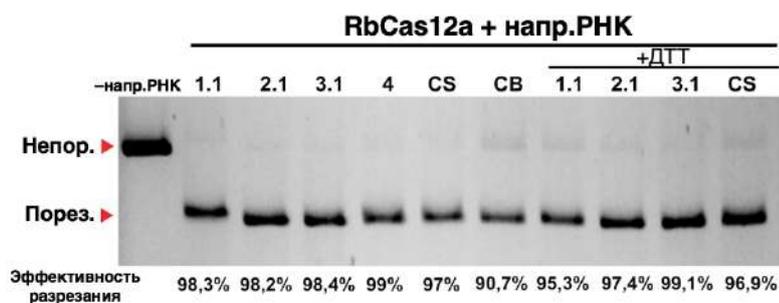


Рисунок 7 - Эффективности разрезания RbCas12a в различных буферах. Разрезания проводили с помощью RbCas12a и 500 нМ crРНК, при соотношении ДНК:RbCas12a:РНК 1:3:30 в различных буферах. Исследовали влияние NEBuffer 1.1 (1.1), NEBuffer 2.1 (2.1), NEBuffer 3.1 (3.1), NEBuffer 4 (4), CutSmart (CS) и буфера для разрезания (CB) с добавлением или без добавления 5 мМ дитиотрейтола (ДТТ). Снизу показаны средние значения эффективности разрезания по результатам трех независимых экспериментов. Непор., порез., напр.РНК — соответственно непорезанные и порезанные фрагменты ДНК и направляющая (crРНК).

Как видно из Рисунка 7, по результатам трех независимых экспериментов буфер для разрезания оказался наименее эффективным, в то время как сравнительные эффективности разрезания, определяемые с помощью программного обеспечения Bio-Rad ImageLab при подсчете интенсивностей полос на электрофореграммах, для коммерческих буферов оказались примерно равными вне зависимости от наличия в смеси ДТТ. Последнее, вероятнее всего, говорит о том, что белок способен формировать нативную структуру без дополнительного поддержания дисульфидных мостиков сульфгидрильными группами ДТТ, несмотря на сравнительно высокое содержание цистеиновых остатков в полипептидной цепи (12 цистеинов) RbCas12a при конкретных значениях рН (от 7 до 7.9), а также связывать в этих условиях направляющую РНК и осуществлять таргетный гидролиз двуцепочечной ДНК в конкретном соотношении компонентов реакции (1:3:30).

3.4 Определение температурного оптимума работы RbCas12a

Чтобы уточнить температурный диапазон активности RbCas12a, стабильность фермента в Nebuffer 2 измеряли с помощью термофлуоресцентного анализа. Этот метод с использованием флуоресцентного красителя позволяет определить точку плавления белка (T_m) в определенных условиях. Как показано на Рисунке 8, рассчитанная по результатам трех экспериментов T_m RbCas12a составляет $45,5^\circ\text{C}$. Далее проводились реакции разрезания *in vitro* при различных температурах, показавшие неспособность рибонуклеопротеинового комплекса к сборке и расщеплению дцДНК при температурах выше $45,5^\circ\text{C}$ из-за процесса денатурации RbCas12a и, возможно, структуры crRNA.

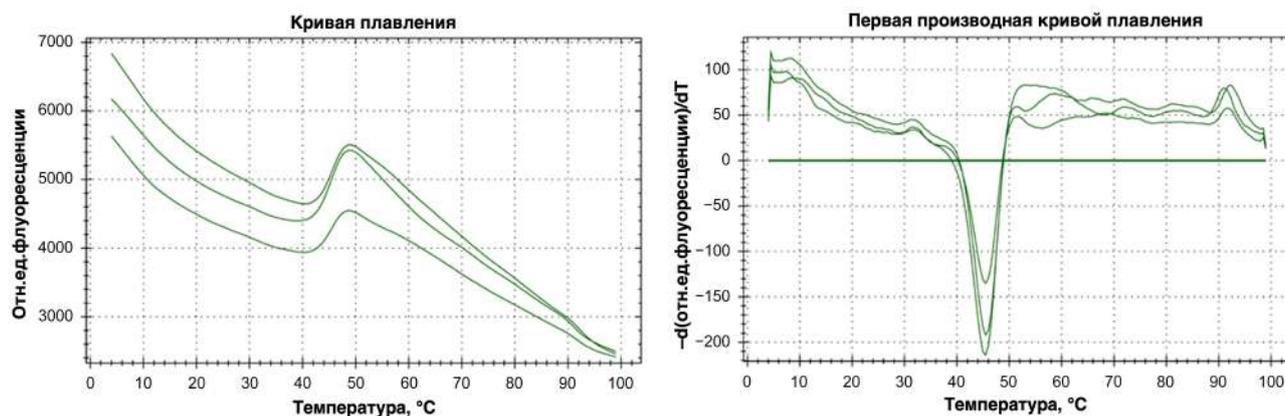


Рисунок 8 - Термофлуоресцентный анализ RbCas12a. Слева показаны кривые плавления, справа показана первая производная кривых плавления. Анализ проводили в трех независимых повторах. Температура плавления RbCas12a составляет $45,5^\circ\text{C}$. Отн.ед.флуоресценции — относительные единицы флуоресценции.

На начальных этапах определения температурного оптимума работы RbCas12a были получены первичные данные об аномально широком рабочем диапазоне фермента (от 4°C до 85°C), однако при более тщательном анализе далее выяснилось, что наблюдаемые ранее явления объясняются лишь крайне

высокой скоростью прохождения реакции разрезания *in vitro*, а не способностью фермента поддерживать нативную структуру с сохранением активности в данном диапазоне температур. Дело в том, что при смешивании компонентов реакция успевает пройти на комнатной температуре еще до инкубации смеси на соответствующей температуре уже менее, чем через 1 секунду. Влияние времени инкубации на эффективность разрезания RbCas12a сравнивали с гомологом из *Acidaminococcus sp. BV3L6* (AsCas12a). Как известно, AsCas12a также проявляет активность на комнатной температуре, что позволяет сравнить скорости направленного гидролиза матрицы для двух белков. Полученные данные позволяют сделать вывод о более эффективном разрезании дцДНК *in vitro* в реакциях с RbCas12a, что выгодно отличает его от гомолога AsCas12a (Рисунок 9).

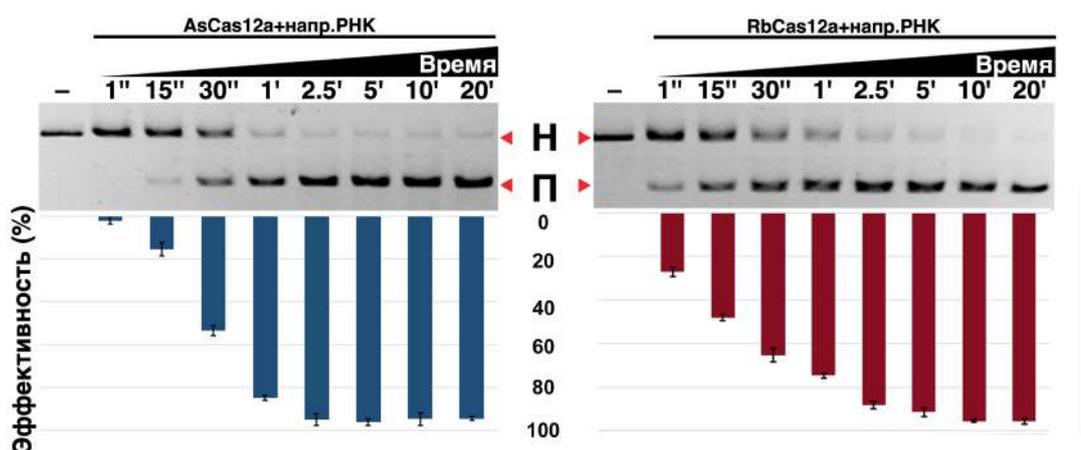


Рисунок 9 - Влияние времени инкубации на эффективность разрезания дцДНК комплексами RbCas12a/crРНК и AsCas12a/crРНК. Разрезание ДНК-мишени проводили рекомбинантными AsCas12a или RbCas12a в присутствии 500 нМ crРНК при соотношении ДНК:Cas12a:РНК 1:3:30. Показаны средние эффективности разрезания и стандартные отклонения, рассчитанные по трем независимым экспериментам. Н., П., напр.РНК — соответственно непорезанные, порезанные фрагменты ДНК и направляющая (crРНК). Дорожки с реакциями без crРНК обозначены как «←».

Из Рисунка 9 следует, что уже более 20% матрицы гидролизовано через 1 сек в случае RbCas12a, и менее 1% в случае AsCas12a при эквимолярной концентрации компонентов реакции. Через 15 сек более половины всей матрицы разрезано в случае RbCas12a, и менее 20% в случае AsCas12a. Таким образом, было обнаружено, что определение температурного диапазона работы RbCas12a стандартным способом не представляется возможным. Поэтому было предложено предварительно нагревать по отдельности все компоненты смеси до соответствующих температур, преинкубировать рибонуклеопротеиновый комплекс при данной температуре, а затем добавлять в него ДНК, также предварительно нагретую до соответствующей температуры. Результаты представлены на Рисунке 10.

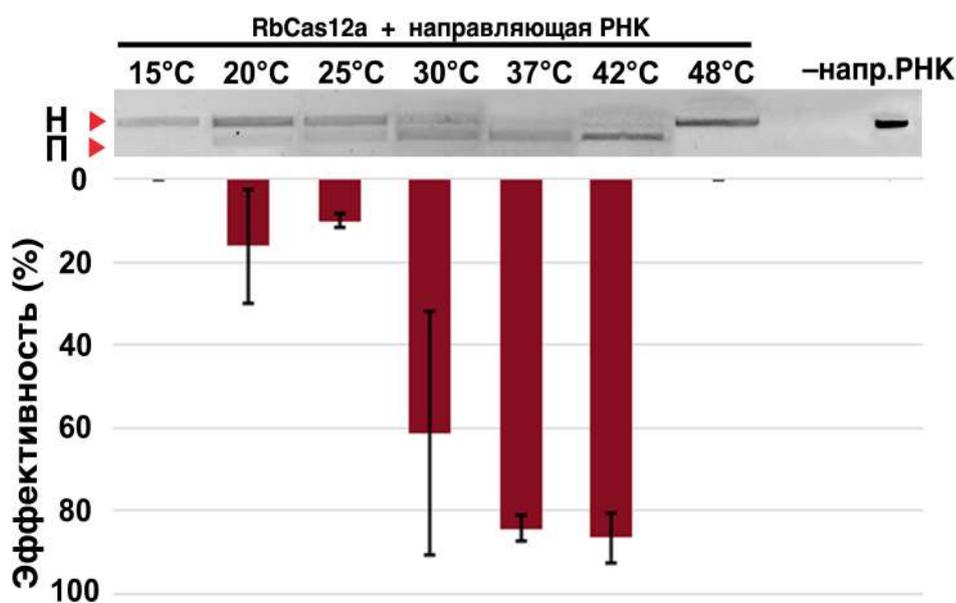


Рисунок 10 - Влияние температуры на эффективность разрезания *in vitro* RbCas12a. Разрезание дцДНК-мишени с помощью RbCas12a, в присутствии 500 нМ crРНК при соотношении ДНК:RbCas12a:РНК 1:3:30. Показаны средние эффективности разрезания и стандартные отклонения, рассчитанные по трем независимым экспериментам. Н., П., напр.РНК — соответственно непорезанные фрагменты, порезанные фрагменты и направляющая (crРНК).

Из Рисунка 10 следует, что наибольшая активность наблюдается при 42 °С, но она не сильно отличается от таковой при 37 °С. При 48 °С активность комплекса не детектируется, что соотносится с полученными данными о точке плавления белка, равной 45,5 °С, полученными в предыдущих экспериментах. При 30 °С активность комплекса снижается в некоторых репликах, в результате по полученным данным наблюдается 61.3 ± 29.2 стандартных отклонений процента разрезания. Интересно, что небольшая активность комплекса RbCas12a/crРНК сохраняется и при 20 °С, однако не детектируется на 15 °С. Вероятнее всего, при 20 °С комплекс становится нестабильным, что отражает большой разброс эффективностей в различных репликах эксперимента — от 5.4 % до 35.5 % разрезания. Крайне важным является тот факт, что комплекс RbCas12a/crРНК сохраняет относительно высокую активность при температурах ниже 37 °С, что говорит о потенциальной возможности использования данного фермента в экспериментах по геномному редактированию видов, чувствительных к температурам. В частности, известно, что для эффективного редактирования растений нужны температуры ниже 37 °С. Таким образом, фермент RbCas12a является потенциальным редактором сельскохозяйственных культур растений. Однако стоит иметь в виду, что текущие данные иллюстрируют лишь результаты экспериментов в конкретных условиях: при использовании конкретной направляющей РНК, конкретной последовательности РАМ и конкретного соотношения ДНК:РНК:белок при отсутствии других компонентов клетки в системе *in vitro*. Хотя данные кажутся обнадеживающими, для более детального изучения этого вопроса требуются дополнительные эксперименты.

3.5 Определение кофактора RbCas12a

Поскольку разные нуклеазы подтипа V-A обладают различными предпочтениями к кофакторам [106], было решено определить, какие катионы могут связываться с RbCas12a для координации гидролиза дцДНК. Катионы

Ca²⁺, Ni²⁺, Mn²⁺, Mg²⁺, Zn²⁺, Cu²⁺ и Co²⁺ тестировали как возможные кофакторы RbCas12a в концентрации 10 мМ при 37 °С. Результаты представлены на Рисунке 11.

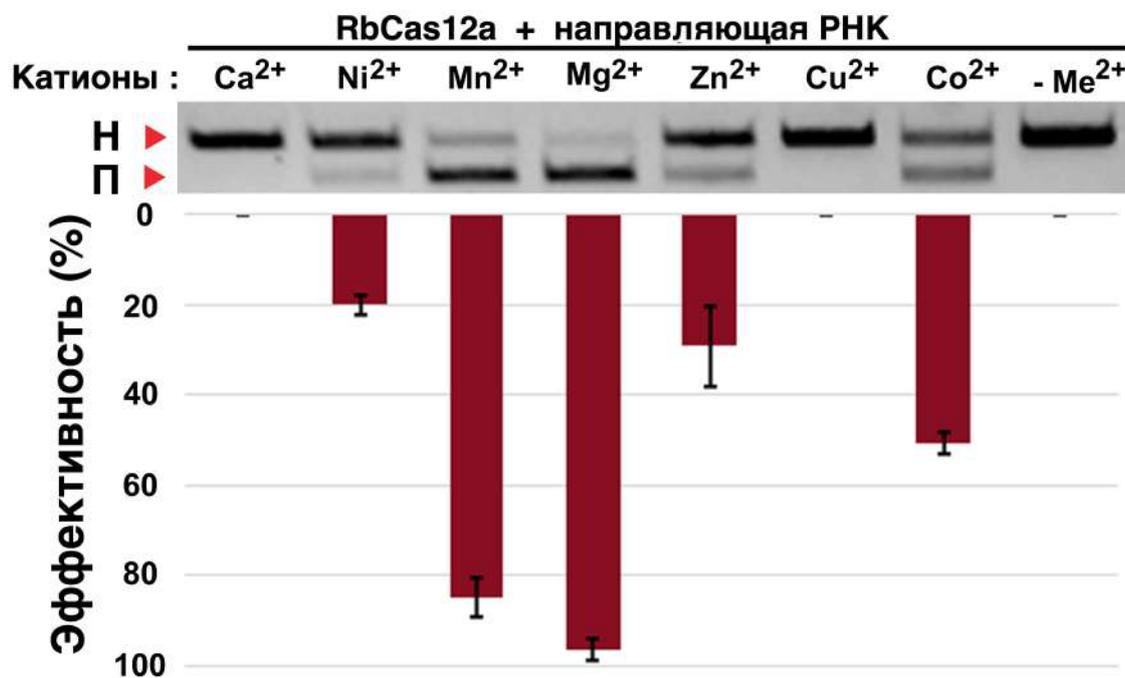


Рисунок 11 - Эффект двухвалентных катионов на разрезание RbCas12a ДНК матрицы *in vitro*. Показан гель и средние значения относительной эффективности разрезания по результатам трех экспериментов в %, черными линиями показано \pm стандартное отклонение. Последняя дорожка (-Me²⁺) - разрезание в том же буфере, но при полном отсутствии двухвалентных катионов. Н., П. — соответственно непорезанные и порезанные фрагменты ДНК.

Как видно из Рисунка 11, Mn²⁺ и Mg²⁺ оказывают схожий эффект на активность RbCas12a, лишь с небольшим преимуществом Mg²⁺, в то время как присутствие вместо них ионов Ni²⁺ или Zn²⁺ практически ингибирует реакции. Хотя RbCas12a и сохраняет активность в присутствии ионов кобальта, только половина молекул ДНК-матрицы гидролизовалась за это время в трех повторных сериях эксперимента. Ионы Ca²⁺ и Cu²⁺ полностью ингибировали реакции разрезания при отсутствии других катионов. Также, как и ожидалось,

таргетной цепи ДНК. Б) Общая схема эксперимента, используемого для идентификации последовательности РАМ RbCas12a. Плазмидную РАМ-библиотеку, содержащую 8 вырожденных нуклеотидов на месте предполагаемого РАМ, гидролизовали в реакциях разрезания *in vitro* с помощью комплекса RbCas12a/crРНК и ряда рестриктаз (сайты рестрикции не показаны). Далее целевые фрагменты разделялись в геле электрофоретически и очищались. Из очищенных гидролизованных и негидролизованных комплексом RbCas12a/crРНК фрагментов ДНК готовили библиотеки для высокопроизводительного секвенирования и секвенировали. В) WebLogo график полученной при анализе данных высокопроизводительного секвенирования консенсусной РАМ-последовательности RbCas12a - 5'-YYN-3'. Нумерация нуклеотидов по оси абсцисс соответствует 5'-3' направлению, начиная с первого вырожденного нуклеотида по нетаргетной цепи РАМ-библиотеки (он же нуклеотид -8). Напр.РНК — направляющая (crРНК).

Как видно из лого-графика (Рисунок 12В), для RbCas12a последовательность РАМ была определена как трехнуклеотидный мотив 5'-YYN-3', где Y - пиримидиновый нуклеотид (либо С, либо Т), а N - любой нуклеотид.

Для подтверждения полученных данных, представленных на лого-графике, была проведена серия реакций *in vitro* разрезания ДНК-матриц с различными РАМ-последовательностями из множества 5'-YYN-3'. Поскольку наиболее распространенным консенсусом РАМ нуклеаз подтипа V-A является 5'-TTTN-3' или в некоторых случаях 5'-TTN-3' [107], в первую очередь было решено проверить, влияет ли последний 3'-концевой нуклеотид РАМ на эффективность протекания *in vitro* разрезов матриц с одним и тем же протоспейсером, но разным РАМ (5'-TTN-3' в случае RbCas12a). Как видно из Рисунка 13А, существенных различий по результатам трех независимых экспериментов между 5'-TTC-3', 5'-TTG-3', 5'-TTA-3', 5'-TTT-3' РАМ-последовательностями не обнаружено. Это доказывает, что положение -1 в РАМ

не принципиально для эффективного гидролиза RbCas12a, в то время как наличие тимидина в этом положении РАМ негативно сказывается на работе некоторых других ортологов, например, нуклеаз AsCas12a и LbCas12a, предпочитающих РАМ вида 5'-ТТТV-3' (где V представляет собой А, С или G) и некоторые другие субоптимальные РАМ [108].

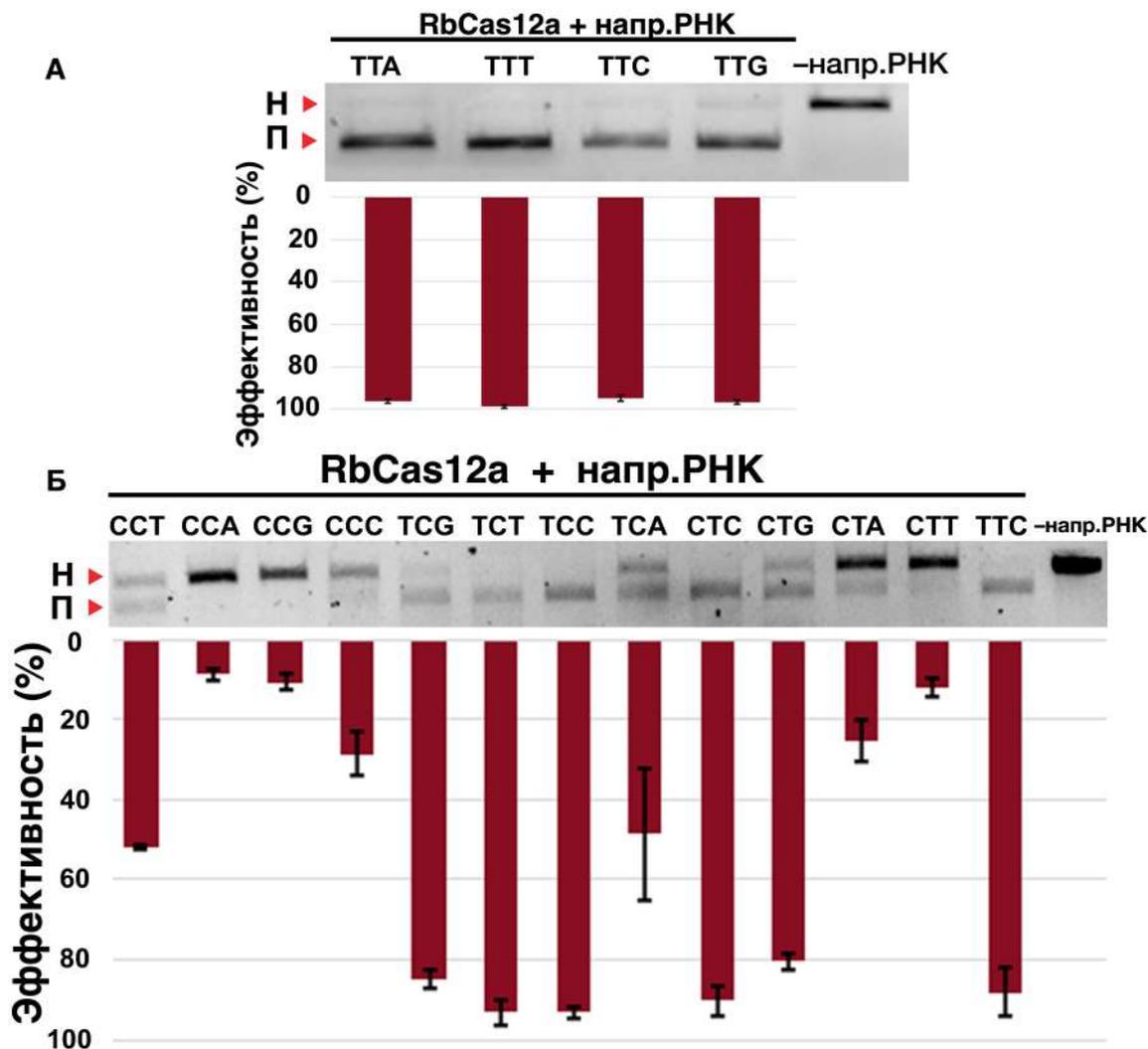


Рисунок 13 - Изучение РАМ-специфичности нуклеазы RbCas12a в условиях *in vitro*. На верхней части (А) показана электрофореграмма разрезов матрицы с одним и тем же протоспейсером, но разными нуклеотидами в последней 3'-концевой позиции РАМ (позиция -1). В качестве контроля использовалась матрица ТТG без добавления сгРНК (-напр.РНК). На нижней части (Б) показана электрофореграмма разрезов рUC119 с использованием различных

направляющих РНК на протоспейсеры с различными последовательностями РАМ вида 5'-YYN-3', на последней дорожке контрольная матрица без добавления направляющей РНК. Под каждой дорожкой на обеих частях рисунка столбики, иллюстрирующие средние значения относительных эффективностей разрезания в % по трем повторам \pm стандартное отклонение. Н., П., напр.РНК — соответственно непорезанные, порезанные фрагменты ДНК и направляющая (crРНК).

Чтобы изучить влияние нуклеотидов в положениях -2 и -3 на эффективность разрезания, мы протестировали активность комплекса RbCas12/crРНК со всеми оставшимися возможными комбинациями 5'-YYN-3' (кроме уже протестированных 5'-TTD-3', где D представляет собой Т, А или G). Результаты показали, что эффективность RbCas12a-опосредованного разрезания ДНК с РАМ вида 5'-TCB-3' и 5'-CTS-3' составила ~80 % (где B представляет собой С, Т или G, а S представляет собой G или С), что соизмеримо с эффективностями разрезаний матриц с РАМ вида 5'-TTN-3' (Рисунок 13Б). Гораздо менее эффективно гидролиз проходил в случае РАМ вида 5'-CCT-3' и 5'-TCA-3' (эффективность ~ 27%), тогда как количество продуктов разрезания матриц с РАМ вида 5'-CCA-3', 5'-CCG-3', 5'-CTT-3', составляли только ~ 10 % от количества интактных фрагментов.

В совокупности эти результаты ясно демонстрируют, что RbCas12a способен эффективно распознавать РАМ, определяемый как 5'-YYN-3', для осуществления направленного гидролиза дцДНК-матриц, по крайней мере, *in vitro*. Следует отметить, что в случае 5'-CCV-3' и 5'-CTW-3' (где V представляет собой С, А или G, а W представляет собой А или Т) РАМ, RbCas12a не способен разрезать ДНК-матрицы с эффективностью более 25% в условиях описанного эксперимента.

3.7 RbCas12a проявляет коллатеральную активность

Способность осуществлять неспецифический гидролиз оцДНК после направленного разрезания дцДНК-матрицы, известная как коллатеральная активность, свойственна всем известным ортологам Cas12a. Чтобы показать, что RbCas12a также проявляет эту активность, проводили реакции разрезания *in vitro* в присутствии ДНК-зонда, меченного флуорофором и гасителем на его 5'- и 3'-концах, соответственно. Результаты флуоресцентной детекции неспецифического гидролиза зонда в присутствии целевой матрицы и при ее отсутствии по итогам трех независимых экспериментов представлены на Рисунке 14.

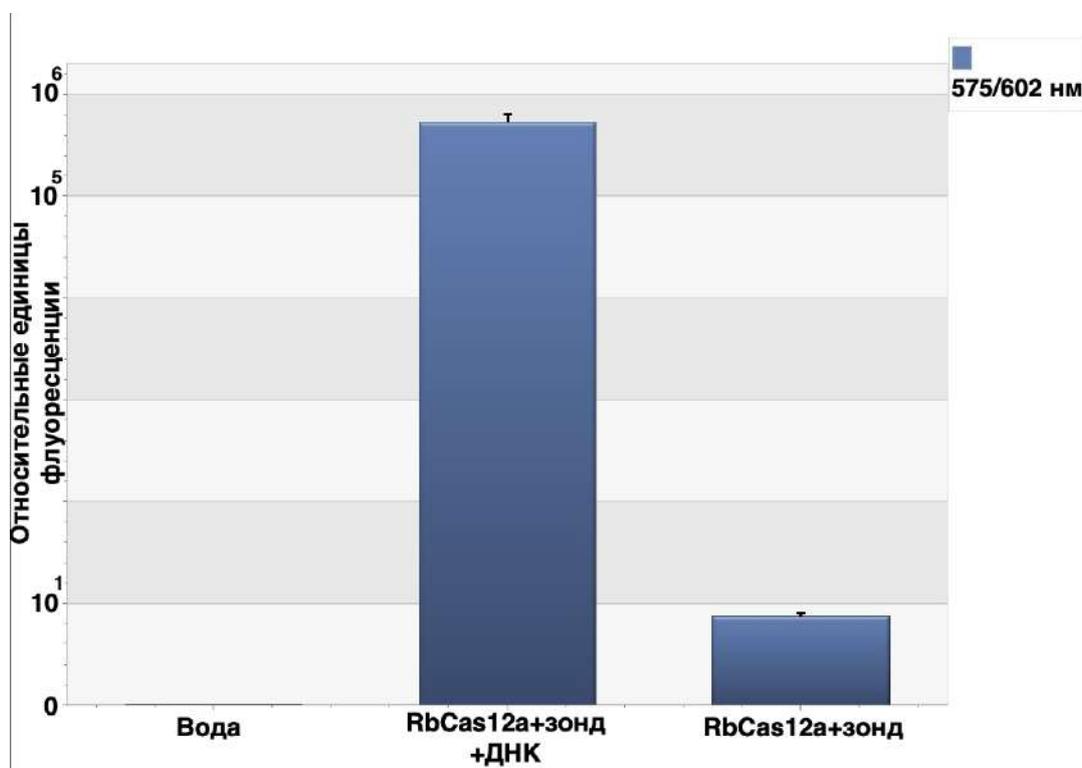


Рисунок 14 - Коллатеральная активность RbCas12a. Данные показывают среднее значение по результатам трех независимых экспериментов. Левый столбик - вода, средний - реакция разрезания *in vitro* RbCas12a целевой ДНК в присутствии crРНК и зонда с ROX. Правый столбик - контроль реакции разрезания *in vitro* RbCas12a без ДНК матрицы, но в присутствии crРНК и зонда с ROX.

Заметно значительное увеличение уровня флуоресценции в реакциях с целевой ДНК-матрицей по сравнению с отрицательным контролем без матрицы, что говорит о способности RbCas12a нуклеазы проявлять коллатеральную активность по отношению к одноцепочечной ДНК. Следовательно, RbCas12a может быть в будущем использован в отечественных разработках диагностических тест-систем на основе CRISPR/Cas.

3.8 RbCas12a формирует выступающие 5'-концы

Как указывалось в Главе 1, все известные ортологи Cas12a формируют при гидролизе выступающие 5'-концы дцДНК, однако длины этих концов варьируют для разных белков данного семейства. Чтобы определить природу и длины формируемых RbCas12a концов, продукты гидролиза *in vitro* подвергали «затуплению», клонировали в составе плазмидного вектора и секвенировали по Сэнгеру. Результаты секвенирования, проанализированные в ПО SeqA6 и SnapGene, представлены на Рисунке 15.

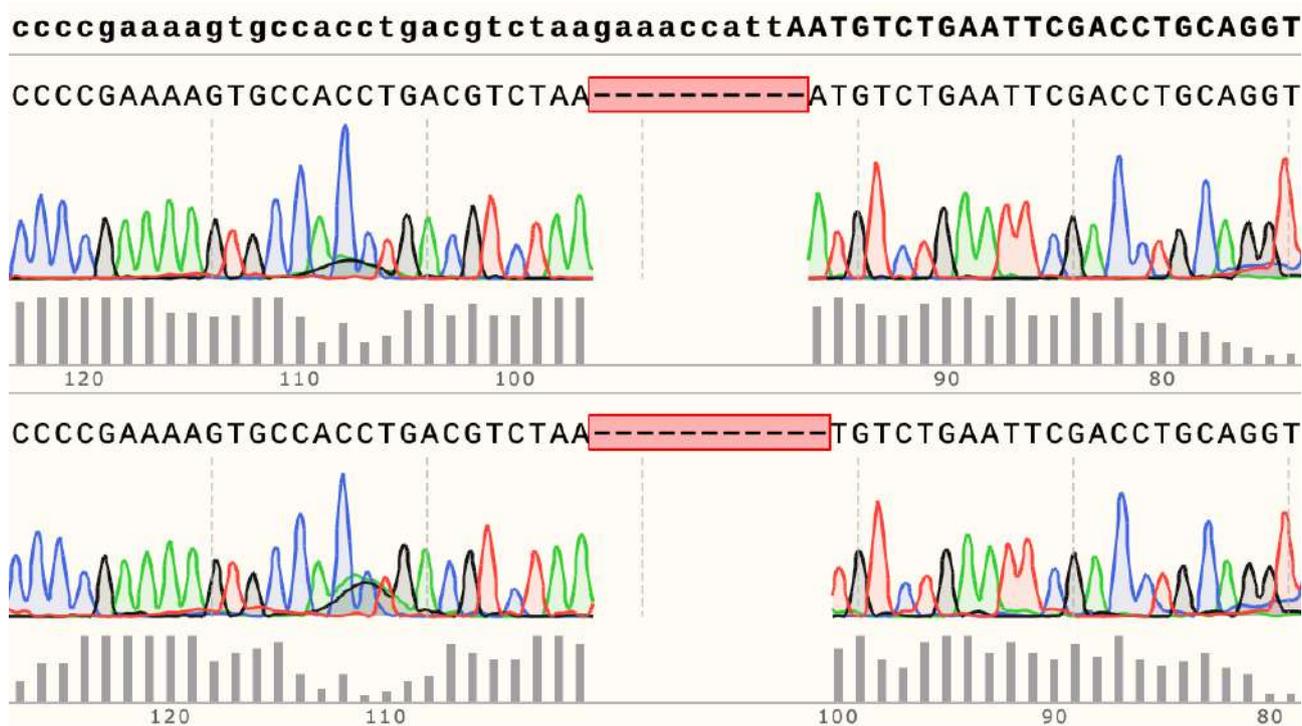


Рисунок 15 - Фрагменты хроматограмм, полученных после лигирования гидролизованного нуклеазами Cas12a фрагмента с обработанными ферментативно выступающими концами в pKan-T вектор. Верхняя панель - после гидролиза FnCas12a, нижняя - после гидролиза RbCas12a. Видно, что выступающий 5'-конец длиннее на 1 нуклеотид в случае RbCas12a.

Среди семейства Cas12a наиболее подробно природа выступающих концов продуктов гидролиза изучена для ортолога из *Francisella novicida* (FnCas12a). Для сравнения с RbCas12a длины концов в реакциях разрезания *in vitro* использовали коммерческий FnCas12a (Applied Biological Materials, США). Как видно из приведенных хроматограмм, выступающий 5'-конец ДНК в случае гидролиза RbCas12a нуклеазой оказался длиннее на 1 нуклеотид, чем тот же 5'-конец, образовавшийся после гидролиза FnCas12a. Стоит отметить, что длина выступающих концов несколько варьирует от клона к клону, что объясняется нативной природой их образования — триммингом целевой цепи, подробно описанной в Главе 1.2.5. Таким образом, приведенные хроматограммы отражают лишь наиболее часто встретившиеся варианты. Более подробную статистику удалось набрать при анализе данных NGS после редактирования генома клеток млекопитающих, о чем будет сказано ниже.

3.9 Сплит направляющей РНК

В ходе работ было изучено влияние укорочения crРНК и расщепление ее на две молекулы на *in vitro* разрезание ДНК матрицы различными нуклеазами Cas12a. Для выяснения, способен ли RbCas12a осуществлять направленный гидролиз ДНК-матрицы, связывая сплит-версию crРНК или одну только спейсерную часть crРНК, проводили разрезания *in vitro* в присутствии указанных компонентов реакций. Как показано на Рисунке 16А, в реакциях, содержащих как полноразмерные, так и сплит crРНК, в условиях эксперимента целевая ДНК гидролизовалась на ~ 90 %. Данные, представленные на Рисунке

16Б, показывают, что увеличение концентрации спейсерной РНК от 0.5 до 5 мкМ вызывает дозозависимое расщепление ДНК-субстрата с помощью RbCas12a. Достаточно эффективное разрезание ДНК-матрицы (~ 50 % после 20-минутной инкубации) удалось наблюдать при соотношении RbCas12a:crРНК 1:500.

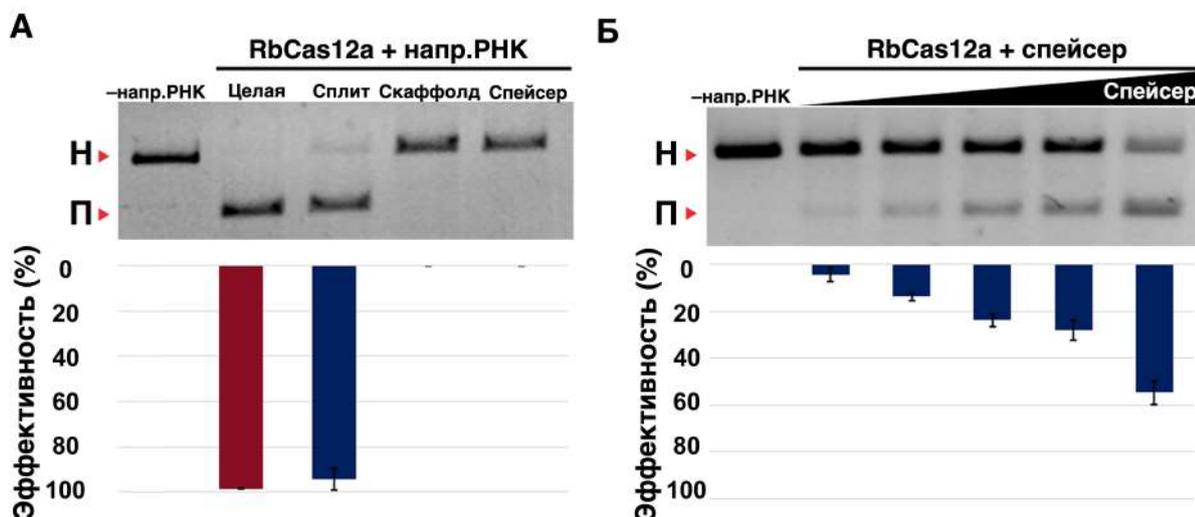


Рисунок 16 - Влияние сплит crРНК и отдельно спейсера crРНК на активность разрезания RbCas12a дцДНК матрицы. А) Разрезание ДНК-мишени с помощью RbCas12a, заряженного 0,3 мкМ полноразмерных (целых), сплит или отдельных скаффолдных и спейсерных частей crRNA при соотношении ДНК:RbCas12a:crРНК 1:3:30. Средняя эффективность разрезания по трем независимым экспериментам показана под гелем. Б) Разрезание дцДНК матрицы с помощью RbCas12a, заряженного с увеличением концентрации (0.5, 1, 2, 2.5 и 5 мкМ спейсерной РНК, что соответствует соотношениям 1:5:50, 1:5:100, 1:5:200, 1:5:250 и 1:5:500). Показаны средние эффективности разрезания и стандартные отклонения, рассчитанные по трем независимым экспериментам. Н., П., напр.РНК — соответственно непорезанные, порезанные фрагменты ДНК и направляющая (crРНК).

Для подтверждения протоспейсер-независимого характера разрезания ДНК со сплит crRNA проводили разрезания *in vitro* на трех различных таргетах дцДНК. В качестве контроля использовали полноразмерные crRNA (Рисунок 17).

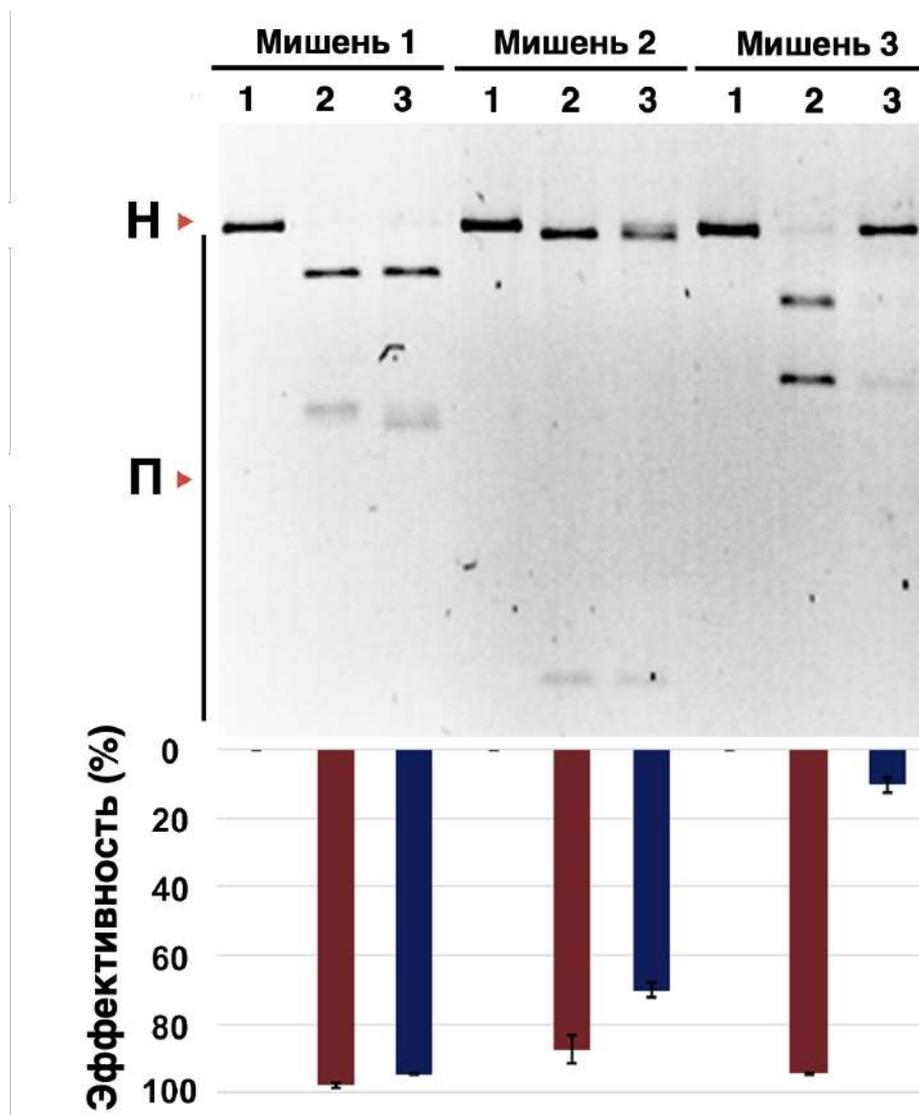


Рисунок 17 - Эффект сплит-версии направляющей РНК на *in vitro* разрезание с помощью RbCas12a. Верх: электрофореграмма продуктов разрезания с использованием трех различных crRNA, их полноразмерной и сплит-версии. В качестве отрицательного контроля проводили реакции без добавления crRNA. Низ: относительные эффективности разрезания в % и стандартные отклонения по трем независимым экспериментам. Н., П.— соответственно непорезанные,

порезанные фрагменты ДНК. 1, 2, 3 – соответственно разрезания без сгРНК, с использованием полноразмерной сгРНК и сплит-версии сгРНК.

Далее проверяли, способен ли другой белок из семейства Cas12a осуществлять гидролиз матрицы, при укорочении длины сгРНК. Выяснилось, что нуклеазная активность ортолога AsCas12a сохраняется после значительного уменьшения длины скаффолдной части направляющей РНК и может быть обнаружена даже при полном удалении скаффолда, то есть в присутствии 20-нуклеотидной РНК, соответствующей только спейсерной части. Кроме того, остаточная нуклеазная активность AsCas12a, заряженного спейсерной РНК, может быть доведена до уровней, типичных для эффекторов, заряженных полноразмерной направляющей РНК, путем добавления *in trans* скаффолдной части РНК из 20 оснований (Рисунок 18).

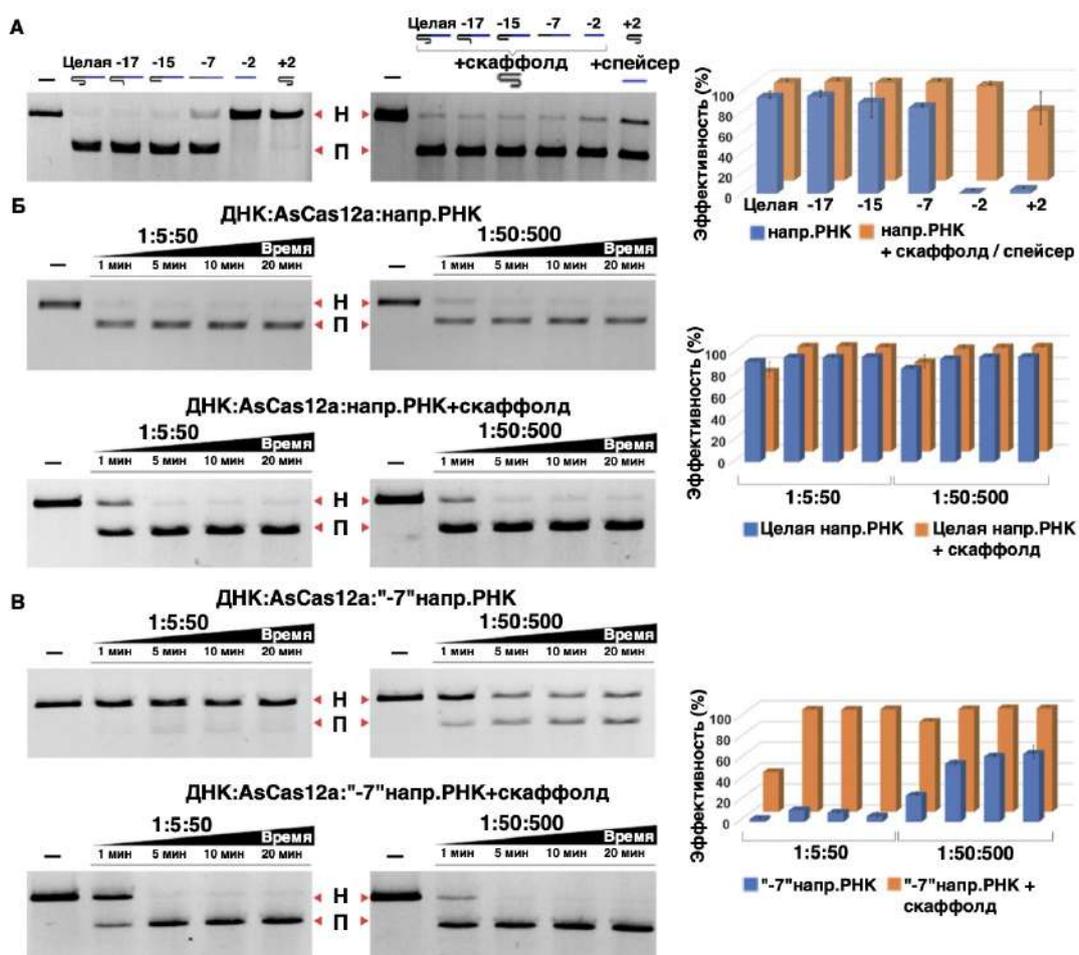


Рисунок 18 - Результаты анализа разрезания ДНК *in vitro* с помощью AsCas12a, запрограммированного укороченными crРНК (указаны над гелем). А) Реакции содержали 10 нМ ДНК, 500 нМ AsCas12a и 5 мкМ указанных crРНК (соотношение 1:50:500). Показаны только неразрезанная ДНК и один из продуктов разрезания (2776 п.н.). На правом геле в реакции добавляли 5 мкМ скаффолдной или спейсерной РНК. Справа показаны средние эффективности разрезания по четырем независимым экспериментам и стандартные отклонения. (Б-В) Кинетика и концентрационная зависимость разрезания ДНК с помощью AsCas12a, заряженного полноразмерными (целыми) (Б) или «-7» (В) crРНК. На гелях отражены результаты анализов разрезов, проведенных в присутствии эквимоллярных количеств скаффолдной части crРНК. Соотношения ДНК:AsCas12a:crРНК подписаны над каждым гелем. Средние значения эффективности разрезания и стандартные отклонения по трем независимым экспериментам показаны справа. Н., П., напр.РНК — соответственно непорезанные, порезанные фрагменты ДНК и направляющая (crРНК). Дорожки с реакциями без добавления crРНК в качестве контроля обозначены как «-».

Чтобы сравнить связывание полноразмерной и сплит направляющей РНК с AsCas12a, был проведен анализ разрезания ДНК *in vitro* в присутствии возрастающих количеств другой направляющей РНК (далее именуемой «конкурентная РНК»), которая нацелена на последовательность, отсутствующую в матрице ДНК (Рисунок 19).

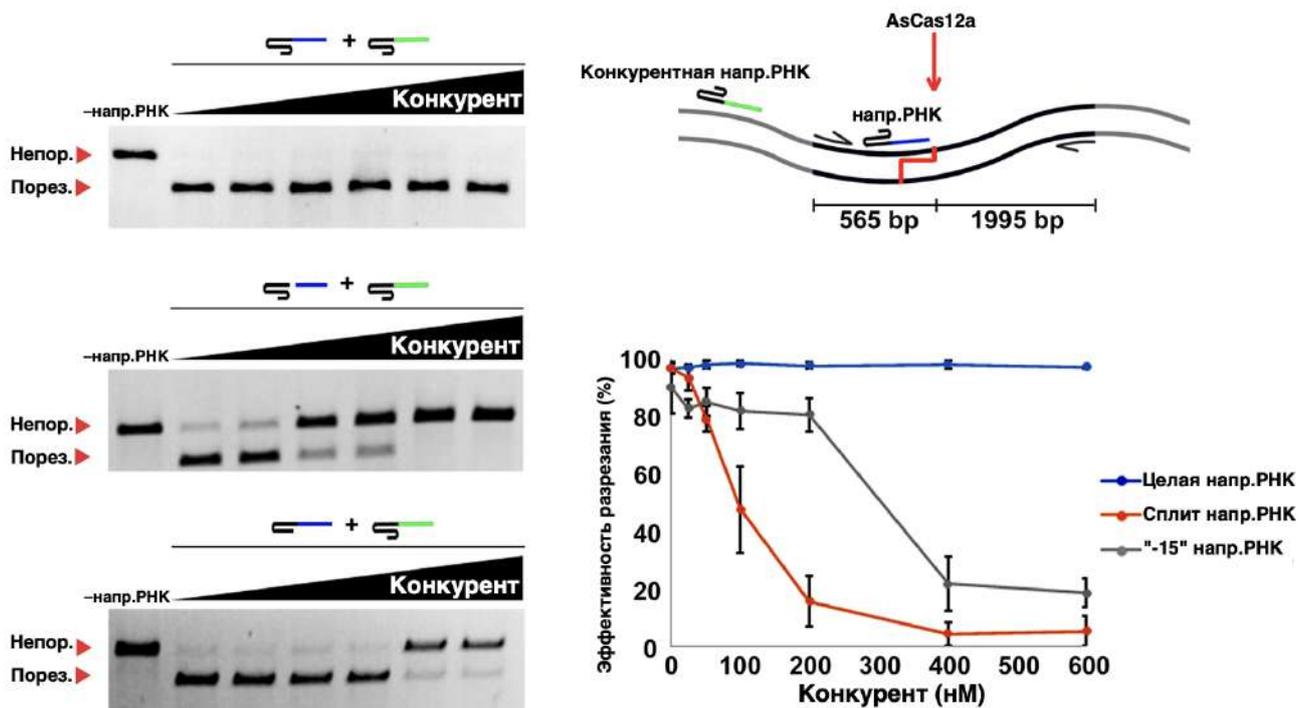


Рисунок 19 - Влияние сплит-направляющей РНК на активность разрезания AsCas12a в присутствии конкурентной направляющей РНК. Вверху справа: схема анализа таргетного разрезания в присутствии конкурентной РНК. Спейсерные фрагменты целевой и конкурентной РНК показаны синим и зеленым соответственно. Ниже приведены результаты анализов разрезания с помощью AsCas12a, связанного с 500 нМ полноразмерной (целой), сплит или «-15» направляющей РНК (соотношение 1:5:50 ДНК:AsCas12a:РНК) в присутствии возрастающих концентраций (0, 25, 50, 100, 200, 400, 600 нМ) конкурентной РНК. Справа показаны средние значения эффективности разрезания и стандартные отклонения, рассчитанные по трем независимым экспериментам. Непор., порез., напр.РНК — соответственно непорезанные, порезанные фрагменты ДНК и направляющая (crРНК).

Как видно из Рисунка 19, добавление до 600 нМ конкурента не имело эффекта на расщепление 10 нМ целевой ДНК эффектором, заряженным полноразмерной направляющей РНК, присутствующей в концентрации 500 нМ. Напротив, разрезание эффектором, заряженным усеченной -15 направляющей РНК,

ингибировалось на 50 % в присутствии 300 нМ конкурирующей РНК, что предполагает 2-кратное снижение аффинности к AsCas12a по сравнению с полноразмерной направляющей РНК. Разрезание мишени эффектором, заряженным сплит-версией РНК, ингибировалось на 50 % в присутствии 100 нМ конкурента, что указывает на то, что сродство сплит РНК к эффектору снижается в ~ 5 раз по сравнению с полноразмерной РНК.

Далее анализировалась способность AsCas12a осуществлять направленный гидролиз как одноцепочечных, так и двуцепочечных ДНК при формировании комплекса со сплит-версией crРНК по сравнению с полноразмерной crРНК. Как видно из Рисунка 20, разрезание ДНК PhiX174 было индуцировано распознаванием двуцепочечных и одноцепочечных мишеней нуклеазой AsCas12a, заряженной контрольной полноразмерной направляющей РНК (дорожки 4 и 6, соответственно) или с помощью сплит-версии направляющей РНК (дорожки 8 и 10). Расщепление PhiX174 не наблюдалось в реакциях без целевой ДНК (дорожки 3 и 7) или в присутствии нецелевой ДНК (дорожки 5 и 9). AsCas12a, заряженный либо полноразмерной, либо сплит-версией направляющей РНК, разрезает двуцепочечные и одноцепочечные таргеты в одних и тех же положениях (дорожки 2 и 4 и дорожки 9 и 11, соответственно). Помимо полос, возникающих в результате сайт-специфического расщепления одноцепочечной ДНК, продукты деградации в результате неспецифической активности наблюдались как с полноразмерными направляющими РНК, так и со сплит-заряженными эффекторами. При активации немеченым таргетным субстратом (активатором), AsCas12a, заряженный полноразмерной или сплит РНК, полностью разрушает меченую одноцепочечную ДНК (дорожки 8 и 15), оставляя двуцепочечную ДНК нетронутой (полосы 7 и 14).

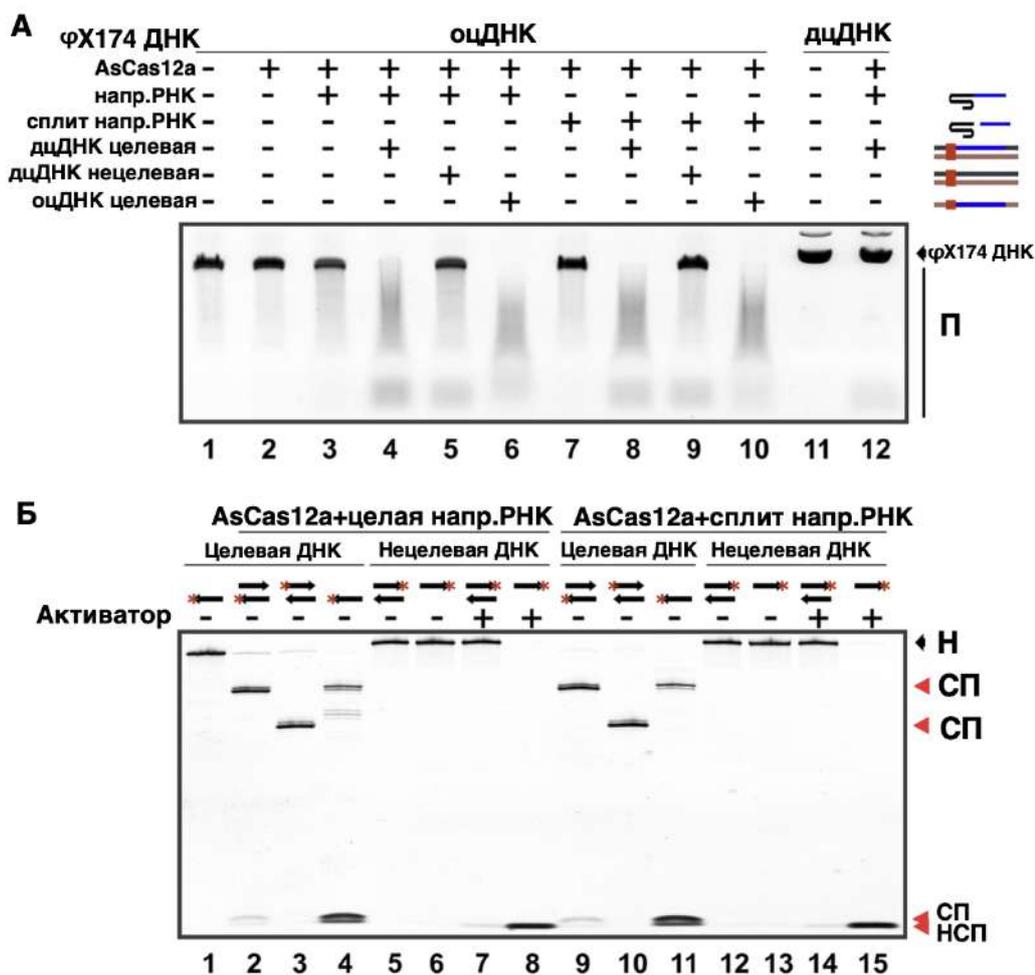


Рисунок 20 - Все известные активности разрезания ДНК с помощью AsCas12a поддерживаются сплит-версией РНК. А) Электрофореграмма *in vitro* разрезов оц- и дцДНК в присутствии и отсутствии различных компонентов реакции с AsCas12a. (Б) Меченные флуоресцеином (красная звездочка) целевые и нецелевые ДНК-субстраты в одноцепочечной или двухцепочечной форме (схематически показаны вверху геля) в реакциях с AsCas12a, заряженного с помощью полноразмерной (целой) или сплит-версии crРНК. «Активатор» представляет собой немеченый одноцепочечный олигонуклеотид, содержащий комплементарную спейсеру РНК и несущий функциональный РАМ. Продукты реакции разделяли в денатурирующем 10 % ПААГ. Указаны непорезанные (Н), порезанные (П), порезанные специфически (СП) и неспецифически (НСП) фрагменты ДНК, напр.РНК — направляющая (crРНК).

Для того, чтобы узнать, является ли способность связывать сплит-версию crРНК общим свойством нуклеаз V-A подтипа, проводили разрезы *in vitro* с использованием FnCas12a и LbCas12a (Рисунок 21).

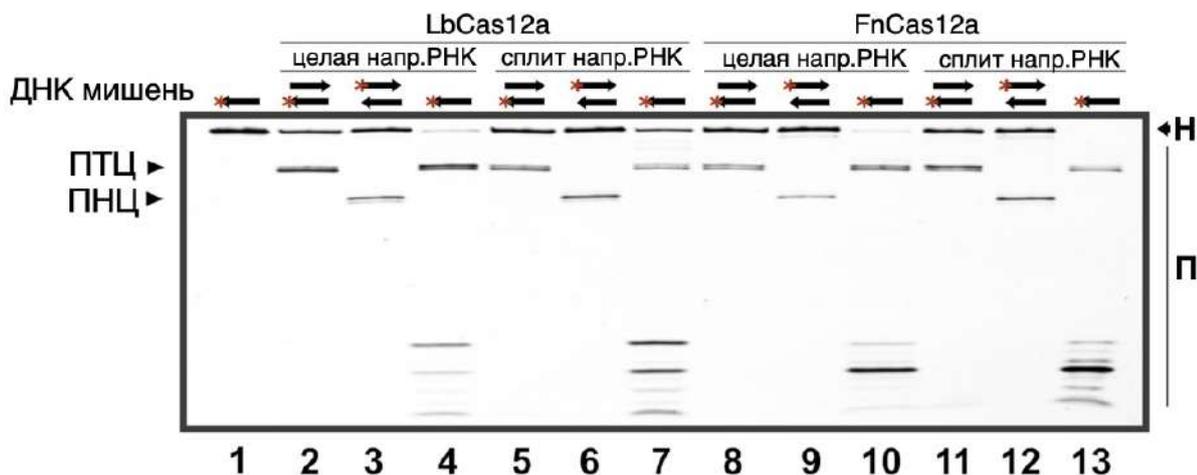


Рисунок 21 - Эффект сплит-версии направляющей РНК на *in vitro* разрезание с помощью LbCas12a и FnCas12a. Красной звездочкой обозначена цепь ДНК, меченная флуоресцеином. Непорезанные (Н) и порезанные (П) фрагменты разделяли в 10 % ПААГ, напр.РНК — направляющая (crРНК). ПТЦ, ПНЦ – соответственно порезанная таргетная и нетаргетная цепи.

Наличие продуктов разрезания на гелях, представленных на Рисунке 16, Рисунке 18 и Рисунке 21 свидетельствует о том, что нуклеазная активность при связывании сплит-версии направляющей РНК сохраняется у всех исследованных ортологов Cas12a в условиях *in vitro* и является, по всей вероятности, общим свойством нуклеаз подтипа V-A. Хотя сплит-версии направляющих РНК работают в более высоких концентрациях, чем полноразмерные направляющие РНК, можно допустить, что использование таких РНК для специфического распознавания и гидролиза ДНК-матриц открывает возможности для будущих механистических исследований и практического применения эффекторов Cas12a, которые требуют

мультиплексирования и\или одновременного связывания нескольких целевых молекул.

3.10 Формирование делеций в геноме клеток линии HEK293T

Чтобы проверить активность кодон-оптимизированного RbCas12a в клетках млекопитающих, вносили делеции в ген *DNMT1* культуры клеток HEK293T с использованием пары sgРНК, закодированных в отдельных плазмидных векторах, и RbCas12a, закодированного в третьем векторе. ДНК-метилтрансфераза 1 (DNMT1) играет основную роль в метилировании ДНК, а также участвует в процессах репликации ДНК и клеточной пролиферации. Мутации в гене *DNMT1* ассоциированы с возникновением аутосомно-доминантной мозжечковой атаксии, глухоты и нарколепсии, а также наследственной сенсорной нейропатией с глухотой и деменцией. Нокаутные линии по гену *DNMT1* являются важными моделями этих заболеваний.

Клетки линии HEK293T снимали через 48 и 96 часов после трансфекции, после чего проверяли наличие делеций методом ПЦР. На Рисунке 22 представлена электрофореграмма геля с полученными ампликонами, а также относительные эффективности формирования делеций в %, и схема полученных данных секвенирования по Сенгеру клонированных ампликонов с делецией. Эффективность формирования делеции в одной реплике достигает 77.1 %, что является довольно высоким показателем для подобных экспериментов среди других ортологов.

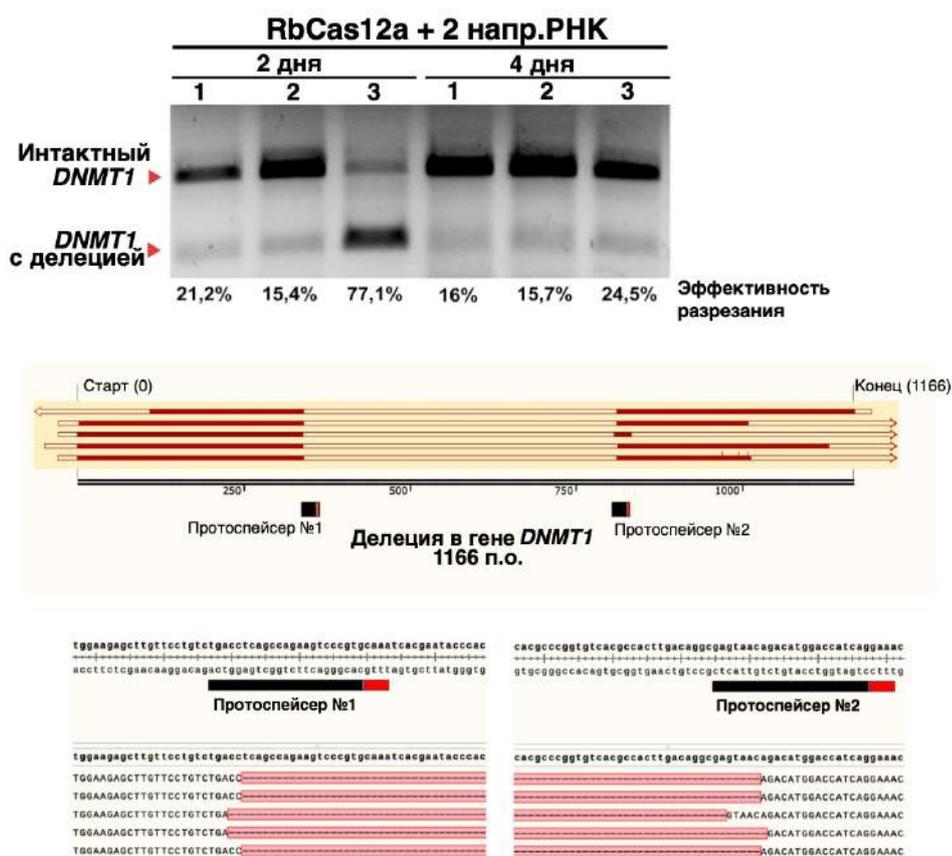


Рисунок 22 - Редактирование гена *DNMT1* двумя направляющими РНК в комплексе с RbCas12a с формированием делеции. Сверху: электрофореграмма ПЦР-продуктов участка гена *DNMT1* после трансфекции культуры клеток НЕК293Т тремя плазмидными векторами, кодирующими соответственно: сгРНК1, сгРНК2 и RbCas12a. Три независимых эксперимента пронумерованы 1–3 на второй и четвертый день после трансфекции. Показана схема ПЦР-продукта с двумя протоспейсерами, узнаваемыми направляющими РНК. Снизу: детальная карта делеции с нуклеотидной последовательностью, выявленной секвенированием по Сэнгеру. Сайты посадки сгРНК выделены черным цветом, РАМ выделены красным. Красно-белыми стрелками показаны выравненные хроматограммы после секвенирования по Сэнгеру; красные части стрелок — выравнивающиеся последовательности, белые — отсутствующие в хроматограммах.

Далее проводили высокопроизводительное секвенирование ампликонов, содержащих делеции. Поскольку каждое прочтение NGS состоит из двух фрагментов ДНК, фланкирующих делецию, связанных друг с другом, мы называем сайты лигирования «сайтами разрыва». Это место перехода от одного протоспейсера к другому. Как видно из Рисунка 23, наиболее распространенные сайты разрыва находятся между 19-м и 20-м нуклеотидами в области протоспейсера левого плеча делеций и между 17-м и 18-м нуклеотидами в зоне протоспейсера правого плеча делеций. Анализ показал, что эффективность вышеупомянутых сайтов разрыва составляла ~ 35 % как для левой, так и для правой crРНК в условиях эксперимента. Интересно отметить, что распределение сайтов разрывов внутри протоспейсера для правой направляющей РНК выглядит унимодальным, а для левой - бимодальным. Это может быть следствием разницы в эффективности разрезания матрицы ДНК комплексом RbCas12a в разных сайтах. Другое пиковое значение для сайтов разрыва в протоспейсере, обозначенном слева, составляло ~ 23 % и локализовалось между 18-й и 19-й парой нуклеотидов. Реакции разрезания проходили на ~ 10 % как для «левой», так и для «правой» crРНК, но в разных сайтах: через 21, 22 и 23 пары нуклеотидов внутри протоспейсера для «левой» направляющей РНК и через 14 и 15 пар нуклеотидов для «правой». Между вторым и четвертым днями эксперимента нет значительной разницы (p -значение критерия хи-квадрат $\gg 0,05$) в эффективности внесенных разрывов. Также нет статистической связи между разрывами в двух протоспейсерах в эксперименте.

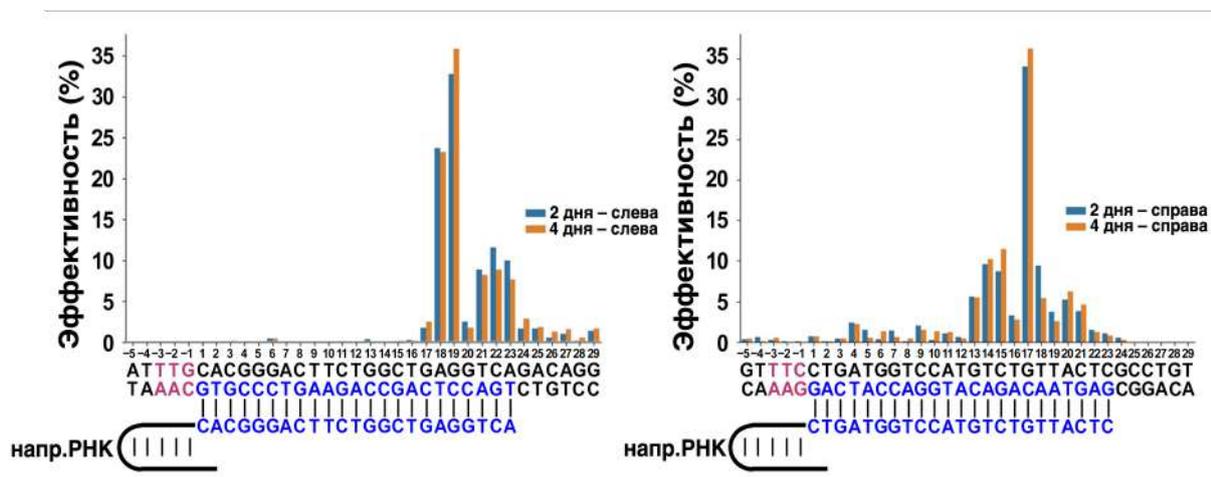


Рисунок 23 - Частоты сайтов разрыва, фланкирующих делецию в *DNMT1*. Слева: последовательность crРНК снизу, которая в комплексе с RbCas12a формирует разрывы в левом плече. Справа: последовательность crРНК снизу, которая в комплексе с RbCas12a формирует разрывы в правом плече. Последовательности PAM на обеих цепях выделены розовым цветом. Протоспейсерная часть таргетной цепи ДНК и спейсер crРНК выделены синим цветом. Относительная глубина покрытия в редактируемой области рассчитывалась как отношение глубины покрытия в определенной позиции к средней глубине покрытия за пределами редактируемой области. Эффективность разрыва цепи в определённой позиции была определена как разность между относительным покрытием в заданной и последующей позициях. Напр.РНК — направляющая (crРНК).

Таким образом, исходя из полученных данных, RbCas12a активен в клетках человека и способен формировать значительные делеции длиной как минимум в 471 нуклеотид в геноме культуры клеток человека.

ЗАКЛЮЧЕНИЕ

Нуклеазы CRISPR/Cas подтипа V-A (Cas12a), также известные как Cpf1, представляют собой инструмент редактирования генома, который рассматривается в качестве альтернативы нуклеазам Cas9. Хотя SpCas9 по-прежнему используется в лабораторной практике чаще, чем нуклеазы семейства Cas12a, последние привлекают все более пристальное внимание, в основном за счет меньшего размера как самих нуклеаз, так и crРНК. В данной работе проведен анализ активностей нового ортолога эффекторных эндонуклеаз систем CRISPR/Cas подтипа V-A из *Ruminococcus bromii* sp. Новый фермент был нами отнесен к семейству Cas12a в соответствии с четырьмя независимыми признаками: 1) RbCas12a кластеризуется с другими ферментами Cas12a на филогенетическом древе (Приложение 1); 2) скаффолд crРНК этой нуклеазы аналогичен скаффолдам ранее опубликованных ортологов Cas12a (Приложение 2); 3) PAM-последовательность RbCas12a схожа с T-богатыми последовательностями PAM других ортологов; 4) аминокислотная последовательность RbCas12a имеет те же консервативные каталитические участки, что и другие ортологи Cas12a (Приложение 1).

Для идентификации условий разрезания *in vitro* ДНК-мишени с помощью RbCas12a, были протестированы различные двухвалентные катионы, добавляемые в реакционный буфер. Известно, что присутствие ионов марганца приводит к почти полной деградации дцДНК нуклеазой AsCas12a, даже в отсутствие crРНК, в отличие, например, от LbCas12a [106]. Полученные в ходе выполненной работы данные свидетельствуют о том, что комплекс RbCas12a/crРНК способен использовать в качестве кофактора магний либо марганец, в результате чего происходит эффективный направленный гидролиз дцДНК без видимой неспецифической деградации матрицы. Более того, цинк и никель могут также быть использованы в качестве кофакторов RbCas12a, хотя и с меньшей эффективностью. Интересно, что кальций не является эффективным

кофактором RbCas12a, хотя с ним работают некоторые другие ортологи, например, FnCas12a [110]. Таким образом, полученные результаты подтверждают высказанный в [106] тезис о том, что разные ортологи Cas12a обладают разным предпочтением к различным кофакторам.

Далее проверялась активность RbCas12a при различных значениях температур протекания реакций разрезания *in vitro*. В отличие от некоторых других похожих работ по характеристике температурных диапазонов нуклеаз, например, [111], в ходе данной работы все компоненты смеси преднагревались до соответствующих температур по отдельности, а не будучи в общем миксе с дальнейшей инкубацией на заданной температуре. Это позволяет более точно охарактеризовать возможность сборки эффекторного комплекса и оценить эффективность гидролиза ДНК-матрицы при конкретной температуре. RbCas12a активен при температурах от 20 до 42 °С, хотя при температурах ниже 25 °С комплекс теряет стабильность. Тем не менее, следует учитывать, что эти данные не следует рассматривать как результаты, отражающие возможную активность нуклеазы *in vivo*, а вместо этого могут быть применены в отношении ряда методов *in vitro*, основанных на использовании CRISPR/Cas12a. Более того, скорости разрезания, полученные в данной работе и в аналогичных исследованиях, отражают только конкретные условия эксперимента (буфер, соотношение crРНК:белок:ДНК, конкретный спейсер crРНК, PAM-последовательность и ее нуклеотидное окружение — то есть специфичность конкретной crРНК) и не должны транслироваться на любые другие эксперименты. Несмотря на эти ограничения, вышеупомянутые эксперименты могут помочь лучше понять природу фермента и его активность *in vitro*. Кроме того, имеется определенная биотехнологическая и сельскохозяйственная потребность в ферментах, активных при более низких температурах, и надо полагать, что относительно эффективное разрезание при температурах вплоть до 20 °С может предоставить возможность для

дальнейших исследований, направленных, например, на редактирование генома растений, поскольку данные эксперименты по-прежнему очень чувствительны к температурным условиям.

В многочисленных работах было показано, что для разных Cas12a нуклеаз требуются схожие, но все же различные T-богатые PAM. Несмотря на попытки расширения PAM-специфичности [112]–[114], по-прежнему остается сложным редактировать не только любой желаемый, но даже любой T-богатый локус генома. Известно, что AsCas12a и LbCas12a распознают PAM-последовательность 5'-TTTV, а FnCas12a и MbCas12a - консенсус 5'-TTN [115]. Стоит дополнить, что помимо указанных канонических PAM, некоторые Cas12a также слабо взаимодействуют с матрицами, содержащими некоторые субоптимальные PAM. По результатам проведенных в работе экспериментов, нуклеаза RbCas12a распознает 5'-YYN PAM, что, насколько можно судить по имеющейся литературе, является наименее строгим нативным PAM из известных.

Согласно ранее опубликованным данным, эффективности разрезания комплексами Cas12a/crPНК могут различаться для экспериментов *in vitro* и *in vivo*. Доступность хроматина для Cas12a, по-видимому, является одной из причин этих различий, хотя этот вопрос требует дополнительного изучения. Насколько известно из имеющейся литературы, существует только один протокол с использованием нуклеаз Cas12a, составленный группой под руководством Jinek [116] и описывающий внесение делеций в геном клеток млекопитающих. В указанном эксперименте были внесены делеции ~500 п.о. в геном мышинных эмбриональных стволовых клеток, трансфецированных RNP комплексами, заряженными двумя разными crPНК. В ходе данной работы была продемонстрирована возможность использования другого ортолога Cas12a для внесения делеций в геном клеток млекопитающих, подтверждая эффективность RbCas12a в качестве инструмента редактирования генома клеток человека.

ВЫВОДЫ

1. По результатам биоинформатического поиска в геноме *Ruminococcus bromii* sp. обнаружен ген ранее не описанной в литературе нуклеазы семейства Cas12a (RbCas12a) длиной 3735 п.о.

2. Показана активность рекомбинантного белка RbCas12a в составе комплексов с различными crРНК в реакциях разрезания *in vitro* на различных ДНК-матрицах, содержащих PAM-последовательности с консенсусом 5'-YYN-3'.

3. Нуклеаза RbCas12a способна использовать в качестве кофактора ионы магния и марганца, а также, хоть с меньшей эффективностью, ионы кобальта, цинка и никеля. Рабочий температурный диапазон RbCas12a варьирует от 20 до 42 °С, с наибольшей активностью в диапазоне 30-42 °С.

4. Нуклеаза RbCas12a способна проявлять коллатеральную активность в отношении оцДНК.

5. Определена природа выступающих концов двуцепочечных продуктов гидролиза ДНК-матрицы с помощью нуклеазы RbCas12a: продукты имеют 5'-концевые выступы, длина которых зависит от экзонуклеазной активности фермента в отношении таргетной цепи и может составлять 4-6 нуклеотидов.

6. Нуклеазы Cas подтипа V-A (RbCas12a, AsCas12a, FnCas12a, LbCas12a) формируют комплексы с двумя молекулами РНК, представляющими из себя отдельные спейсерную и скаффолдную части crРНК (сплит РНК), и осуществляют в их составе направленный гидролиз ДНК-матриц.

7. С помощью RbCas12a получена направленная делеция в гене *DNMT1* культуры клеток HEK293T.

СПИСОК ЛИТЕРАТУРЫ

1. Ishino Y., Shinagawa H., Makino K., Amemura M., Nakamura A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product // *J. Bacteriol.* – 1987. – V. 169. – № 12. – P. 5429–5433.
2. Mitchell D.A., Teague W.J., Jarosz E., Wang D.W., Res Lett G., Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., Horvath P. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes // *Science.* – 2007. – V. 315. – № 5819. – P. 1709–1712.
3. Makarova K.S., Haft D.H., Barrangou R., Brouns S.J.J., Charpentier E., Horvath P., Moineau S., Mojica F.J.M., Wolf Y.I., Yakunin A.F., Van Der Oost J., Koonin E. V. Evolution and classification of the CRISPR-Cas systems // *Nat. Rev. Microbiol.* – 2011. – V. 9. – № 6. – P. 467–477.
4. Makarova K.S., Grishin N. V., Shabalina S.A., Wolf Y.I., Koonin E. V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action // *Biol. Direct.* – 2006. – V. 1. – № 1. – P. 1–26.
5. Elitza D., Krzysztof C., Sharma Cynthia M, Karine G., Yanjie C., Pirzada Zaid A, Eckert Maria R, Jörg V., Emmanuelle C. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III // *Nature.* – 2011. – V. 471. – № 7340. – P. 602–607.
6. Mali P., Yang L., Esvelt K.M., Aach J., Guell M., DiCarlo J.E., Norville J.E., Church G.M. RNA-guided human genome engineering via Cas9 // *Science.* – 2013. – V. 339. – № 6121. – P. 823–826.
7. 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. Multiplex genome engineering using CRISPR/Cas systems // *Science.* – 2013. – V. 339. – № 6121. – P. 819–823.
8. Scott D.A., Weinstein J.A., Ran F.A., Konermann S., Li Y., Fine E.J., Wu X., Shalem O., Cradick T.J., Luciano A. DNA targeting specificity of RNA-guided Cas9 nucleases // *Nat. Biotechnol.* – 2013. – V. 31. – № 9. – P. 827–832.

9. Shalem O., Sanjana N.E., Hartenian E., Shi X., Scott D.A., Mikkelsen T.S., Heckl D., Ebert B.L., Root D.E., Doench J.G., Zhang F. Genome-scale CRISPR-Cas9 knockout screening in human cells // *Science*. – 2014. – V. 343. – № 6166. – P. 84–87.
10. 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. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System // *Cell*. – 2015. – V. 163. – № 3. – P. 759–771.
11. Min Y.L., Li H., Rodriguez-Caycedo C., Mireault A.A., Huang J., Shelton J.M., McAnally J.R., Amoasii L., Mammen P.P.A., Bassel-Duby R., Olson E.N. CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells // *Sci. Adv.* – 2019. – V. 5. – № 3. – P. eaav4324
12. Ohmori T., Mizukami H., Ozawa K., Sakata Y., Nishimura S. New approaches to gene and cell therapy for hemophilia // *J. Thromb. Haemost.* – 2015. – V. 13. – № S1. – P. S133–S142.
13. Khosravi M.A., Abbasalipour M., Concordet J.P., Berg J. Vom, Zeinali S., Arashkia A., Azadmanesh K., Buch T., Karimipour M. Targeted deletion of BCL11A gene by CRISPR-Cas9 system for fetal hemoglobin reactivation: A promising approach for gene therapy of beta thalassemia disease // *Eur. J. Pharmacol.* – 2019. – V. 854. – P. 398–405.
14. György B., Nist-Lund C., Pan B., Asai Y., Karavitaki K.D., Kleinstiver B.P., Garcia S.P., Zaborowski M.P., Solanes P., Spataro S., Schneider B.L., Joung J.K., Géléoc G.S.G., Holt J.R., Corey D.P. Allele-specific gene editing prevents deafness in a model of dominant progressive hearing loss // *Nat. Med.* – 2019. – V. 25. – № 7. – P. 1123–1130.
15. Hou P., Chen S., Wang S., Yu X., Chen Y., Jiang M., Zhuang K., Ho W., Hou W., Huang J., Guo D. Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection // *Sci. Rep.* – 2015. – V. 5. – № 1. – P. 1–12.
16. Gergen J., Coulon F., Crenequy A., Elain-Duret N., Gutierrez A., Pinkenburg O., Verhoeyen E., Anegon I., Nguyen T.H., Halary F.A., Haspot F. Multiplex CRISPR/Cas9 system impairs HCMV replication by excising an essential viral gene // *PLoS One* – 2018. – V. 13. – № 2. – P. e0192602

17. van Diemen F.R., Kruse E.M., Hooykaas M.J.G., Bruggeling C.E., Schürch A.C., van Ham P.M., Imhof S.M., Nijhuis M., Wiertz E.J.H.J., Lebbink R.J. CRISPR/Cas9-Mediated Genome Editing of Herpesviruses Limits Productive and Latent Infections // *PLoS Pathog.* – 2016. – V. 12. – № 6. – P. e1005701
18. Nidhi S., Anand U., Oleksak P., Tripathi P., Lal J.A., Thomas G., Kuca K., Tripathi V. Novel crispr–cas systems: An updated review of the current achievements, applications, and future research perspectives // *Int. J. Mol. Sci.* – 2021. – V. 22. – № 7. – P. 3327
19. Gao L., Altae-Tran H., Böhning F., Makarova K.S., Segel M., Schmid-Burgk J.L., Koob J., Wolf Y.I., Koonin E. V., Zhang F. Diverse enzymatic activities mediate antiviral immunity in prokaryotes // *Science.* – 2020. – V. 369. – № 6507. – P. 1077–1084.
20. Makarova K.S., Wolf Y.I., Snir S., Koonin E. V. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems // *J. Bacteriol.* – 2011. – V. 193. – № 21. – P. 6039–6056.
21. Koonin E. V., Makarova K.S., Wolf Y.I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria // *Annu. Rev. Microbiol.* – 2017. – V. 71. – P. 233–261.
22. Makarova K.S., Wolf Y.I., Koonin E. V. Comparative genomics of defense systems in archaea and bacteria // *Nucleic Acids Res.* – 2013. – V. 41. – № 8. – P. 4360–4377.
23. Isaev A.B., Musharova O.S., Severinov K. V. Microbial Arsenal of Antiviral Defenses – Part I // *Biochem.* – 2021. – V. 86. – № 3. – P. 319–337.
24. Mruk I., Kobayashi I. To be or not to be: Regulation of restriction-modification systems and other toxin-antitoxin systems // *Nucleic Acids Res.* – 2014. – V. 42. – № 1. – P. 70–86.
25. Zhang Y., Zhang Z., Zhang H., Zhao Y., Zhang Z., Xiao J. PADS Arsenal: A database of prokaryotic defense systems related genes // *Nucleic Acids Res.* – 2020. – V. 48. – № D1. – P. D590–D598.
26. Wang L., Chen S., Xu T., Taghizadeh K., Wishnok J.S., Zhou X., You D., Deng Z., Dedon P.C. Phosphorothioation of DNA in bacteria by dnd genes // *Nat. Chem. Biol.* – 2007. – V. 3. – № 11. – P. 709–710.

27. Thiaville J.J., Kellner S.M., Yuan Y., Hutinet G., Thiaville P.C., Jumpathong W., Mohapatra S., Brochier-Armanet C., Letarov A. V., Hillebrand R., Malik C.K., Rizzo C.J., Dedon P.C., De Crécy-Lagard V. Novel genomic island modifies DNA with 7-deazaguanine derivatives // *Proc. Natl. Acad. Sci. U. S. A.* – 2016. – V. 113. – № 11. – P. E1452–E1459.
28. Lopatina A., Tal N., Sorek R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy // *Annu. Rev. Virol.* – 2020. – V. 7. – P. 371–384.
29. Kronheim S., Daniel-Ivad M., Duan Z., Hwang S., Wong A.I., Mantel I., Nodwell J.R., Maxwell K.L. A chemical defence against phage infection // *Nature.* – 2018. – V. 564. – № 7735. – P. 283–286.
30. Rousset F., Depardieu F., Miele S., Dowding J., Laval A.-L., Lieberman E., Garry D., Rocha E.P.C., Bernheim A., Bikard D. Phages and their satellites encode hotspots of antiviral systems // *Cell Host Microbe.* – 2022. – V. 30. – № 5. – P. 740–753.
31. Mestre M.R., González-Delgado A., Gutiérrez-Rus L.I., Martínez-Abarca F., Toro N. Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems // *Nucleic Acids Res.* – 2020. – V. 48. – № 22. – P. 12632–12647.
32. Kuzmenko A., Oguienko A., Esyunina D., Yudin D., Petrova M., Kudinova A., Maslova O., Ninova M., Ryazansky S., Leach D., Aravin A.A., Kulbachinskiy A. DNA targeting and interference by a bacterial Argonaute nuclease // *Nature.* – 2020. – V. 587. – № 7835. – P. 632–637.
33. Hutvagner G., Simard M.J. Argonaute proteins: Key players in RNA silencing // *Nat. Rev. Mol. Cell Biol.* – 2008. – V. 9. – № 1. – P. 22–32.
34. Kuzmenko A., Yudin D., Ryazansky S., Kulbachinskiy A., Aravin A.A. Programmable DNA cleavage by Ago nucleases from mesophilic bacteria *Clostridium butyricum* and *Limnothrix rosea* // *Nucleic Acids Res.* – 2019. – V. 47. – № 11. – P. 5822–5836.
35. Kaya E., Doxzen K.W., Knoll K.R., Wilson R.C., Strutt S.C., Kranzusch P.J., Doudna J.A. A bacterial Argonaute with noncanonical guide RNA specificity // *Proc. Natl. Acad. Sci. U. S. A.* – 2016. – V. 113. – № 15. – P. 4057–4062.

36. Bernheim A., Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource // *Nat. Rev. Microbiol.* – 2020. – V. 18. – № 2. – P. 113–119.
37. Isaev A.B., Musharova O.S., Severinov K. V. Microbial Arsenal of Antiviral Defenses. Part II // *Biochem.* – 2021. – V. 86. – № 4. – P. 449–470.
38. Makarova K.S., Wolf Y.I., Koonin E. V. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? // *Cris. J.* – 2018. – V. 1. – № 5. – P. 325–336.
39. Özcan A., Krajeski R., Ioannidi E., Lee B., Gardner A., Makarova K.S., Koonin E. V., Abudayyeh O.O., Gootenberg J.S. Programmable RNA targeting with the single-protein CRISPR effector Cas7-11 // *Nature.* – 2021. – V. 597. – № 7878. – P. 720–725.
40. Lee H., Sashital D.G. Creating memories: molecular mechanisms of CRISPR adaptation // *Trends Biochem. Sci.* – 2022. – V. 47. – № 6. – P. 464–476.
41. Wimmer F., Beisel C.L. CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers // *Front. Microbiol.* – 2020. – V. 10. – P. 3078.
42. Amitai G., Sorek R. CRISPR-Cas adaptation: Insights into the mechanism of action // *Nat. Rev. Microbiol.* – 2016. – V. 14. – № 2. – P. 67–76.
43. Subramaniam S., Smith G.R. Myths and mechanisms: RecBCD and Chi hotspots as determinants of self vs . non-self // *BioRxiv* 2021.07.08.451572.
44. Gussow A.B., Park A.E., Borges A.L., Shmakov S.A., Makarova K.S., Wolf Y.I., Bondy-Denomy J., Koonin E. V. Machine-learning approach expands the repertoire of anti-CRISPR protein families // *Nat. Commun.* – 2020. – V. 11. – № 1. – P. 1-12.
45. Datsenko K.A., Pougach K., Tikhonov A., Wanner B.L., Severinov K., Semenova E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system // *Nat. Commun.* – 2012. – V. 3. – № 1. – P. 1-7.
46. Zheng Y., Li J., Wang B., Han J., Hao Y., Wang S., Ma X., Yang S., Ma L., Yi L., Peng W. Endogenous Type I CRISPR-Cas: From Foreign DNA Defense to Prokaryotic Engineering // *Front. Bioeng. Biotechnol.* – 2020. – V. 8. – P. 62.

47. Charpentier E., Richter H., van der Oost J., White M.F. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity // *FEMS Microbiol. Rev.* – 2015. – V. 39. – № 3. – P. 428–441.
48. Musharova O., Medvedeva S., Klimuk E., Guzman N.M., Titova D., Zgoda V., Shiriaeva A., Semenova E., Severinov K., Savitskaya E. Prespacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase // *Proc. Natl. Acad. Sci. U. S. A.* – 2021. – V. 118. – № 22. – P. e2021291118.
49. Kim S., Loeff L., Colombo S., Jergic S., Brouns S.J.J., Joo C. Selective loading and processing of prespacers for precise CRISPR adaptation // *Nature.* – 2020. – V. 579. – № 7797. – P. 141–145.
50. Heler R., Samai P., Modell J.W., Weiner C., Goldberg G.W., Bikard D., Marraffini L.A. Cas9 specifies functional viral targets during CRISPR-Cas adaptation // *Nature.* – 2015. – V. 519. – № 7542. – P. 199–202.
51. González-Delgado A., Rodríguez Mestre M., Martínez-Abarca F., Toro N. Spacer acquisition from RNA mediated by a natural reverse transcriptase-Cas1 fusion protein associated with a type III-D CRISPR–Cas system in *Vibrio vulnificus* // *Nucleic Acids Res.* – 2019. – V. 47. – № 19. – P. 10202–10211.
52. Artamonova D., Karneyeva K., Medvedeva S., Klimuk E., Kolesnik M., Yasinskaya A., Samolygo A., Severinov K. Spacer acquisition by Type III CRISPR–Cas system during bacteriophage infection of *Thermus thermophilus* // *Nucleic Acids Res.* – 2020. – V. 48. – № 17. – P. 9787–9803.
53. Newire E., Aydin A., Juma S., Enne V.I., Roberts A.P. Identification of a Type IV-A CRISPR-Cas System Located Exclusively on IncHI1B/IncFIB Plasmids in Enterobacteriaceae // *Front. Microbiol.* – 2020. – V. 11. – P. 1937.
54. Hoikkala V., Ravanti J., Díez-Villaseñor C., Tiirola M., Conrad R.A., McBride M.J., Moineau S., Sundberg L.R. Cooperation between different crispr-cas types enables adaptation in an RNA-targeting system // *Mbio.* – 2021. – V. 12. – № 2. – P. e03338-20.
55. 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., ... Koonin E. V. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants // *Nat. Rev. Microbiol.* – 2020. – V. 18. – № 2. – P. 67–83.
56. Tong B., Dong H., Cui Y., Jiang P., Jin Z., Zhang D. The Versatile Type V CRISPR Effectors and Their Application Prospects // *Front. Cell Dev. Biol.* – 2021. – V. 8. – P. 622103.
57. Karvelis T., Gasiunas G., Miksys A., Barrangou R., Horvath P., Siksnys V. crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus* // *RNA Biol.* – 2013. – V. 10. – № 5. – P. 841–851.
58. Zhang Y., Heidrich N., Ampattu B.J., Gunderson C.W., Seifert H.S., Schoen C., Vogel J., Sontheimer E.J. Processing-Independent CRISPR RNAs Limit Natural Transformation in *Neisseria meningitidis* // *Mol. Cell.* – 2013. – V. 50. – № 4. – P. 488–503.
59. Taylor H.N., Warner E.E., Armbrust M.J., Crowley V.M., Olsen K.J., Jackson R.N. Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease // *RNA Biol.* – 2019. – V. 16. – № 10. – P. 1438–1447.
60. Fonfara I., Richter H., Bratovič M., Le Rhun A., Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA // *Nature.* – 2016. – V. 532. – № 7600. – P. 517–521.
61. Swarts D.C., van der Oost J., Jinek M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a // *Mol. Cell.* – 2017. – V. 66. – № 2. – P. 221-233.e4.
62. Yan W.X., Hunnewell P., Alfonse L.E., Carte J.M., Keston-Smith E., Sothivelvam S., Garrity A.J., Chong S., Makarova K.S., Koonin E. V., Cheng D.R., Scott D.A. Functionally diverse type V CRISPR-Cas systems // *Science.* – 2019. – V. 363. – № 6422. – P. 88–91.
63. Zhang H., Li Z., Xiao R., Chang L. Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease // *Nat. Struct. Mol. Biol.* – 2020. – V. 27. – № 11. – P. 1069–1076.
64. Harrington L.B., Ma E., Chen J.S., Witte I.P., Gertz D., Paez-Espino D., Al-Shayeb B., Kyrpides N.C., Burstein D., Banfield J.F., Doudna J.A. A

- scoutRNA Is Required for Some Type V CRISPR-Cas Systems // *Mol. Cell.* – 2020. – V. 79. – № 3. – P. 416-424.e5.
65. Pausch P., Al-Shayeb B., Bisom-Rapp E., Tsuchida C.A., Li Z., Cress B.F., Knott G.J., Jacobsen S.E., Banfield J.F., Doudna J.A. Crispr-casf from huge phages is a hypercompact genome editor // *Science.* – 2020. – V. 369. – № 6501. – P. 333–337.
 66. Kick L.M., von Wrisberg M.K., Runtzsch L.S., Schneider S. Structure and mechanism of the RNA dependent RNase Cas13a from *Rhodobacter capsulatus* // *Commun. Biol.* – 2022. – V. 5. – № 1. – P. 1-9.
 67. Knott G.J., East-Seletsky A., Cofsky J.C., Holton J.M., Charles E., O’Connell M.R., Doudna J.A. Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme // *Nat. Struct. Mol. Biol.* – 2017. – V. 24. – № 10. – P. 825–833.
 68. Liu L., Li X., Ma J., Li Z., You L., Wang J., Wang M., Zhang X., Wang Y. The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a // *Cell.* – 2017. – V. 170. – № 4. – P. 714-726.e10.
 69. Molina R., Sofos N., Montoya G. Structural basis of CRISPR-Cas Type III prokaryotic defence systems // *Curr. Opin. Struct. Biol.* – 2020. – V. 65. – P. 119–129.
 70. Kolesnik M. V., Fedorova I., Karneyeva K.A., Artamonova D.N., Severinov K. V. Type III CRISPR-Cas Systems: Deciphering the Most Complex Prokaryotic Immune System // *Biochem.* – 2021. – V. 86. – № 10. – P. 1301–1314.
 71. Jiang F., Taylor D.W., Chen J.S., Kornfeld J.E., Zhou K., Thompson A.J., Nogales E., Doudna J.A. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage // *Science.* – 2016. – V. 351. – № 6275. – P. 867–871.
 72. Anders C., Niewoehner O., Duerst A., Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease // *Nature.* – 2014. – V. 513. – № 7519. – P. 569–573.
 73. Jiang F., Doudna J.A. CRISPR-Cas9 Structures and Mechanisms // *Annu. Rev. Biophys.* – 2017. – V. 46. – P. 505–529.

74. Westra E.R., van Erp P.B.G., Künne T., Wong S.P., Staals R.H.J., Seegers C.L.C., Bollen S., Jore M.M., Semenova E., Severinov K., de Vos W.M., Dame R.T., de Vries R., Brouns S.J.J., van der Oost J. CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3 // *Mol. Cell.* – 2012. – V. 46. – № 5. – P. 595–605.
75. Cofsky J.C., Karandur D., Huang C.J., Witte I.P., Kuriyan J., Doudna J.A. CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks // *Elife.* – 2020. – V. 9. – P. 1–32.
76. Zhang B., Luo D., Li Y., Perčulija V., Chen J., Lin J., Ye Y., Ouyang S. Mechanistic insights into the R-loop formation and cleavage in CRISPR-Cas12i1 // *Nat. Commun.* – 2021. – V. 12. – № 1. – P. 1-13.
77. Stella S., Alcón P., Montoya G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage // *Nature.* – 2017. – V. 546. – № 7659. – P. 559–563.
78. Swarts D.C., Jinek M. Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a // *Mol. Cell.* – 2019. – V. 73. – № 3. – P. 589-600.e4.
79. van Aelst K., Martínez-Santiago C.J., Cross S.J., Szczelkun M.D. The effect of DNA topology on observed rates of R-loop formation and DNA strand cleavage by CRISPR Cas12a // *Genes (Basel).* – 2019. – V. 10. – № 2. – P. 169.
80. Naqvi M.M., Lee L., Montaguth O.E.T., Diffin F.M., Szczelkun M.D. CRISPR–Cas12a-mediated DNA clamping triggers target-strand cleavage // *Nat. Chem. Biol.* – 2022. – V. 18. – № 9. – P. 1014–1022.
81. Yang H., Gao P., Rajashankar K.R., Patel D.J. PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease // *Cell.* – 2016. – V. 167. – № 7. – P. 1814-1828.e12.
82. Kurihara N., Nakagawa R., Hirano H., Okazaki S., Tomita A., Kobayashi K., Kusakizako T., Nishizawa T., Yamashita K., Scott D.A., Nishimasu H., Nureki O. Structure of the type V-C CRISPR-Cas effector enzyme // *Mol. Cell.* – 2022. – V. 82. – № 10. – P. 1865-1877.

83. Liu J.J., Orlova N., Oakes B.L., Ma E., Spinner H.B., Baney K.L.M., Chuck J., Tan D., Knott G.J., Harrington L.B., Al-Shayeb B., Wagner A., Brötzmann J., Staahl B.T., Taylor K.L., Desmarais J., Nogales E., Doudna J.A. CasX enzymes comprise a distinct family of RNA-guided genome editors // *Nature*. – 2019. – V. 566. – № 7743. – P. 218–223.
84. 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. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes // *Science*. – 2018. – V. 362. – № 6416. – P. 839–842.
85. Takeda S.N., Nakagawa R., Okazaki S., Hirano H., Kobayashi K., Kusakizako T., Nishizawa T., Yamashita K., Nishimasu H., Nureki O. Structure of the miniature type V-F CRISPR-Cas effector enzyme // *Mol. Cell*. – 2021. – V. 81. – № 3. – P. 558-570.e3.
86. Xiao R., Li Z., Wang S., Han R., Chang L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease // *Nucleic Acids Res*. – 2021. – V. 49. – № 7. – P. 4120–4128.
87. Querques I., Schmitz M., Oberli S., Chanez C., Jinek M. Target site selection and remodelling by type V CRISPR-transposon systems // *Nature*. – 2021. – V. 599. – № 7885. – P. 497–502.
88. Li Z., Zhang H., Xiao R., Han R., Chang L. Cryo-EM structure of the RNA-guided ribonuclease Cas12g // *Nat. Chem. Biol*. – 2021. – V. 17. – № 4. – P. 387–393.
89. Hall R.A., Macdonald J. Synthetic Biology Provides a Toehold in the Fight against Zika // *Cell Host Microbe*. – 2016. – V. 19. – № 6. – P. 752–754.
90. Takahashi M.K., Tan X., Dy A.J., Braff D., Akana R.T., Furuta Y., Donghia N., Ananthakrishnan A., Collins J.J. A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers // *Nat. Commun*. – 2018. – V. 9. – № 1. – P. 1-12.
91. Broughton J.P., Deng X., Yu G., Fasching C.L., Servellita V., Singh J., Miao X., Streithorst J.A., Granados A., Sotomayor-Gonzalez A., Zorn K., Gopez A., Hsu E., Gu W., Miller S., Pan C.Y., Guevara H., Wadford D.A., Chen J.S., Chiu C.Y. CRISPR–Cas12-based detection of SARS-CoV-2 // *Nat. Biotechnol*. – 2020. – V. 38. – № 7. – P. 870–874.

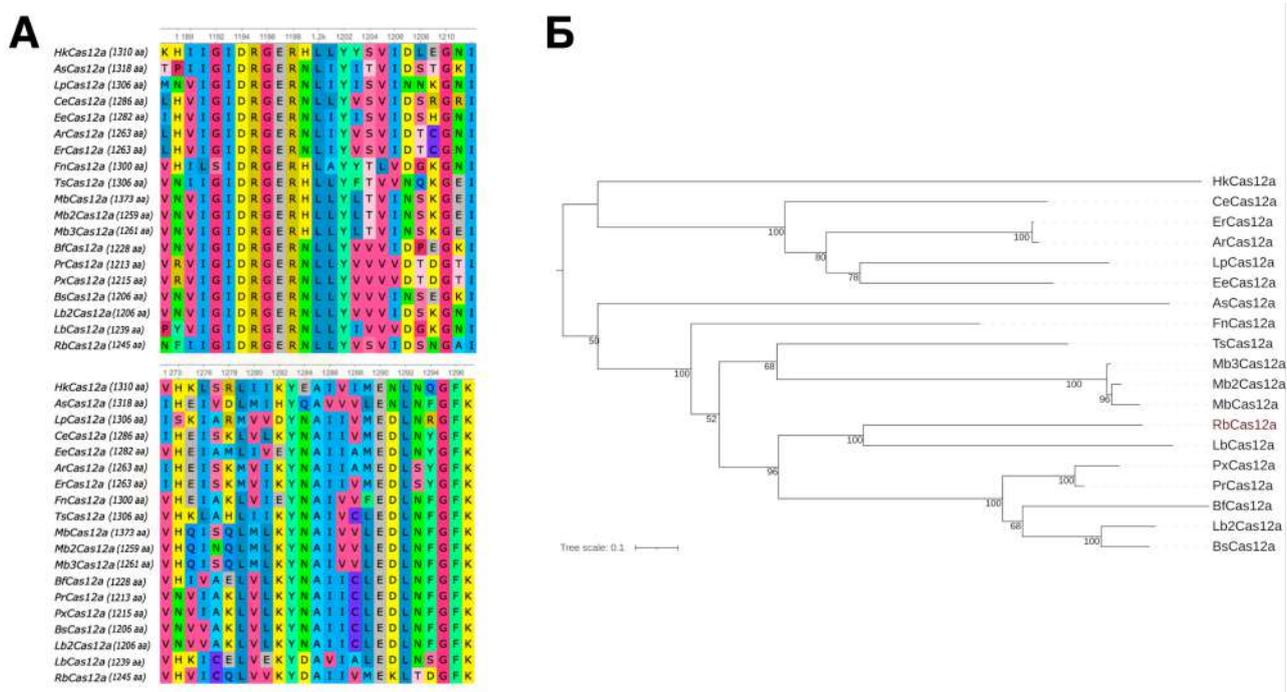
92. Kellner M.J., Koob J.G., Gootenberg J.S., Abudayyeh O.O., Zhang F. SHER-LOCK: nucleic acid detection with CRISPR nucleases // *Nat. Protoc.* – 2019. – V. 14. – № 10. – P. 2986–3012.
93. Li L., Li S., Wu N., Wu J., Wang G., Zhao G., Wang J. HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation // *ACS Synth. Biol.* – 2019. – V. 8. – № 10. – P. 2228–2237.
94. Li S.Y., Cheng Q.X., Li X.Y., Zhang Z.L., Gao S., Cao R.B., Zhao G.P., Wang J., Wang J.M. CRISPR-Cas12a-assisted nucleic acid detection // *Cell Discov.* – 2018. – V. 4. – № 1. – P. 1-4.
95. Teng F., Guo L., Cui T., Wang X.G., Xu K., Gao Q., Zhou Q., Li W. CDetection: CRISPR-Cas12b-based DNA detection with sub-attomolar sensitivity and single-base specificity // *Genome Biol.* – 2019. – V. 20. – № 1. – P. 1-7.
96. Jiang Y., Hu M., Liu A.A., Lin Y., Liu L., Yu B., Zhou X., Pang D.W. Detection of SARS-CoV-2 by CRISPR/Cas12a-Enhanced Colorimetry // *ACS Sensors.* – 2021. – V. 6. – № 3. – P. 1086–1093.
97. Arizti-Sanz J., Freije C.A., Stanton A.C., Petros B.A., Boehm C.K., Siddiqui S., Shaw B.M., Adams G., Kosoko-Thoroddsen T.S.F., Kembell M.E., Uwanibe J.N., Ajogbasile F. V., Eromon P.E., Gross R., Wronka L., Caviness K., Hensley L.E., Bergman N.H., MacInnis B.L., ... Myhrvold C. Streamlined inactivation, amplification, and Cas13-based detection of SARS-CoV-2 // *Nat. Commun.* – 2020. – V. 11. – № 1. – P. 1-9.
98. Russel J., Pinilla-Redondo R., Mayo-Muñoz D., Shah S.A., Sørensen S.J. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci // *Cris. J.* – 2020. – V. 3. – № 6. – P. 462–469.
99. Biswas A., Gagnon J.N., Brouns S.J.J., Fineran P.C., Brown C.M. CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets // *RNA Biol.* – 2013. – V. 10. – № 5. – P. 817–827.
100. Sambrook J., Russell D.W. The Inoue Method for Preparation and Transformation of Competent E. Coli : “Ultra-Competent” Cells // *Cold Spring Harb. Protoc.* – 2006. – V. 2006. – № 1. – P. 10-1101.
101. Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. WebLogo: A sequence logo generator // *Genome Res.* – 2004. – V. 14. – № 6. – P. 1188–1190.

102. Schneider TD, Stephens RM Sequence logos: a new way to display consensus sequences // *Nucleic Acids Res.* – 1990. – V. 18. – № 20. – P. 6097–6100.
103. Enghiad B., Huang C., Guo F., Jiang G., Wang B., Tabatabaei S.K., Martin T.A., Zhao H. Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination // *Nat. Commun.* – 2021. – V. 12. – № 1. – P. 1-11.
104. Oliveros J.C., Franch M., Tabas-Madrid D., San-León D., Montoliu L., Cubas P., Pazos F. Breaking-Cas-interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes // *Nucleic Acids Res.* – 2016. – V. 44. – № 1. – P. W267–W271.
105. Okonechnikov K., Golosova O., Fursov M., Varlamov A., Vaskin Y., Efremov I., German Grehov O.G., Kandrov D., Rasputin K., Syabro M., Tleukenov T. Unipro UGENE: A unified bioinformatics toolkit // *Bioinformatics.* – 2012. – V. 28. – № 8. – P. 1166–1167.
106. Li B., Yan J., Zhang Y., Li W., Zeng C., Zhao W., Hou X., Zhang C., Dong Y. CRISPR-Cas12a Possesses Unconventional DNase Activity that Can Be Inactivated by Synthetic Oligonucleotides // *Mol. Ther. - Nucleic Acids.* – 2020. – V. 19. – P. 1043–1052.
107. Teng F., Li J., Cui T., Xu K., Guo L., Gao Q., Feng G., Chen C., Han D., Zhou Q., Li W. Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds // *Genome Biol.* – 2019. – V. 20. – № 1. – P. 1-6.
108. Yamano T., Zetsche B., Ishitani R., Zhang F., Nishimasu H., Nureki O. Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1 // *Mol. Cell.* – 2017. – V. 67. – № 4. – P. 633-645.e3.
109. Ahn W.C., Park K.H., Bak I.S., Song H.N., An Y., Lee S.J., Jung M., Yoo K.W., Yu D.Y., Kim Y.S., Oh B.H., Woo E.J. In vivo genome editing using the Cpf1 ortholog derived from *Eubacterium eligens* // *Sci. Rep.* – 2019. – V. 9. – № 1. – P. 1-7.
110. Singh D., Mallon J., Poddar A., Wang Y., Tippiana R., Yang O., Bailey S., Ha T. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a) // *Proc. Natl. Acad. Sci. U. S. A.* – 2018. – V. 115. – № 21. – P. 5444–5449.

111. Teng F., Cui T., Feng G., Guo L., Xu K., Gao Q., Li T., Li J., Zhou Q., Li W. Repurposing CRISPR-Cas12b for mammalian genome engineering // *Cell Discov.* – 2018. – V. 4. – № 1. – P. 1-15.
112. Gao L., Cox D.B.T., Yan W.X., Manteiga J.C., Schneider M.W., Yamano T., Nishimasu H., Nureki O., Crosetto N., Zhang F. Engineered Cpf1 variants with altered PAM specificities // *Nat. Biotechnol.* – 2017. – V. 35. – № 8. – P. 789–792.
113. Kleinstiver B.P., Sousa A.A., Walton R.T., Tak Y.E., Hsu J.Y., Clement K., Welch M.M., Horng J.E., Malagon-Lopez J., Scarfò I., Maus M. V., Pinello L., Aryee M.J., Joung J.K. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing // *Nat. Biotechnol.* – 2019. – V. 37. – № 3. – P. 276–282.
114. Zhang L., Zuris J.A., Viswanathan R., Edelstein J.N., Turk R., Thommandru B., Rube H.T., Glenn S.E., Collingwood M.A., Bode N.M., Beaudoin S.F., Lele S., Scott S.N., Wasko K.M., Sexton S., Borges C.M., Schubert M.S., Kurgan G.L., McNeill M.S., ... Vakulskas C.A. AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines // *Nat. Commun.* – 2021. – V. 12. – № 1. – P. 1-15.
115. Tóth E., Czene B.C., Kulcsár P.I., Krausz S.L., Tálas A., Nyeste A., Varga É., Huszár K., Weinhardt N., Ligeti Z., Borsy A., Fodor E., Welker E. Mb- And FnCpf1 nucleases are active in mammalian cells: Activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants // *Nucleic Acids Res.* – 2018. – V. 46. – № 19. – P. 10272–10285.
116. Kissling L., Monfort A., Swarts D.C., Wutz A., Jinek M. Preparation and electroporation of Cas12a/Cpf1-guide RNA complexes for introducing large gene deletions in mouse embryonic stem cells // *Methods Enzymol.* – 2019. – V. 616. – P. 241–263.

ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

Приложение 1. Связь RbCas12a с другими ортологами Cas12a. А) Показаны множественные выравнивания последовательности RbCpf1 с известными ортологами Cas12a, которые функционируют в клетках человека. Каталитические Asp1194, Glu1290 остатки являются консервативными (нумерация для целого выравнивания). Полные названия микроорганизмов, по которым даны сокращенные названия Cas12a приведены в Приложении 2. Адаптировано из [109]. (Б) Реальное филогенетическое дерево (real tree), построенное по методу ближайших соседей без поправок на расстояния.



Приложение 2. Ортологи Cas12a с подтвержденной активностью в отношении генома млекопитающих и последовательности скаффолдов их crRNA.

Хозяин	Название	Скаффолд, 5'→3'
<i>Lachnospira eligens</i> ATCC 27750	EeCas12a	UAAUUUCUACU UUGUAGAU
<i>Moraxella bovoculi</i> 237	MbCas12a	UAAUUUCUACU UUUGUAGAU
<i>Eubacterium rectale</i>	ErCas12a	UAAUUUCUACU CUUGUAGAU

<i>Acidaminococcus sp</i>	AsCas12a	UAAUUUCUACU CUU GUAGAU
<i>Ruminococcus bromii sp.</i>	RbCas12a	UAAUUUCUACU AUU GUAGAU
<i>Butyrivibrio sp. NC3005</i>	BsCas12a	UAAUUUCUACU AUU GUAGAU
<i>Lachnospiraceae bacterium MA2020</i>	Lb2Cas12a	UAAUUUCUACU AUU GUAGAU
<i>Helcococcus kunzii ATCC 51366</i>	HkCas12a	UAAUUUCUACU AUU GUAGAU
Pseudobutyrvibrio xylanivorans strain DSM 10317	PxCas12a	UAAUUUCUACU AUU GUAGAU
<i>Francisella novicida</i>	FnCas12a	UAAUUUCUACU GUU GUAGAU
Thiomicrospira sp. XS5	TsCas12a	UAAUUUCUACU GUU GUAGAU
<i>Agathobacter rectalis strain 2789STDY5834884</i>	ArCas12a	UAAUUUCUACU GUU GUAGAU
<i>Butyrivibrio fibrisolvens MD2001</i>	BfCas12a	UAAUUUCUACU GUU GUAGAU
<i>Lachnospiraceae bacterium ND2006</i>	LbCas12a	UAAUUUCUACU UAAGU GUAGAU
<i>Moraxella bovoculi AAX08_00205</i>	Mb2Cas12a	UAAUUUCUACU UGUUU GUAGAU
<i>Moraxella bovoculi sp.</i>	Mb3Cas12a	UAAUUUCUACU UGUUU GUAGAU
<i>Coprococcus eutactus sp.</i>	CeCas12a	UAAUUUCUACU UCG GUAGAU
Lachnospira pectinoschiza strain 2789STDY5834886	LpCas12a	UAAUUUCUACU UGUGU GUAGAU
<i>Pseudobutyrvibrio ruminis CF1b</i>	PrCas12a	UAAUUUCUACU UGUGU GUAGAU

Приложение 3. Использованные в работе последовательности, кодирующие RbCas12a

Название	Последовательность 5'->3'
----------	---------------------------

Нуклеотидная последовательность RbCas12a

ATGCAAGAGCGTAAAAAATATCGCATCT
TACACACAGAAATTCAGTTAAAAAACA
ATTAGGATGCAACTGAATCCTGTAGGTAA
ACAATGGATTATTTTCAAGCAAAGCAA
ATTCTTGAAAATGATGAAAAGCTTAAAG
AGAACTATCAGAAAATCAAGGAAATAGC
AGACAGGTTTTACAGAAATTTAAATGAG
GATGTACTTTCAAAAACCGGGTTAGATAA
ATTTAAAGATTATGCTGAAATTTACTATCA
CTGTAATACGGATGCAGACCGAAAAAGA
CTTGATGAATGTGCATCGGAATTAAGGAA
AGAAATCGTCAAAAATTTTAAGAATAGA
GATGAGTATAACAACTATTCGATAAAAG
GATGATTGAGATAGTTCTTCCCCAGCATC
TTAAAAACGAGGACGAAAAGGAAGTTGT
AGCCTCATTAAAAATTTCAACAATACT
TTACAGGTTTTCTTACTAACAGAAAAAT
ATGTATTCGGACGGAGAAGAATCCACGG
CAATCGCATATAGATGCATTAACGAAAAT
TTGCCTAACATCTTGACAATGTCAAGGC
CTTTGAAAAAGCAATTTCTAACTATCCA
AAAACGCAATTGATGATTTAGATGCCACT
TATTCCGGCTTATGCGGTACAAATTTGTAT
GATGTTTTTACAGTTGATTATTTAACTTT
TTGCTTCCACAATCCGGAATTACCGAATA
TAACAAAATCATCGGCGGTTACACAACA
AGCGACGGTACAAAAGTTAAGGGTATTA
ACGAATATATAAATTTGTACAATCAACAA
GTATCCAAACGGGATAAAATTCCTAATCT
TCAAATTTTGTATAAACAAATTTTAAGTG
AGAGTGAAAAGGTATCATTACATCCGCC
AAAGTTTGAAGATGACAACGAACTTTTA
TCGGCTGTTTCAGAGTTTTACGCAAACG
ACGAAACCTTTGACGGGATGCCATTA
AAAAGCAATTGATGAAACAAAGCTATTAT
TCGGCAATTTAGATAATTCCTCTCTAATG
GAATTTACATTCAAATGACCGATCCGTG
ACAAATCTGTCAAACAGTATGTTCCGGTTC
TTGGTCGGTAATAGAAGATTTATGGAACA
AAAATTATGACTCCGTTAATTCAAACAGC
AGAATCAAAGATATTCAAAGCGTGAAG
ACAAAAGAAAAAAGCATACAAAGCAG
AAAAGAACTTTCACTTTCATTTTTACAG
GTTTTGATTTCCAATCCGAAAATGATGA
AATCAGAAAAAAGTCTATCGTAGATTACT
ACAAGACTTCTTTAATGCAACTTACCAAC
AATTTATCAGACAAATACAACGAAGCAG
CACCTCTGTTTCAGTGAAAATTACGATAAT
GAAAAGGTTTGAAAATGACGATAAAT
CTATTTCATTAATAAAAATTTTCTTGATG

CCATAAAAGAAATTGAAAAATTCATAAA
GCCTTTGTCCGAACTAATATTACAGGTG
AGAAAAATGATTTGTTTTACAGTCAGTTC
ACACCATTACTTGATAATATCAGCAGAAT
AGACATATTATATGATAAGGTCAGAACT
ATGTTACACAAAAACCGTTTTCAACCGAT
AAAATCAAGCTTAACTTTGACAATTACCA
GCTATTAACGGCTGGGATAAAGACAAA
GAAAGAGAGTACGGAGCCGTTTTGCTTT
GTAAAGATGAAAAGTATTATCTTGCAATC
ATAGATAAAAGCAATAATCGTATTTTGG
AAATATTGATTTTCAAGACTGCGATGAAA
GCGATTGTTACGAAAAGATAATTTACAAG
CTTCTCCCCACTCCAAATAAAATGCTTCC
AAAAGTTTTCTTTGCAAAAAAGCACAAA
AACTTTTGTACCGTCAGACGAAATAC
TTAAAATTTATAAAAGCGGCACTTTCAA
AAAGGTGATAAGTTCAGCCTTGATGATTG
CCATAAGTTAATTGATTTCTACAAAGAAT
CATTCAAAAAGTACCCAAAATGGTTAATT
TATAACTTTAAATTCAAAAAACAAACG
AATATAACGATATCCGCGAATTTTATAATG
ATGTTGCTTTACAGGGATATAATTTCAA
AAATGAAAATCCCGACATCATTATTGAC
AACTTGTAGATGAAGGAAAAATCTATCT
TTTCCAACCTCTACAACAAAGACTTTTCAC
CGCATAGCAAGGGTACTCCTAATCTGCAT
ACACTTTATTTTAAAATGTTATTTGATGAA
AGAAATCTTGAAGATGTGGTGTATAGGCT
TAACGGTGAGGCAGAAATGTTTTATCGTC
CTGCAAGTATAAAATATGACAAACCCACT
CATCCCAAAAACACACCGATAAAAAATA
AAAATACACTCAATGATAAAAAAGCAAG
CACTTTTCCTTATGACTTAATTAAGATAA
ACGCTACACTAAATGGCAGTTTTCACTTC
ACTTCCCTATTACCATGAATTTTAAAGCT
CCGGATAGGGCAATGATCAATGATGATGT
CAGAAATCTGCTGAAATCCTGCAACAAC
AATTCATCATAGGAATTGACAGAGGCG
AAAGAACTTGCTTTATGTCAGCGTAATT
GACAGCAACGGTGCTATAATATATCAGCA
CTCACTCAATATTATCGGAAACAAGTTTA
AAGGAAAACATACGAACTAACTACCA
GGAAAACTTGCAACAAGAGAAAAAGA
GCGTACGGAACAGCGCCGTAAGTGGAAA
GCAATTGAGAGTATAAAAGAACTCAAAG
AGGGCTATATCAGTCAGGCTGTGCATGTT
ATATGTCAGCTTGTTGTCAAGTACGATGC
AATCATCGTTATGGAAAAGCTGACTGAC
GGATTCAAACGAGGCAGAACAAGTTTG

AAAAACAGGTTTATCAGAAATTTGAAAA
AATGCTGATTGACAACTTAATTACTATG
TTGACAAAAAGCTTGATCCCGATGAAGA
AGGCGGTTTACTTCATGCCTACCAGCTTA
CGAACAAGCTTGAGAGCTTTGATAAGCT
TGGTACGCAAAGCGGTTTTATTTTCTATG
TTCGTCTGATTTTACAAGCAAATTGAT
CCCGTTACCGGCTTTGTAAATTTGTTGTA
CCCTCGATATGAAAACATTGACAAAGCC
AAAGATATGATTTCAAGATTTGACGATAT
AAGATACAATGCCGGCGAGGACTTTTTT
GAATTTGACATTGATTACGATAAGTTTCC
AAAGACTGCGTCTGACTATCGCAAAAAG
TGGACAATCTGTACTAACGGCGAAAGGA
TTGAAGCTTTCAGAAATCCCGCAAACAA
TAACGAATGGAGTTATCGTACAATAATTC
TTGCAGAAAAATTCAAAGAATTATTTGAT
ACAATTCTATAAATTATCGTGATTCTGAC
GATTTGAAAGCTGAAATTCCTTTCACAGA
CAAAGGGCAAATTTTTTGAGGATTTCTTC
AAATTATTAAGACTTACCCTACAGATGCG
AAACAGTAACCCTGAAACAGGCGAGGA
CCGTATTCTTTCTCCCGTCAAGGACAAAA
ACGGCAATTTTTACGACAGTTCAAATAT
GATGAAAAGAGCAAGCTTCCGTGTGACG
CCGATGCAAACGGTGCGTACAACATTGC
CCGCAAAGGTTTGTGGATTGTTGAACAA
TTCAAAAATCCGATAATGTTTCAACTGT
CGAACCGGTAATTCACAATGACAAATGG
CTGAAATTTGTTTCAGGAGAATGATATGGC
GAATAATCTCGAG

<p>Аминокислотная последовательность RbCas12a</p>	<p>MQERKKISHLTHRNSVKKTIRMQLNPV GK TMDYFQAKQILENDEKLKENYQKIKEIAD RFYRNLNEDVLSKTGLDKLKDYAEIYYHC NTDADRKRLDECASELRKEIVKNFKNRDE YNKLFDKRMIEIVLPQHLKNEDEKEVVAS FKNFTTYFTGFFTNRKNMYS DGEESTAIA Y RCINENLPKHLDNVKA FEKAISKLSKNAID DLDATYSGLCGTNLYDVFTVDYFNFLLPQ SGITEYNKIIGGYTTSDGTKVKGINEYINLY NQQVSKRDKIPNLQILYKQILSESEKVSFIP PKFEDDNELLSAVSEFYANDETFDGMPLK KAIDETKLLFGNLDNSSLNGIYIQNDRSVT NLSNSMFGSWSVIEDLWNKNYDSVNSNSR IKDIQKREDKRKKAYKAEKKLSLSFLQVLI SNSENDEIRKKSIVDYKTSMLQLTNNLSD KYNEAAPLFSENYDNEKGLKNDDKSISLIK NFLDAIKEIEKFIKPLSETNITGEKNDFYS QFTPLLDNISRIDILYDKVRNYVTQKPFSTD KIKLNF DNYQLLNGW DDKEREY GAVLL CKDEKY YLAIDKSNRILENIDFQDCDES DCYEKIIYKLLPTPNKMLPKVFFAKKHKK LLSPSDEILKIYKSGTFKKGDKFSLDDCHK LIDFYKESFKKYPKWLIYNFKFKKTNEYN DIREFYNDVALQGYNISKMKIPTSFIDKLV DEGKIYLFQLYNKDFSPHSGT PNLHTLYF KMLFDERNLEDVVYRLNGEAEMFYRPASI KYDKPTHPKNTPIKNKNTLNDDKASTFPY DLIKDKRYTKWQFSLHFPITMNFKAPDRA MINDDVRNLLKSCNNNFII GIDRGERNLLY VSVIDSNGAIIYQHSLNII GNKFKGKTYETN YQEKLATREKERTEQRRNWKAIESIKELKE GYISQAVHVICQLVVKYDAIIVMEKLT DGF KRGRTKFEKQVYQKFEKMLIDKLNYYVD KKLDPDEEGLLHAYQLTNKLESFDKLTGT QSGFIFYVRP DFTSKIDPVTGFVNLLYPRYE NIDKAKDMISRFD DIRYNAGEDFFFEFDIDY DKFPKTASDYRKKWTICTNGERIEAFRNPA NNNEWSYRTIILA EKFKELFDNNSIN YRDS DDLKAEILSQT KGKFFEDFFKLLRLTLQMR NSNPETGEDRILSPVKDKNGNFYDSSKYD EKSKLPCDADANGAYNIARKGLWIVEQFK KSDNVSTVEPVIHNDKWLK FVQENDMAN NLE</p>
<p>Нуклеотидная последовательность кодон- оптимизированного RbCas12a с тремя сигналами ядерной локализации на N-конце</p>	<p>ATGCCGAAGAAAAAGCGCAAGGTCAAG CTCCGAAGAAAAAGCGCAAGGTCAAG CTCCGAAGAAAAAGCGCAAGGTCAAG</p>

(3xSV40 NLS — выделены красным) и
сигналом ядерной локализации на С-конце
(nucleoplasmin NLS — выделен синим)

CTTATGCAAGAACGCAAGAAGATTAGTC
ATCTGACCCATAGAAACTCCGTGAAGAA
GACCATCCGTATGCAATTAACCCGGTCCG
GCAAGACCATGGACTACTTCCAAGCCAA
GCAAATCCTGGAGAACGACGAGAAGCT
GAAGGAAAAC TACCAAAGATCAAGGA
GATTGCCGACCGTTTCTATAGAACTTAA
ACGAAGACGTCCTGAGCAAGACTGGTTT
GGACAAGTTAAAGGACTACGCTGAGATC
TATTACCATTGTAACACGGACGCCGACCG
TAAGAGACTGGACGAGTGTGCCAGTGAG
TTACGTAAGGAGATCGTAAAGA ACTTCA
AGAACAGAGACGAATACAACAAGCTATT
TGACAAGCGTATGATCGAAATTGTGCTGC
CGCAACATCTGAAGAACGAAGACGAGA
AGGAGGTGGTCGCGTCTTTCAAGAACTT
TACCACCTACTTCACCGGCTTTTTTACTA
ACAGAAAGAACATGTACAGTGACGGAG
AGGAGTCGACGGCCATCGCCTACAGATG
TATCAACGAGAACTTACCGAAGCATCTG
GACAACGTAAAGGCATTTCGAGAAGGCTA
TCAGTAAGCTATCGAAGAACGCCATCGA
CGACTTAGACGCAACTTACTCGGGCTTAT
GTGGAACCAACTTGTACGACGTGTTTAC
CGTGGACTACTTCAACTTCTTGCTGCCGC
AATCGGGAATCACTGAGTACAACAAGAT
CATCGGCGGCTACACCACCTCTGACGGC
ACCAAGGTGAAAGGTATCAACGAGTACA
TAACTTGTACAACCAACAAGTCTCGAA
GCGTGACAAGATCCCGAACCTGCAAATC
TTGTACAAGCAAATCTTAAGTGAAAGTG
AGAAGGTCAGCTTTATTCCACCGAAGTT
CGAGGACGACAACGAGCTGTTAAGTGCT
GTGAGCGAATTTTACGCCAACGACGAGA
CTTTCGACGGTATGCCGTTAAAGAAGGC
CATCGACGAGACCAAGCTATTATTTGGCA
ACTTAGACA ACTCGAGTCTGAACGGAAT
CTACATCCAAAACGACCGTTCGGTTACC
AACTTAAGCAACAGTATGTTTGGCAGTT
GGAGTGTGATTGAGGACTTATGGAACAA
GAACTACGACTCGGTGAACTCTAACTCT
AGAATCAAGGACATCCAAAAGCGCGAG
GACAAGAGAAAGAAGGCCTACAAGGCC
GAGAAGAAGCTGAGCCTGAGCTTCTTAC
AAGTGTTGATCTCGAACTCGGAGAACGA
CGAGATCAGAAAGAAGAGTATCGTTCGAC
TACTACAAGACTAGTTTAATGCAACTGAC
TAACA ACTTAAGCGACAAGTACAACGAG
GCCGCCCGTTATTTAGTGAGA ACTACGA
CAACGAGAAGGGCTTGAAGAACGACGA

CAAGAGTATCAGCTTAATCAAGAACTTCC
TGGACGCAATTAAGGAGATCGAGAAGTT
TATTAAGCCGTTGTCGGAGACTAACATCA
CCGGCGAAAAGAACGACTTGTTCTACAG
TCAATTTACCCCGTACTGGACAACATCT
CTAGAATTGACATTTTATACGACAAGGTA
AGAACTACGTGACCCAAAAGCCATTCA
GCACTGACAAGATCAAGCTGAACTTCGA
CAACTACCAACTATTAACGGCTGGGAC
AAGGACAAGGAGAGAGAATACGGAGCG
GTGTTGCTGTGTAAGGACGAGAAGTACT
ACCTGGCCATCATTGACAAGTCTAACAA
CCGCATCTTGGAGAACATCGACTTCCAA
GACTGTGACGAGTCTGACTGTTACGAGA
AGATTATCTACAAGCTGCTTCCGACTCCG
ACAAGATGCTGCCGAAGGTGTTTTTCG
CCAAGAAGCATAAGAAGCTGTTGAGCCC
AAGCGACGAGATTCTGAAGATCTACAAG
TCTGGCACTTTTAAGAAGGGCGACAAGT
TTTCTCTGGACGACTGTCATAAGTTAATC
GACTTTTACAAGGAGAGCTTTAAGAAGT
ACCCGAAGTGGTTAATCTACAACCTCAA
GTTTAAGAAGACCAACGAGTACAACGAC
ATCCGTGAGTTCTACAACGACGTGGCTTT
ACAAGGTTACAACATCAGCAAGATGAAG
ATCCCGACCTCTTTCATCGACAAGCTGGT
CGACGAGGGAAAGATCTACCTGTTTCAA
CTTTACAACAAGGACTTCAGCCCACATT
CTAAGGGCACTCCGAACCTTACATACCCTG
TACTTCAAGATGTTATTCGACGAGAGAA
ACCTGGAGGACGTTGTTTACCGTCTGAA
CGGCGAAGCCGAGATGTTCTACCGCCCCG
GCCAGTATTAAGTACGACAAGCCGACTC
ATCCGAAGAACACCCCAATTAAGAACAA
GAACACCCTTAACGACAAGAAGGCCTCT
ACTTTCCCGTACGACTTAATCAAGGACA
AGCGTTACACTAAGTGGCAATTCAGCCT
GCATTTTCCGATCACTATGAACTTCAAGG
CTCCAGACCGTGCCATGATTAACGACGA
CGTAAGAACTTATTAAGTCGTGTAACA
ACAACCTTATCATTGGAATCGACAGAGG
CGAGAGAACTTGCTGTACGTATCTGTCA
TCGACTCTAACGGCGCTATTATTTACCAA
CATAGCCTTAACATCATCGGAAACAAGTT
CAAGGGAAAGACCTACGAGACTAACTAC
CAAGAGAAGCTGGCCACCAGAGAGAAG
GAACGCACGGAGCAACGTCGCAACTGG
AAGGCCATCGAAAGTATTAAGGAGCTTA
AGGAAGGCTACATCAGTCAAGCTGTTCA
TGTGATTTGTCAACTGGTGGTAAAGTACG

ACGCCATCATCGTGATGGAGAAGTTAAC
TGACGGATTTAAGCGTGGCAGAACCAAG
TTCGAGAAGCAAGTGTACCAAAGTTTCG
AGAAGATGTTAATCGACAAGCTGAACTA
CTACGTGGACAAGAAGCTGGACCCGGAC
GAGGAGGGCGGCTTACTGCATGCATACC
AACTGACGAACAAGCTGGAATCTTTCGA
CAAGCTGGGCACGCAATCTGGCTTCATC
TTTTACGTGCGCCCGGACTTCACCTCTAA
GATCGACCCGGTGACTGGCTTCGTCAAC
TTGTTGTACCCGCGTTACGAGAACATCGA
CAAGGCAAAGGACATGATCAGCAGATTC
GACGACATTAGATAACAACGCAGGCGAAG
ACTTCTTCGAGTTCGACATCGACTACGAC
AAGTTCCCGAAGACTGCGAGTGACTACC
GTAAGAAGTGGACCATCTGTACTAACGG
CGAGCGTATCGAGGCTTTTAGAAACCCG
GCCAACAAACGAGTGGAGTTACCGCA
CCATTATCCTGGCCGAGAAGTTTAAGGA
GTTATTCGACAACAACAGTATTAACTACC
GCGACAGTGACGACTTGAAGGCTGAGAT
CCTGAGCCAAACCAAGGGCAAGTTCTTC
GAAGACTTTTTTAAGTTATTAAGACTGAC
TCTACAAATGCGTAACAGTAACCCGGAG
ACCGGGCAAGACCGCATCCTGAGTCCGG
TAAAGGACAAGAACGGAACTTCTACGA
CAGTAGCAAGTACGACGAGAAGTCTAAG
CTGCCATGTGACGCAGACGCCAACGGCG
CGTACAACATCGCACGTAAGGGCTTGTG
GATCGTGGAGCAATTTAAGAAGTCGGAC
AACGTGAGCACTGTAGAGCCAGTCATCC
ATAACGACAAGTGGTTAAAGTTCGTGCA
AGAAAACGACATGGCGAACAACAAGCG
GCCTGCGGCTACTAAGAAGGCTGGGCAG
GCTAAGAAGAAGAAG