

(19)



**Евразийское  
патентное  
ведомство**

(21) **201991913** (13) **A1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки  
2020.01.22

(51) Int. Cl. *C12N 9/22* (2006.01)

(22) Дата подачи заявки  
2018.02.14

---

(54) **ВЫСОКОТОЧНЫЕ ВАРИАНТЫ Cas9 И ИХ ПРИМЕНЕНИЯ**

---

(31) 102017000016321

(32) 2017.02.14

(33) IT

(86) PCT/EP2018/053717

(87) WO 2018/149888 2018.08.23

(71) Заявитель:

**УНИВЕРСИТА' ДЕЛЪИ СТУДИ ДИ  
ТRENTO (IT)**

(72) Изобретатель:

**Черезето Анна, Казини Антонио,  
Петрис Джанлука, Инга Альберто,  
Оливьери Микеле (IT)**

(74) Представитель:

**Медведев В.Н. (RU)**

(57) Для того чтобы обратиться к ограничениям, возникающим из неспецифического расщепления генома Cas9 *Streptococcus pyogenes* (SpCas9), и идентифицировать варианты с более высокой точностью расщепления, данное изобретение описывает анализ на основе дрожжей, который позволяет одновременно оценивать активность в и вне мишени, направленную на две сконструированные геномные мишени. Скрининг вариантов SpCas9, полученных путем случайного мутагенеза домена Rec1-II, делал возможной идентификацию попаданий с увеличенными соотношениями в/вне мишени. Наиболее эффективную нуклеазу evoCas9 выделяли через комбинацию идентифицированных мутаций в одном варианте. Сравнительные анализы с использованием рационально сконструированных вариантов, о которых сообщали ранее, демонстрировали значительное усовершенствование точности evoCas9 по данному изобретению.

**A1**

**201991913**

**201991913**

**A1**

## ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-558319EA/019

### ВЫСОКОТОЧНЫЕ ВАРИАНТЫ CAS9 И ИХ ПРИМЕНЕНИЯ

#### ОБЛАСТЬ ИЗОБРЕТЕНИЯ

Настоящее изобретение относится к области ферментов и связывающих нуклеиновые кислоты белков, в частности к вариантам Cas9.

#### СУЩЕСТВУЮЩИЙ УРОВЕНЬ ТЕХНИКИ

В последние годы наблюдают существенный рост биотехнологических применений с использованием системы CRISPR-Cas9, который обусловлен гибкостью и эффектом этого нового инструмента редактирования генома. Cas9 обычно распознает целевые участки через так называемую последовательность «гидовой РНК» (гРНК), комплементарную целевой нуклеиновой кислоте, которая содержит последовательность протоспейсера. Для распознавания мишени также необходимо присутствие короткой соседней последовательности PAM (прилегающие к протоспейсеру мотивы). Целевая нуклеиновая кислота обычно представляет собой ДНК, но в некоторых обстоятельствах также может представлять собой РНК. Гидовые РНК можно формировать с помощью одной или нескольких малых РНК. Редактирование генома с помощью подхода CRISPR-Cas9 успешно применяли к различным типам клеток и видам, ясно демонстрируя эффективность и устойчивость, которые должны характеризовать инновационную технологию. Что важно, как фундаментальные исследовательские, так и терапевтически ориентированные применения, помимо эффекта, требуют высокой специфичности направленного воздействия для редактирования. Однако некоторые исследования показали, что Cas9 расщепление генома не всегда направлено на предполагаемые участки и нежелательные повреждения можно вводить в области ДНК, обладающие различными уровнями сходства с выбранной мишенью. Кроме того, предсказание такой нежелательной активности является сложным и часто ненадежным из-за отсутствия простых правил, определяющих феномен. Кроме того, оценка эффектов вне мишени не всегда является простой, а получаемые результаты применения различных способов часто не находятся в соответствии. Таким образом, увеличение специфичности арсенала CRISPR-Cas9 является очень желательным усовершенствованием этой ключевой технологии, допускающим ее безопасное применение во всех областях применения, в частности, в терапевтических применениях у

человека.

Различные стратегии предложены для того, чтобы снижать введение нежелательных мутаций вне мишени с помощью CRISPR-Cas9, такие как строгий контроль внутриклеточных уровней Cas9 *Streptococcus pyogenes* (SpCas9), введение сконструированных гРНК, отличающихся более короткими протоспейсерами с меньшей комплементарностью с целевой последовательностью (усеченная гРНК), слияние SpCas9 со специфическими ДНК-связывающими доменами, чтобы направлять его связывание, или эксплуатация спаренных нуклеаз SpCas9 и спаренных каталитически неактивных SpCas9, слитых с доменом эндонуклеазы FokI. Однако ни один из этих подходов не свободен от эффектов вне мишени, и из-за свойственной им молекулярной сложности часто они имеют дефектную активность в мишени.

В последнее время, две группы сообщали о рациональном структурном конструировании вариантов SpCas9, отличающихся более низкой предрасположенностью к расщеплению участков вне мишени.

Slaymaker IM et al. (Science. 2016, 351(6268):84-8) создавали три мутанта SpCas9 как с высокой эффективностью (близкой к уровням дикого типа для инсерций-делеций в мишени, инсерция-делеция, формирование) и специфичностью (не поддающееся обнаружению формирование инсерций-делеций в участках EMX(1) и VEGFA(1) вне мишени, стандартные локусы для оценки специфичности): SpCas9 (K855A), SpCas9 (K810A/K1003A/R1060A) [также обозначаемый как eSpCas9(1.0)] и SpCas9 (K848A/K1003A/R1060A) [также обозначаемый как eSpCas9(1.1)].

Kleinstiver BP et al. (Nature. 2016, 529(7587):490-5) создавали 15 различных вариантов SpCas9, несущих все возможные одинарные, двойные, тройные и четверные комбинации замен N497A, R661A, Q695A и Q926A. Как тройной мутантный вариант (R661A/Q695A/Q926A), так и вариант с четверной заменой (N497A/R661A/Q695A/Q926A, далее обозначаемый как SpCas9-HF1) демонстрировали минимальное разрушение EGFP на близких к фону уровнях с использованием четырех несовпадающих sgRNA.

Также из этих недавних попыток очевидно, что основная потребность в этой области состоит в создании систем редактирования генома без активности вне мишени.

Цель настоящего изобретения состоит в том, чтобы предоставить новые, по меньшей мере альтернативные, высокоточные варианты Cas9.

## СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Объект изобретения по настоящему изобретению представляет собой выделенную модифицированную молекулу Cas9, содержащую по меньшей мере одну мутацию, расположенную в положениях аминокислотных остатков, выбранных в группе, состоящей из: K377, E387, D397, R400, D406, A421, L423, R424, Q426, Y430, K442, P449, V452, A456, R457, W464, M465, K468, E470, T474, P475, W476, F478, K484, S487, A488, T496, F498, L502, N504, K506, P509, F518, N522, E523, K526, L540, S541, I548, D550, F553, V561, K562, E573, A589, L598, D605, L607, N609, N612, E617, D618, D628, R629, R635, K637, L651, K652, R654, T657, G658, L666, K673, S675C, I679V, L680, L683, N690, R691, N692, F693, S701, F704, Q712, G715, Q716, H723, I724, L727, I733, L738 и Q739;

где положение модифицированной аминокислотной последовательности идентифицируют посредством обращения к нумерации аминокислот в соответствующем положении немодифицированного зрелого Cas9 *Streptococcus pyogenes* (SpCas9), как идентифицируют с помощью SEQ ID NO: 1.

В предпочтительном варианте осуществления модифицированный Cas9 содержит по меньшей мере одну мутацию в положении K526.

Варианты SpCas9 в соответствии с изобретением изначально получали посредством случайного мутагенеза его домена REC1-II и осуществляли скрининг для идентификации попаданий с увеличенными соотношениями в/вне мишени с помощью анализа на основе дрожжей, который позволяет одновременно оценивать активность в и вне мишени относительно двух сконструированных геномных мишеней. После дополнительной валидации в клетках млекопитающих, создавали варианты Cas9 в соответствии с изобретением. К удивлению, модифицированный SpCas9 в соответствии с изобретением демонстрировал значительно сниженную активность вне мишени по сравнению с SpCas9 дикого типа и сравнительные анализы с использованием рационально разработанных вариантов, о которых сообщали ранее, демонстрировал значительное усовершенствование в точности варианта SpCas9 по изобретению. Дополнительно, модифицированный SpCas9 в соответствии с изобретением, имеющий дополнительный мутации D10A и H840A, слитый с доменом транскрипционной активации (VP64), демонстрировал значительно сниженную активность вне мишени по сравнению с вариантом Cas9 дикого типа, содержащим мутации D10A и H840A.

## ПОДРОБНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

Настоящее изобретение описывает выделенные молекулы Cas9 с увеличенной специфичностью, получаемые через случайный мутагенез домена REC1-II в Cas9 и скрининг с использованием анализа на основе дрожжей, чтобы оценивать одновременно активность в и вне мишени у каждого созданного варианта. Выбранные попадания дополнительно уточняли посредством скрининга в системе млекопитающего.

В первом аспекте изобретения варианты Cas9 содержат по меньшей мере одну мутацию, расположенную в положениях аминокислотных остатков, выбранных в группе, состоящей из K526, K377, E387, D397, R400, D406, A421, L423, R424, Q426, Y430, K442, P449, V452, A456, R457, W464, M465, K468, E470, T474, P475, W476, F478, K484, S487, A488, T496, F498, L502, N504, K506, P509, F518, N522, E523, L540, S541, I548, D550, F553, V561, K562, E573, A589, L598, D605, L607, N609, N612, E617, D618, D628, R629, R635, K637, L651, K652, R654, T657, G658, L666, K673, S675, I679, L680, L683, N690, R691, N692, F693, S701, F704, Q712, G715, Q716, H723, I724, L727, I733, L738 и Q739;

где положение модифицированной аминокислотной последовательности идентифицируют посредством обращения к нумерации аминокислот в соответствующем положении немодифицированного зрелого Cas9 *Streptococcus pyogenes* (SpCas9), как идентифицируют с помощью SEQ ID NO: 1.

Предпочтительно в соответствии с изобретением мутацию в положении K526 выбирают в группе, состоящей из K526N и K526E; более предпочтительно K526E.

В соответствии с предпочтительным вариантом осуществления изобретения модифицированный Cas9, имеющий мутантный K526, содержит одну или несколько дополнительных мутаций, расположенных в положениях аминокислотных остатков, выбранных в группе, состоящей из:

K377, E387, D397, R400, Q402, R403, F405, D406, N407, A421, L423, R424, Q426, Y430, K442, P449, Y450, V452, A456, R457, W464, M465, K468, E470, T472, I473, T474, P475, W476, F478, K484, S487, A488, M495, T496, N497, F498, L502, N504, K506, P509, Y515, F518, N522, E523, L540, S541, I548, D550, F553, V561, K562, E573, A589, L598, D605, L607, N609, N612, E617, D618, D628, R629, R635, K637, L651, K652, R654, T657, G658,

W659, R661, L666, K673, S675, I679, L680, L683, N690, R691, N692, F693, Q695, H698, S701, F704, Q712, G715, Q716, H721, H723, I724, L727, A728, I733, L738, Q739.

Одна или несколько дополнительных мутаций составляют число, находящееся между 1 и 8.

Предпочтительно одну или несколько дополнительных мутаций выбирают в группе, состоящей из:

K377E, E387V, D397E, R400H, Q402R, R403H, F405L, D406Y, D406V, N407P, N407H, A421V, L423P, R424G, Q426R, Y430C, K442N, P449S, Y450A, Y450S, Y450H, Y450N, V452I, A456T, R457P, R457Q, W464L, M465R, K468N, E470D, T472A, I473F, I473V, T474A, P475H, W476R, F478Y, F478V, K484M, S487Y, A488V, M495V, M495T, T496A, N497A, F498I, F498Y, L502P, N504S, K506N, P509L, Y515N, F518L, F518I, N522K, N522I, E523K, E523D, L540Q, S541P, I548V, D550N, F553L, V561M, V561A, K562E, E573D, A589T, L598P, D605V, L607P, N609D, N609S, N612Y, N612K, E617K, D618N, D628G, R629G, R635G, K637N, L651P, L651H, K652E, R654H, T657A, G658E, W659R, R661A, R661W, R661L, R661Q, R661S, L666P, K673M, S675C, I679V, L680P, L683P, N690I, R691Q, R691L, N692I, F693Y, Q695A, Q695H, Q695L, H698Q, H698P, S701F, F704S, Q712R, G715S, Q716H, H721R, H723L, I724V, L727H, A728G, A728T, I733V, L738P, Q739E, Q739P и Q739K.

В предпочтительных вариантах осуществления общее число вышеуказанных мутаций составляет между 1 и 9; предпочтительно между 1 и 5; наиболее предпочтительно между 1 и 4.

SEQ ID NO: 1 представляет собой последовательность, имеющую номер доступа NP\_269215 (NCBI), относящийся к SpCas9.

В соответствии с изобретением, модифицированный полипептид, за исключением мутаций, предпочтительно обладает идентичностью по меньшей мере 50%, 60%, 70%, 80%, 90%, 95%, 99% или 100% с SEQ ID NO: 1.

Процент идентичности между двумя полипептидами или последовательностями нуклеиновой кислоты могут определять специалисты в данной области с использованием программного обеспечения для выравнивания (т. е. программы BLAST).

Предпочтительно модифицированный Cas9 представляет собой Cas9 *S. pyogenes*. В некоторых вариантах осуществления Cas9 представляет собой ортологи SpCas9 (т. е. *S. thermophilus*, *S. aureus*, *N. meningitides*). В некоторых вариантах осуществления ортолог Cas9 обладает по меньшей мере 10% или 25% аминокислотной идентичностью с домен Rec1-II в SpCas9 и полной аминокислотной

идентичностью с любой процентной долей между 10% или 25% и 100% с SpCas9. Специалисты в данной области могут определять подходящие гомологичные остатки, подлежащие модификации, посредством выравнивания последовательностей и/или структурного выравнивания. Идентифицированные аминокислоты можно модифицировать консервативно с заменами в пределах следующих групп: глицин, аланин; валин, изолейцин, лейцин; аспарагиновая кислота, глутаминовая кислота, аспарагин, глутамин; серин, треонин; лизин, аргинин; фенилаланин, тирозин.

Модифицированный полипептид сохраняет способность взаимодействовать с гРНК и/или с целевой ДНК или РНК.

В соответствии с изобретением мутация X1nnnX2 обозначает, что в положении nnn аминокислота X2 присутствует вместо аминокислоты X1, которая присутствует в полипептиде дикого типа; так, например, K526E обозначает, что аминокислота в положении 526 соответствует глутаминовой кислоте (Glu или E) вместо аминокислоты лизина (Lys или K), которая присутствует в полипептиде дикого типа.

В соответствии с предпочтительным вариантом осуществления изобретения, модифицированный полипептид Cas9 содержит мутацию в положении K526 и одну или несколько дополнительных мутаций в положении, выбранном в группе, состоящей из Y450, M495, Y515, R661, N690, R691, Q695, H698; предпочтительно M495, Y515, R661, H698.

В соответствии с предпочтительным вариантом осуществления изобретения, по меньшей мере одну дополнительную мутацию выбирают в группе, состоящей из Y450S, M495V, Y515N, R661X, N690I, R691Q, Q695H, H698Q; предпочтительно выбирают в группе, состоящей из M495V, Y515N, K526E, R661X, H698Q; где X представляет собой аминокислоту, выбранную в группе, состоящей из L, Q и S; предпочтительно X представляет собой Q или S.

В соответствии с предпочтительным вариантом осуществления изобретения модифицированный полипептид Cas9 содержит двойную мутацию, выбранную в группе, состоящей из K526E+Y450S, K526E+M495V, K526E+Y515N, K526E+R661X, K526E+N690I, K526E+R691Q, K526E+Q695H и K526E+H698Q; где X представляет собой аминокислоту, выбранную в группе, состоящей из L, Q и S; предпочтительно X представляет собой Q или S.

В соответствии с предпочтительным вариантом осуществления изобретения модифицированный полипептид Cas9, как описано выше,

содержит тройную мутацию, выбранную в группе, состоящей из M495V+K526E+R661X, Y515N+K526E+R661X, K526E+R661X+H698Q и M495V+Y515N+K526E; где X представляет собой аминокислоту, выбранную в группе, состоящей из L, Q и S; предпочтительно X представляет собой Q или S.

В соответствии с предпочтительным вариантом осуществления изобретения, модифицированный полипептид Cas9, как описано выше, содержит четверную мутацию, выбранную в группе, состоящей из M495V+Y515N+K526E+R661X и M495V+K526E+R661X+H698Q; где X представляет собой аминокислоту, выбранную в группе, состоящей из L, Q и S; предпочтительно X представляет собой Q или S.

Наиболее предпочтительным является модифицированный полипептид Cas9, как описано выше, содержащий четверную мутацию M495V+Y515N+K526E+R661Q (далее здесь также обозначают evoCas9) или M495V+Y515N+K526E+R661S (далее здесь также обозначают evoCas9-II).

Для определенного аспекта, объект изобретения по настоящему изобретению представляет собой выделенный модифицированный полипептид Cas9, содержащий по меньшей мере одну мутацию, выбранную в группе, состоящей из K377E, E387V, D397E, R400H, Q402R, R403H, F405L, D406Y, D406V, N407P, N407H, A421V, L423P, R424G, Q426R, Y430C, K442N, P449S, Y450S, Y450H, Y450N, V452I, A456T, R457P, R457Q, W464L, M465R, K468N, E470D, T472A, I473F, I473V, T474A, P475H, W476R, F478Y, F478V, K484M, S487Y, A488V, M495V, M495T, T496A, F498I, F498Y, L502P, N504S, K506N, P509L, Y515N, F518L, F518I, N522K, N522I, E523K, E523D, K526E, K526N, L540Q, S541P, I548V, D550N, F553L, V561M, V561A, K562E, E573D, A589T, L598P, D605V, L607P, N609D, N609S, N612Y, N612K, E617K, D618N, D628G, R629G, R635G, K637N, L651P, L651H, K652E, R654H, T657A, G658E, W659R, R661W, R661L, R661Q, R661S, L666P, K673M, S675C, I679V, L680P, L683P, N690I, R691Q, R691L, N692I, F693Y, Q695H, Q695L, H698Q, H698P, S701F, F704S, Q712R, G715S, Q716H, H721R, H723L, I724V, L727H, A728G, A728T, I733V, L738P, Q739E, Q739P и Q739K.

В соответствии с предпочтительным вариантом осуществления изобретение относится к модифицированному полипептиду Cas9, содержащем:

одну мутацию, выбранную в группе, состоящей из D406Y, W464L, T474A, K526E, N612K, L683P; или

двойную мутацию, выбранную в группе, состоящей из



R400H+Y450S, D406V+E523K, A421V+R661W, R424G+Q739P, W476R+L738P, P449S+F704S, N522K+G658E, E523D+E617K, L540Q+L607P, W659R+R661W, S675C+Q695L и I679V+H723L; или

тройную мутацию, выбранную в группе, состоящей из K377E+L598P+L651H, D397E+Y430C+L666P, Q402R+V561M+Q695L, N407P+F498I+P509L, N407H+K637N+N690I, Y450H+F553L+Q716H, Y450N+H698P+Q739K, T472A+P475H+A488V, I473F+D550N+Q739E, F478Y+N522I+L727H, K484M+Q695H+Q712R, S487Y+N504S+E573D, T496A+N609D+A728G, R654H+R691Q+H698Q и R691L+H721R+I733V;

четверную мутацию, выбранную в группе, состоящей из F405L+F518L+L651P+I724V, L423P+M465R+Y515N+K673M, R457P+K468N+R661W+G715S, E470D+I548V+A589T+Q695H, A488V+D605V+R629G+T657A и M495V+K526N+S541P+K562E; или

пять мутаций, выбранных в группе, состоящей из R403H+N612Y+L651P+K652E+G715S;

шесть мутаций, выбранных в группе, состоящей из E387V+V561A+D618N+D628G+L680P+S701F, R403H+K526E+N612Y+L651P+K652E+G715S, R403H+M495T+N612Y+L651P+K652E+G715S, R403H+L502P+N612Y+L651P+K652E+G715S, R403H+K506N+N612Y+L651P+K652E+G715S, R403H+N612Y+L651P+K652E+N692I+G715S; или

семь мутаций, выбранных в группе, состоящей из R403H+A456T+N612Y+L651P+K652E+G715S+G728T, R403H+F498Y+N612Y+L651P+K652E+R661L+G715S, R403H+Q426R+F478V+N612Y+L651P+K652E+G715S; или

восемь мутаций, выбранных в группе, состоящей из R403H+R442N+V452I+N609S+N612Y+R635G+L651P+K652E+F693Y+G715S; или

девять мутаций, выбранных в группе, состоящей из R403H+R457Q+F518I+N612Y+R635G+L651P+K652E+F693Y+G715S.

Предпочтительно, в соответствии с изобретением, модифицированный полипептид Cas9 содержит по меньшей мере одну мутацию, выбранную в группе, состоящей из Y450S, M495V, Y515N, K526E, R661X, N690I, R691Q, Q695H, H698Q; где X выбирают в группе, состоящей из L, Q и S; предпочтительно X представляет собой Q или S.

В некоторых вариантах осуществления указанные мутации, идентифицированные для Cas9, подходят для усовершенствования специфичности других вариантов нуклеазы Cas9, о которых сообщали до настоящего момента (SpCas9-HF1-4, eSpCas9(1.0)-(1.1)).

Следовательно, необязательно описанный выше вариант Cas9 дополнительно может содержать одну или несколько дополнительных мутаций в остатках L169A, K810A, K848A, Q926A, R1003A, R1060A, D1135E.

В некоторых вариантах осуществления указанные мутации, идентифицированные для Cas9, подходят для усовершенствования специфичности другой нуклеазы Cas9, dCas9-FokI или dCas9. Следовательно, необязательно описанный выше вариант Cas9 дополнительно может содержать по меньшей мере одну дополнительную мутацию в остатке, выбранном в группе, состоящей из D10, E762, D839, H840, N863, H983 и D986, чтобы снижать активность нуклеазы. Предпочтительно такие дополнительные мутации представляют собой D10A или D10N и H840A, H840N или H840Y. Предпочтительно указанные мутации ведут к никазе Cas9 или к каталитически неактивной Cas9 (Ran F et al. Cell. 2013, 154(6):1380-1389; Maeder M et al. Nature Methods. 2013, 10(10):977-979).

В некоторых вариантах осуществления указанные мутации, идентифицированные для Cas9, подходят для усовершенствования специфичности вариантов нуклеазы Cas9, распознающей альтернативные последовательности PAM. Следовательно, необязательно описанный выше вариант Cas9 дополнительно может содержать одну или несколько дополнительных мутаций в остатках D1135V/R1335Q/T1337R (вариант QVR), D1135E/R1335Q/T1337R (вариант EVR), D1135V/G1218R/R1335Q/T1337R (вариант VRQR), D1135V/G1218R/R1335E/T1337R (вариант VRER) в соответствии с патентом США US20160319260.

Дополнительный объект изобретения по настоящему изобретению также представляет собой вариант белка SpCas9, как описано выше, слитый с другими полипептидными последовательностями.

Предпочтительно вариант Cas9 сливают с аминокислотными последовательностями, которые кодируют белковые метки (т. е. метку V5, метку FLAG, метку мус, метку HA, метку GST, полигистидиновую метку, метку MBP), белки, белковые домены, модуляторы транскрипции, ферменты, действующие на низкомолекулярные субстраты, ДНК, РНК и ферменты модификации белков (т. е. аденозиндезаминаза, цитидиндезаминаза, гуанозилтрансфераза, ДНК метилтрансфераза, РНК метилтрансферазы, ДНК деметилазы, РНК деметилазы, диоксигеназы, полиаденилатполимеразы, псевдоуридинсинтазы, ацетилтрансферазы,

деацетилаза, убиквитин-лигазы, дезубиквитиназы, киназы, фосфатазы, NEDD8-лигазы, дез-NEDD-илазы, SUMO-лигазы, дез-SUMO-илазы, гистондеацетилазы, гистонацетилтрансферазы гистонметилтрансферазы, гистондеметиلاзы), ДНК-связывающие домены белков, РНК-связывающие белки, полипептидные последовательности с конкретными биологическими функциями (т. е. сигналы ядерной локализации, сигналы митохондриальной локализации, сигналы пластидной локализации, сигналы субклеточной локализации, дестабилизирующие сигналы, мотивы бокса разрушения геминина), домены биологического прикрепления (т. е. MS2, Csy4 и белок  $\lambda$ N).

Дополнительный объект изобретения по настоящему изобретению представляет собой способ получения варианта Cas9, как описано выше, указанный способ включает восстановление варианта Cas9 из одного или нескольких его фрагментов; предпочтительно в соответствии с чем интеин или интрон белка или димеризующийся домен включают в полипептид Cas9.

В определенном варианте осуществления стадию восстановления можно осуществлять *in vitro*, в другом определенном варианте осуществления ее можно осуществлять *in vivo* (см. далее).

Предпочтительно такие фрагменты можно включать для того, чтобы восстанавливать белок Cas9 посредством димеризации расщепленного Cas9 (Wright AV et al. PNAS 2015 12(10):2984-9; Liu KI et al. Nat Chem Biol. 2016, 12(11):980-987).

Предпочтительно такие фрагменты можно индуцировать к восстановлению каталитически активного белка Cas9 посредством димеризации интеина в расщепленном Cas9 (Truong DJ et al. Nucleic Acids Res 43(13):6450-8).

В соответствии с изобретением вектор представляет собой систему, подходящую для доставки или экспрессии нуклеотидной или белковой последовательности.

Дополнительный объект изобретения по настоящему изобретению также представляет собой белковые или рибонуклеопротеиновые комплексы или смешанные белковые, рибонуклеопротеиновые и липидные комплексы, содержащие модифицированный полипептид Cas9 (комплексы Cas9-sgRNA и их конъюгация с дополнительными белками, нуклеиновыми кислотами или липидными компонентами, такими как, но не ограничиваясь этим, проникающие в клетку пептиды, аптамеры нуклеиновых кислот и липидные везикулы).

Дополнительный объект изобретения по настоящему изобретению

представляет собой белковый или белковый рибонуклеотидный вектор, содержащий модифицированный полипептид Cas9. В некоторых вариантах осуществления такой вектор представляет собой природный или искусственный комплекс или везикулу (см. выше). В некоторых вариантах осуществления такой вектор получают из пакующей или высвобождающей клетки. В некоторых вариантах осуществления такой вектор представляет собой внеклеточные структуры на основе везикул (т. е., но не ограничиваясь этим, экзосомы и экзосомоподобные структуры) или вирусные частицы или вирусоподобные частицы, содержащие модифицированный полипептид Cas9 в соответствии с изобретением.

Дополнительный объект изобретения по настоящему изобретению представляет собой последовательность нуклеотидов, кодирующую модифицированный полипептид Cas9, как описано выше, и его фрагменты.

Предпочтительно, в соответствии с изобретением, последовательность нуклеотидов, кодирующая модифицированный Cas9 как описано выше, или его фрагменты, основана на SEQ ID NO: 2, которая представляет собой последовательность, имеющую номер доступа NC\_002737, или SEQ ID NO: 3, которая получена через оптимизацию кодонов для экспрессии в клетках человека, и представляет замены оснований, соответствующие описанным выше мутациям. Последовательность нуклеотидов по изобретению, кодирующая модифицированный полипептид Cas9, предпочтительно обладает, за исключением мутаций, идентичностью по меньшей мере 50%, 60%, 70%, 80%, 90%, 95%, 99% или 100% с SEQ ID NO: 2 или SEQ ID NO: 3.

Дополнительный объект изобретения по настоящему изобретению представляет собой нуклеиновую кислоту, содержащую последовательность нуклеотидов как описано выше.

Способ получения модифицированного полипептида Cas9, как описано выше, в соответствии с чем модифицированный полипептид Cas9 экспрессируют с помощью нуклеиновой кислоты, как описано выше.

Дополнительный объект изобретения по настоящему изобретению представляет собой вектор, содержащий нуклеиновую кислоту, как описано выше; где указанный вектор подходит для экспрессии гена в прокариотических клетках или эукариотических клетках (например, дрожжах, клетках млекопитающих, клетках насекомого, клетках растения). Предпочтительно вектор может представлять

собой, но не ограничиваясь этим, плазмиду, фагмиду, бактериальную или дрожжевую искусственную хромосому, фрагмент ДНК или РНК или вектор на основе *Agrobacterium* или вирусный вектор.

Нуклеиновую кислоту по изобретению предпочтительно доставляют через *lentiSLiCES*, что делает возможной дополнительную специфичность через схему самоограничения, как описано в итальянской патентной заявке IT 102016000102542. Нуклеиновую кислоту по изобретению предпочтительно доставляют через ретровирусный вектор, вектор EIAV, вектор SIV, аденовирусный вектор, вектор AAV, герпес-вектор, бакуловирусный вектор, вектор вируса осповакцины, вектор вируса Сендай или бактериофаг. Предпочтительно бактериофаговый вектор основан на фаге  $\lambda$ , фаге M13 или фаге P1.

Дополнительный объект изобретения по настоящему изобретению представляет собой нуклеиновую кислоту, содержащую фрагменты последовательности нуклеотидов, как описано выше.

Предпочтительно транслируемые полипептиды, кодируемые двумя или больше различными нуклеиновыми кислотами, содержащими фрагменты последовательности нуклеотидов, как описано выше, можно использовать *in vitro* или *in vivo* для того, чтобы восстанавливать каталитически активный вариант Cas9, как описано выше.

Предпочтительно такие фрагменты Cas9 можно использовать для того, чтобы восстанавливать экспрессию белка Cas9 на уровне ДНК, используя рекомбинацию между различными вирусными векторами (т. е. Wu Z et al. *Mol Ther.* 2010, 18(1)80–86).

Предпочтительно такие фрагменты Cas9 можно использовать для того, чтобы восстанавливать белок Cas9 на уровне транскрипции посредством exploiting trans-splicing (Fine EJ et al. *Sci Rep.* 2015, 5:10777).

Дополнительный объект изобретения по настоящему изобретению также представляет собой клетку, сконструированную для того, чтобы кодировать нуклеиновую кислоту или вектор, как описано выше, или клетку, перманентно модифицированную вариантом Cas9 по изобретению.

Предпочтительно сконструированная клетка представляет собой прокариотическую клетку, более предпочтительно бактерию.

Предпочтительно сконструированная клетка представляет собой эукариотическую клетку. Предпочтительно представляет собой

клетку животного. Предпочтительно сконструированная клетка представляет собой клетку млекопитающего. Более предпочтительно представляет собой клетку человека. Предпочтительно сконструированная клетка представляет собой соматическую клетку, более предпочтительно представляет собой опухолевую клетку, наиболее предпочтительно представляет собой стволовую клетку или индуцированную плюрипотентную стволовую клетку.

Дополнительный объект изобретения по настоящему изобретению также представляет собой животное, сконструированное для того, чтобы кодировать нуклеиновую кислоту или вектор, как описано выше, или животное, перманентно модифицированное вариантом Cas9 по изобретению. Предпочтительно животное представляет собой модельный организм (т. е. *Drosophila melanogaster*, мышь, москита, крысу), или животное представляет собой сельскохозяйственное животное или искусственно выращиваемую рыбу или домашнее животное. Предпочтительно животное представляет собой вектор для по меньшей мере заболевания. Более предпочтительно организм представляет собой вектор для заболеваний человека (т. е. москиты, клещ, птицы).

Дополнительный объект изобретения по настоящему изобретению представляет собой растение, сконструированное с использованием нуклеиновой кислоты или вектора, как описано выше, или растение, перманентно модифицированное вариантом Cas9 по изобретению. Предпочтительно, растение представляет собой сельскохозяйственное растение (т. е. рис, соевые бобы, пшеница, табак, хлопок, люцерна, канола, кукуруза, сахарная свекла).

Дополнительный объект изобретения по настоящему изобретению также представляет собой способ перманентной модификации клетки, животного или растения, указанный способ включает использование молекулы Cas9 по изобретению для редактирования ДНК клетки, животного или растения.

Дополнительный объект изобретения по настоящему изобретению также представляет собой последовательность нуклеотидов или нуклеиновую кислоту или вектор, как описано выше, для применения в качестве лекарственного средства для генной терапии. Дополнительный объект изобретения по настоящему изобретению также представляет собой фармацевтическую композицию, содержащую последовательность нуклеотидов или нуклеиновую кислоту или вектор, как описано выше, и по меньшей мере один фармацевтически приемлемый эксципиент.

Дополнительный объект изобретения по настоящему изобретению также представляет собой фармацевтическую композицию, содержащую рекомбинантный полипептид Cas9, содержащий описанные выше мутации, и по меньшей мере один фармацевтически приемлемый эксципиент.

Дополнительный объект изобретения по настоящему изобретению также представляет собой использование *in vitro* последовательности нуклеотидов или нуклеиновой кислоты или вектора, как описано выше, для конструирования генома, клеточной инженерии, экспрессии белка или других биотехнологических применений.

Дополнительный объект изобретения по изобретению представляет собой использование *in vitro* рекомбинантного полипептида Cas9, содержащего описанные выше мутации, вместе с гРНК, для конструирования генома, клеточной инженерии, экспрессии белка или других биотехнологических применений.

Дополнительный объект изобретения по изобретению представляет собой набор частей, для одновременного, раздельного или последовательного использования, который содержит последовательность нуклеотидов или нуклеиновую кислоту или вектор или рекомбинантный полипептид Cas9 как описано выше.

Способ изменения генома клетки *in vitro* или *in vivo*, способ включает экспрессию в клетке модифицированного Cas9, как описано выше, вместе с гидовой РНК, направленной на конкретную геномную последовательность.

Способ изменения транскрипта клетки *in vitro* или *in vivo*, способ включает экспрессию в клетке модифицированного транскрипционного регулятора на основе dCas9, как описано выше, вместе с гидовой РНК, направленной на конкретную геномную последовательность.

Способ изменения эпигенома клетки *in vitro* или *in vivo*, способ включает экспрессию в клетке модифицированного средства редактирования эпигенома на основе dCas9, как описано выше, вместе с гидовой РНК, направленной на конкретную геномную последовательность.

Настоящее изобретение легче понять в свете следующих экспериментов.

#### КРАТКОЕ ОПИСАНИЕ ФИГУР

Фиг. 1. Схема анализа на дрожжах *in vivo* для того, чтобы количественно определять активность SpCas9 в и вне мишени.

Создание репортерных штаммов дрожжей. Локусы TRP1 и ADE2 модифицировали посредством инсерции репортерной кассеты, содержащей участок в мишени (TRP1) или другие последовательности вне мишени (ADE2, последовательности приведены в верхнем правом блоке). Присутствие областей гомологии (HR) по обеим сторонам от мишени делает возможной эффективную репарацию посредством отжига одной цепи при расщеплении с помощью Cas9. Используя подходящие селективные чашки, можно проследить состояние редактирования двух локусов. Выживаемость колонии указывает на расщепление TRP1 в мишени, тогда как цвет колонии позволяет оценивать расщепление ADE2 вне мишени.

Фиг. 2. Валидация штамма  $\gamma$ АСМО с использованием SpCas9 дикого типа. Количественное определение средней процентной доли красных (в мишени) и белых (вне мишени) колоний после трансформации wt SpCas9 в репортерные штаммы  $\gamma$ АСМО-off1/off4, стабильно экспрессирующие sgRNA, совпадающую с последовательностью в мишени в локусе TRP1. Экспрессию Cas9 индуцировали в течение 24 ч перед высеванием. Планки ошибки представляют s.e.m. для n=3.

Фиг. 3. Отбор целевого домена для создания случайной библиотеки. Схемы доменов SpCas9. Домен Rec1-II (выделен) является частью  $\alpha$ -спиральной распознающей доли. ВН: мостиковая спираль.

Фиг. 4. Дрожжевой скрининг для высокоспецифичных вариантов SpCas9. (a) Схематическое представление потока операций дрожжевого скрининга. (b) Оценка активности в мишени и специфичность вариантов SpCas9, полученных из дрожжевого скрининга через анализ с использованием репортерного штамма  $\gamma$ АСМО-off4. Экспрессию Cas9 индуцировали в течение 5 ч перед высеванием. См. таблицу приложения 1. (c) Оценка специфичности варианта C13. C13 трансформировали в репортерные штаммы  $\gamma$ АСМО-off1/off4, экспрессирующие sgRNA в мишени, и оценивали процентную долю красных (в мишени) и белых (вне мишени) колоний после 24 ч экспрессии Cas9. Планки ошибки представляют s.e.m. для n=3.

Фиг. 5. Отбор оптимизированных вариантов SpCas9 в клетках млекопитающих. Клетки 293, стабильно экспрессирующие EGFP, трансфицировали одинарными и двойными (a), а также тройными (b) мутантами, созданными посредством иерархической комбинации мутаций, получаемых из наиболее эффективных вариантов,



выделенных из дрожжей, вместе с sgRNA в мишени (sgGFPon) или каждой из несовпадающих гидовых. Потерю флуоресценции EGFP измеряли посредством FACS анализа через 7 суток после трансфекции. (с) Параллельное сравнение лучших созданных вариантов с предварительно опубликованными мутантами. Анализ нокаута EGFP в клеточной линии 293 использовали для того, чтобы оценивать специфичность лучших выделенных вариантов (варианты VNEL, VNEQ и VNES) и чтобы сравнивать их эффективность с предварительно опубликованными высокоточными мутантами. Потерю флуоресценции EGFP измеряли посредством FACS анализа через 7 суток после трансфекции. sgGFP1314 содержит несовпадения в положении 13 и 14 из PAM спейсерной последовательности; sgGFP1819 содержит несовпадения в положениях 18 и 19; sgGFP18 содержит одно несовпадение в положении 18. Пунктирные линии показывают соотношения в мишени/вне мишени, вычисляемые для указанных пар в мишени/вне мишени. Штриховые линии показывают фоновую потерю флуоресценции EGFP. Планки ошибки представляют s.e.m. для  $n \geq 2$ .

Фиг. 6. Активность evoCas9 в мишени против EGFP. (а) Активность evoCas9 против локусов EGFP. Клетки 293, стабильно экспрессирующие EGFP, трансфицировали с использованием wt SpCas9, evoCas9 (VNEQ) или варианта VNEL вместе с sgRNA, направленными на различные области кодирующей последовательности EGFP. Потерю флуоресценции EGFP измеряли посредством FACS анализа через 7 суток после трансфекции. Штриховые линии показывают фоновую потерю флуоресценции EGFP. Планки ошибки представляют s.e.m. для  $n \geq 2$ . (b) Соотношение активность в мишени для evoCas9 и варианта VNEL относительно SpCas9 дикого типа вычисляли в локусах EGFP. Показаны медиана и межквартильный размах. Уровень активности в мишени выше 70% от белка wt показан заштрихованной областью. (с) Внутриклеточная экспрессия evoCas9. Репрезентативный вестерн-блоттинг лизатов из клеток 293T, трансфицированных wt SpCas9, evoCas9 или других высокоточных вариантов. Тубулин использовали в качестве контроля загрузки. SpCas9 обнаруживают с использованием антитела против FLAG.

Фиг. 7. Активность evoCas9 в эндогенных локусах. (а) Активность wt SpCas9, evoCas9 и SpCas9-HF1 против эндогенных локусов сравнивали посредством трансфекции клеток 293T и посредством измерения формирования инсерций-делеций через 7 суток после трансфекции с использованием инструмента TIDE. (b)

Соотношение активности в мишени для *evoCas9* и *SpCas9-HF1* относительно *SpCas9* дикого типа, вычисляемое для эндогенных локусов. Представлены медиана и межквартильный размах. Уровень активности в мишени выше 70% белка *wt* показан заштрихованной областью. Планки ошибки представляют *s.e.m.* для  $n=2$ .

Фиг. 8. Параллельное сравнение специфичности *evoCas9* и *SpCas9-HF1*. (а) Активность вне мишени для *evoCas9* в выбранных локусах. Клетки 293Т трансфицировали с использованием *wt SpCas9*, *evoCas9* или *SpCas9-HF1* вместе с *sgRNA*, направленными на участок 2 *FANCF* или локусы *CCR5 sp11*. Формирование инсерций-делеций в двух предварительно валидированных участках вне мишени оценивали через 7 суток после трансфекции с использованием инструмента TIDE. Планки ошибки представляют *s.e.m.* для  $n=2$ . (b) Соотношение в/вне мишени, вычисленное по средним процентным долям инсерций-делеций, полученным в (а). Штриховые линии показывают и соотношение в/вне мишени, равное 1. (с) Схематическое представление локуса *CCR5* и его участка вне мишени в высоко гомологичном гене *CCR2*. Одновременное расщепление двух участков создает хромосомную делецию приблизительно в 16 т. о. Осуществляли полуколичественную ПЦР геномной ДНК клеток 293Т, трансфицированных с использованием *wt SpCas9*, *evoCas9* или *SpCas9-HF1* и гидовой РНК *CCR5* для того, чтобы оценивать количество хромосомной делеции, создаваемой при каждом условии. Локус *FANCF* использовали в качестве внутреннего нормализатора. Количество делеции определяли с использованием денситометрии и ImageJ. Планки ошибки представляют *s.e.m.* для  $n=2$ .

Фиг. 9. Валидация специфичности *evoCas9* посредством направленного глубокого секвенирования. Направленное глубокое секвенирование предварительно валидированных участков вне мишени относительно локуса участка 1 *EMX1* (а) и локуса участка 3 *VEGFA* (b) осуществляли на геномной ДНК клеток 293Т, экспрессирующих или *wt SpCas9* или *evoCas9* вместе с каждой специфической *sgRNA*. Клетки, не экспрессирующие *Cas9*, секвенировали для того, чтобы определять фоновые уровни инсерций-делеций. Геномную ДНК из трех биологических реплик смешивали перед получением библиотеки.

Фиг. 10. Общегеномная специфичность *evoCas9*. (а) анализ GUIDE-seq участков вне мишени по сравнению с локусом участка 2 *VEGFA* выполняли как для *wt SpCas9*, так и для *evoCas9* в клетках 293Т. Черный квадрат показывает участок в мишени. (b) Общее число обнаруживаемых участков вне мишени для *wt SpCas9* и

evoCas9. Геномную ДНК из трех биологических реплик смешивали перед получением библиотеки.

Фиг. 11. Транскрипционная активация evo-dCas9. (a) Схематическое представление репортер транскрипционного активатора на основе Tet чувствительного элемента (TRE)-EGFP. При связывании dCas9-VP64 с повторами TetO активируют экспрессию EGFP. (b) Активацию EGFP измеряли в клетках 293T, трансфицированных с использованием транскрипционных активаторов на основе dCas9 или evo-dCas9 вместе с совпадающими sgRNA (как с 5' несовпадающим G, так и без него) или несовпадающими гидовыми. TetO-off6 содержит несовпадение в положении 6 из PAM, TetO-off1314 содержит два несовпадения в положении 13-14 и TetO-off1819 содержит два несовпадения в положении 18-19. (c) Кратность активации экспрессии EGFP относительно контроля вне мишени, вычисляемая по данным в (b). Планки ошибки представляют s.e.m. для n=2. Экспрессию EGFP измеряли посредством FACS анализа через 2 суток после трансфекции.

Фиг. 12. Долгосрочная специфичность evoCas9. SpCas9 (a), evoCas9 (b) и SpCas9-NF1 (c) стабильно экспрессировали через лентивирусную доставку в клетки 293T, стабильно экспрессирующие EGFP. Каждый лентивирусный вектор нес sgRNA в мишени к кодирующей последовательности EGFP или другие несовпадающие гидовые, представленные на фиг. 5. Нокаут EGFP оценивали посредством FACS анализа в моменты времени, указанные на графиках. Трансдуцированные клетки культивировали в селективной среде с пурамицином. Планки ошибки представляют s.e.m. для n=2.

#### ЭКСПЕРИМЕНТАЛЬНЫЙ РАЗДЕЛ

Разработка репортерного штамма дрожжей для обнаружения активности Cas9

*Saccharomyces cerevisiae* использовали в качестве экспериментальной модели для разработки скрининга направленной эволюции для того, чтобы выделять высокоспецифичные варианты SpCas9. Преимущество использования платформы анализа на основе дрожжей состоит с одной стороны в сходстве дрожжей и бактерий, таком как высокая скорость удвоения, возможность без труда выделять отдельные клоны и доступность быстрых и надежных протоколов трансформации; с другой стороны, организация ДНК и метаболизма дрожжей схожа с таковой у высших эукариотических клеток. Следовательно, дрожжевая модель предлагает гибкую платформу для высокопропускного скрининга в комбинации со

сходством с системой млекопитающего, что увеличивает устойчивость исхода скрининга. Изначально разрабатывали стратегию создания ауксотрофных репортерных штаммов дрожжей для одновременного измерения активности Cas9 в и вне мишени. Этот подход состоял в модификации двух геномных локусов дрожжей: TRP1 (хромосома IV) и ADE2 (хромосома XV). Используя подход *delitto perfetto*, кодирующие последовательности дикого типа для двух локусов разделяли на две половины, разделенные конкретной целевой последовательностью, совпадающей с sgRNA в мишени в случае локуса TRP1, или различными последовательностями вне мишени в случае локуса ADE2 (схематично на фиг. 1). Каждое вне мишени (ADE2off1-off4, таблица приложения 4) содержало одно несовпадение, расположенное на возрастающих расстояниях от последовательности PAM. Дупликацию 100 п. о. добавляли по обеим сторонам от целевой последовательности (TRP1 или ADE2) и терминирующий кодон располагали непосредственно выше по направлению считывания, между двумя областями гомологии, чтобы обеспечивать преждевременное прерывание трансляции (фиг. 1). Нокаут генов TRP1 и ADE2 посредством инсерции репортерной кассеты создает дефекты в метаболических путях триптофана и аденина, что подавляет рост в отсутствие триптофана и ведет к накоплению в клеточной вакуоли промежуточного продукта биосинтеза аденина, когда клетки растут при низких концентрациях аденина, таким образом придавая характерную красную пигментацию колониям на чашках с агаром. После формирования двухцепочечных разрывов, индуцированных с помощью Cas9, каждый локус можно эффективно репарировать с помощью клеток дрожжей, используя путь репарации отжигом одной цепи благодаря присутствию двух фланкирующих гомологичных областей, достигая реверсной мутации прототрофности по двум питательным веществам. Успешное событие редактирования в каждом из двух локусов можно в свою очередь визуализировать с использованием подходящих репортерных чашек, которые истощают по триптофану и содержат только низкие концентрации аденина (SDluta<sub>5</sub>, фиг. 1). Полное прочтение анализа состоит из двухстадийного процесса. Первая стадия состоит в оценке эффективности расщепления в мишени посредством сравнения числа колоний, получаемых в репортерных чашках SDluta<sub>5</sub>, с таковым, получаемым при отборе только по общему числу трансформантов (чашки SDlu). Вторая стадия состоит в подсчете числа красных колоний (TRP1<sup>+</sup>/ADE2<sup>-</sup>), соответствующих расщеплению в мишени, и

белых колоний (TRP1<sup>+</sup>/ADE2<sup>+</sup>), в которых также проводили редактирование локуса вне мишени, на репортерных чашках для оценки активности в и вне мишени. Создавали четыре штамма дрожжей, содержащих четыре потенциальных участка вне мишени (ADE2off1-off4), которые называли  $\gamma$ АСМО-off1/off4. Затем оценивали уровень самопроизвольной реверсной мутации для локусов ADE2 и TRP1 различных штаммов  $\gamma$ АСМО для того, чтобы избежать введения любой искажающей переменной при прочтении анализа. Реверсная мутация локуса TRP1 может фактически вести к выделению ложноположительных клонов, тогда как самопроизвольная рекомбинация гена ADE2 может создавать ложноотрицательные колонии. Чтобы приближать экспериментальные условия, используемые в ходе анализа авторов изобретения, каждый из четырех штаммов отдельно трансформировали плазмидой, кодирующей SpCas9; после 24 часов инкубации в селективной среде, клетки распределяли по селективным чашкам для того, чтобы измерять общее число трансформантов и в 1000 раз больше клеток высевали на селективных чашках, истощенных по триптофану или аденину, чтобы подсчитывать число ревертантов для каждого локуса. Посредством сравнения числа колоний, получаемых в различных селективных условиях, возможно оценивать среднюю частоту реверсных мутаций около  $1-1,5 \times 10^{-5}$  для обоих локусов TRP1 и ADE2, таким образом показывая незначительную самопроизвольную реверсную мутацию.

#### Валидация репортерного штамма $\gamma$ АСМО

Функциональность репортерного анализа валидировали посредством тестирования четырех репортерных штаммов ( $\gamma$ АСМО-off1/off4) в комбинации с SpCas9 дикого типа. Для того чтобы максимизировать общую эффективность, перед испытанием с использованием SpCas9, каждый из штаммов стабильно трансформировали плазмидой, экспрессирующей sgRNA-on, превосходно совпадающую с последовательностью в мишени в локусе TRP1. Затем четыре штамма трансформировали плазмидой для экспрессии SpCas9 дикого типа под управлением галактозного индуцибельного промотора и, после 4 часов восстановительной инкубации, индуцировали в течение ночи в содержащих галактозу средах перед высеванием на чашки SDlu и репортерные чашки SDluta<sub>5</sub>. В этих экспериментальных условиях авторы изобретения согласованно достигали 100% расщепления в мишени, при этом активность вне мишени, измеряемая как процентная доля белых

колоний (TRP1<sup>+</sup>/ADE2<sup>+</sup>) на репортерных чашках, возрастала в соответствии с расстоянием от несовпадающего основания до последовательности PAM, как ожидали (фиг. 2). Для двух наиболее PAM-дистальных вне мишени (off3 и off4), SpCas9 был полностью неспособен различать совпадающую и две несовпадающих последовательности (фиг. 2).

Учитывая эти результаты, осуществляли скрининг вариантов SpCas9 с использованием штамма *yACMO-off4*, содержащего сильнейшую последовательность вне мишени, чтобы выбирать мутанты с заметным увеличением точности.

Скрининг на основе дрожжей для высокоспецифичных вариантов SpCas9

В отличие от опубликованных исследований (Slaymaker IM et al., *Science*. 2016, 351(6268):84-8; Kleinstiver BP et al., *Nature*. 2016, 529(7587):490-5), авторы изобретения полагали, что объективный подход может вести к выделению нетривиальной замены аминокислоты, которая увеличивает вероятность получения варианта SpCas9 с более высокой точностью. Чтобы найти подходящую мишень для случайного мутагенеза, доступные структурные данные анализировали для того, чтобы идентифицировать, какой домен SpCas9 может быть больше вовлечен в формирование взаимодействий такого типа. Долю нуклеазы Cas9 исключали из этого анализа, поскольку она содержит два каталитических сайта, которые нужно сохранять для того, чтобы поддерживать активность расщепления. Сообщалось, что распознающая доля, содержащая домены Rec1, Rec2 и мостиковой спирали, образует несколько контактов с дуплексом гРНК:ДНК. Кроме того, распознающая доля в целом представляет собой одну из наименее консервативных областей среди всех трех семейств Cas9, относящихся к системам CRISPR II типа, что указывает на высокую степень пластичности последовательности. Мостиковая спираль, напротив, представляет собой одну из наиболее консервативных областей среди различных ортологов Cas9, что подсказывает, что ее последовательность очень важна для функции нуклеазы. Область Rec1-Rec2 охватывает более чем 600 аминокислот, такой размер не подходит для случайного мутагенеза, но большинство взаимодействующих остатков располагаются в последней части домена Rec1-II, приблизительно между остатками 400 и 700 (фиг. 3).

Библиотеку вариантов REC1-II, созданных с помощью ПЦР, подверженной ошибкам, чтобы они содержали приблизительно 4-5

мутаций на молекулу, непосредственно собирали в репортерном штамме дрожжей  $\gamma$ АСМО-off4, используя гомологичную рекомбинацию между мутировавшими фрагментами REC1-II, содержащими подходящие гомологичные плечи, и плазмиду, экспрессирующую индуцируемый галактозой SpCas9, где ту же область предварительно удаляли. Общий поток операций скрининга схематически изображен на фиг. 4a. После совместной трансформации и восстановительной инкубации в течение ночи в среде SDlu для того, чтобы сделать возможной репарацию Cas9-кодирующей плазмиды посредством гомологичной рекомбинации и отбора трансформированных клеток, культуры индуцировали в течение 5 часов с использованием галактозы для того, чтобы экспрессировать SpCas9, и затем высевали на несколько репортерных чашек SDluta<sub>5</sub>. Время индукции укорачивали относительно предыдущих экспериментов, чтобы получать варианты, которые поддерживали высокую активность в мишени, поскольку авторы изобретения наблюдали, что SpCas9 дикого типа может полностью расщеплять последовательность в мишени в этом ограниченном временном интервале (данные не приведены). Через двое суток получали множество колоний, и красные колонии штрихами сеяли на новые репортерные чашки, содержащие галактозу вместо декстрозы, чтобы повторно активировать экспрессию SpCas9 и поддерживать ее постоянно включенной, чтобы усиливать любой эффект вне мишени. После 48 часов плазмиды извлекали из самых красных штрихов и после амплификации в бактериях проводили секвенирование по Сэнгеру для того, чтобы идентифицировать мутации, введенные в домен REC1-II. Идентифицировали несколько аминокислотных замен, некоторые из которых присутствовали более чем один раз в совокупности мутантов в комбинации с группами различных мутаций. Вероятно следует отметить, что мутанты, содержащие тот же набор вариаций, представляют клоны, происходящие из той же исходной клетки, которую реплицировали в ходе восстановительной инкубации. Однако, учитывая разнообразие полученных замен, этот феномен не влияет на результаты скрининга. Затем осуществляли эксперимент повторного испытания в штамме  $\gamma$ АСМО-off4 с каждым выделенным вариантом для того, чтобы измерять более точно его активность расщепления, отбрасывать те, которые не эффективно разрезали свою мишень по сравнению с SpCas9 дикого типа, и ранжировать оставшиеся в соответствии с последним параметром и их способностью различать участки вне мишени (фиг. 4b). Чтобы дополнительно валидировать результаты

скрининга, специфичность одного из полученных вариантов (вариант C13) оценивали более подробно посредством испытания всех четырех репортерных штаммов уАСМО. После 24 часов экспрессии Cas9, количественное определение белых и красных колоний на репортерных чашках показывало значительно сниженную активность вне мишени по сравнению с SpCas9 дикого типа (сравните фиг. 4с и фиг. 2).

Оптимизация высокоточных вариантов SpCas9 в клетках млекопитающих

Выбирали совокупность замен, относящихся к наиболее эффективным вариантам, выделенным при дрожжевом скрининге в соответствии как с эффективностью расщепления в мишени, так и снижением неспецифической активности. Чтобы достигать значительного увеличения точности в отношении идентифицированных мутантов предпринимали попытку иерархической комбинации этих мутаций, поскольку ожидали, что некоторые из замен в каждом случайно созданном варианте могли быть нейтральными или вредоносными. Относительное положение каждой замены и дуплекса sgRNA:ДНК, в соответствии с доступными структурными данными, использовали в качестве первого критерия фильтрации, идентифицирующего первый поднабор мутаций. Кроме того, внимание привлекал конформационный кластер замен, расположенных на конце домена REC1-II, который находится в контакте с более РАМ-дистальной частью целевой последовательности ДНК (н. 17–20). Таким образом, мутации, относящиеся к этому кластеру, отбирали даже несмотря на то, что ранее не сообщалось о взаимодействии с дуплексом sgRNA:ДНК. В частности, решали добавлять мутации последовательно, начиная с мутанта K526E, который работал особенно хорошо в дрожжевом анализе (фиг. 4с). Стоит отметить, что мутацию K526E включали в указанный выше домен Rec1-II (фиг. 3).

Используя репортерную клеточную линию, стабильно экспрессирующую EGFP (293multiGFP), активность в мишени (sgGFPon) двойных мутантов тестировали посредством измерения потери флуоресценции, индуцированной мутациями со сдвигом рамки в кодирующей последовательности EGFP. Параллельно оценивали их способность избегать расщепления одного и того же участка после введения в sgRNA одного или двух несовпадающих оснований в положениях, дистальных от тринуклеотида РАМ (положение 18 для sgGFP18 и положения 18–19 для sgGFP1819). SpCas9 дикого типа не



был способен различать эти суррогатные последовательности вне мишени, как подтверждали посредством наблюдения о том, что он расщеплял целевую последовательность с равной эффективностью, когда его направляли как совпадающие, так и несовпадающие sgRNA, вызывая такое же снижение процентной доли EGFP<sup>+</sup> клеток (фиг. 5a). После первого раунда отбора, наиболее эффективные замены комбинировали в тройных мутантах, которые использовали для повторения испытания EGFP репортерной клеточной линии (фиг. 5b). Затем осуществляли последний раунд отбора после создания четверного мутанта посредством объединения лучших замен из предыдущего раунда (вариант VNEL). Кроме того, другую sgRNA, содержащую два несовпадения в более PAM-проксимальной области (положения 13 и 14, sgGFP1314) тестировали для того, чтобы верифицировать, что наблюдаемое увеличение точности варианта VNEL сохраняли для мутаций, охватывающих всю спейсерную последовательность. Вариант VNEL вызывал небольшую или нулевую потерю флуоресценции EGFP для всех несовпадающих гидовых РНК, этот результат особенно удивителен для sgRNA, содержащей одну замену в положении 18 из PAM (sgGFP18). С другой стороны, это сильное увеличение специфичности вызывало небольшое, но поддающееся измерению снижение активности в мишени (потеря ~20%, фиг. 5с). Для того чтобы разобраться с этим вопросом, создавали две альтернативные производные варианта VNEL (варианты VNEQ и VNES) посредством рационального конструирования и тестировали их с использованием того же анализа нокаута EGFP. Как ожидали, наблюдали полное восстановление эффективности расщепления в мишени параллельно с небольшим увеличением активности вне мишени. В целом, мутант варианта VNEQ демонстрировал лучшее соотношение в/вне (фиг. 5с).

Параллельное сравнение четверных мутантов (варианты VNEL, VNEQ, VNES) с предварительно опубликованными высокоточными вариантами SpCas9-HF1 и eSpCas9(1.1) с использованием описанной выше EGFP репортерной клеточной линии обнаруживало заметное увеличение точности, которое в частности было очевидно при использовании sgRNA, содержащей одно несовпадение в положении 18. Для этого конкретного суррогата вне мишени, абсолютное снижение неспецифического расщепления приблизительно от 17 до 4 раз измеряли при сравнении вариантов VNEL, VNEQ и VNES с SpCas9-HF1, которые в соответствии с данными экспериментами уже различали несовпадающие участки значительно лучше, чем

eSpCas9(1.1) (фиг. 5с). Это наблюдение дополнительно подтверждали посредством анализа соотношений в/вне мишени для различных вариантов SpCas9, которые вычисляли для двух сильнейших суррогатов вне мишени (sgGFP1819 и sgGFP18) (сравните пунктирные линии на фиг.5с).

Затем активность в мишени для вариантов VNEL и VNEQ оценивали более подробно посредством направленного воздействия на различные области кодирующей последовательности EGFP, используя репортерную клеточную линию 293multiGFP и измеряя потерю флуоресценции EGFP. Вариант VNES исключали из анализа, поскольку он вел себя аналогично мутанту VNEQ. В соответствии с предыдущими результатами (фиг. 5с), авторы изобретения наблюдали уровни активности дикого типа для VNEQ, тогда как VNEL был слегка менее эффективным в некоторых из участков, при значительном падении активности для одного из тестируемых локусов (фиг. 6а и 6b). Чтобы исключить возможность, что различное расщепляющее поведение, измеряемое в отношении участков в и вне мишени, обусловлено изменением внутриклеточных уровней вариантов SpCas9, уровни белка SpCas9 дикого типа, VNEL и VNEQ, а также двух ранее опубликованных высокоточных вариантов eSpCas9(1.1) и SpCas9-HF1 анализировали в клетках 293T, трансфицированных равными количествами экспрессирующих плазмид. Как ясно показывают результаты (фиг. 6с), не наблюдали больших различий в уровнях белка. Учитывая эти результаты, дополнительно определяли характеристики варианта VNEQ, поскольку среди мутантов, анализируемых здесь, он сохранял уровни активности, близкие к дикому типу, и существенно снижал расщепление несовпадающих последовательностей в клеточной модели разрушения EGFP. Мутант VNEQ называли evoCas9 (развитый Cas9).

Активность evoCas9 в отношении эндогенных локусов

Вышеуказанные находки дополнительно валидировали посредством направленного воздействия на эндогенные локусы. Группу предварительно протестированных геномных целевых участков таким образом отбирали для того, чтобы сравнивать активность расщепления evoCas9 с таковой у SpCas9 дикого типа в каждом локусе. Кроме того, SpCas9-HF1 также вводили в сравнение, в качестве дополнительного репера. После трансфекции в клетках 293T каждого варианта SpCas9 вместе с sgRNA, направленными на различные локусы, формирование инсерций-делеций анализировали с использованием пакета программного обеспечения Tracking of

Indels by Decomposition (TIDE) на секвенированных по Сэнгеру ампликонах относительно каждого целевого участка (Brinkman EK et al., *Nucleic Acids Res.* 2014, 42(22):e168). Для большинства локусов авторы изобретения не наблюдали какое-либо большое различие в эффективности направленного воздействия между SpCas9 дикого типа и evoCas9, причем последний был в целом слегка менее активным при общей средней активности, которая составляет 80% от таковой у белка дикого типа (фиг. 7a и фиг. 7b). Для одного целевого участка, локуса ZSCAN2, авторы изобретения наблюдали очень низкую эффективность расщепления, без очевидного объяснения такого поведения (фиг. 7a). Вариант SpCas9-HF1 демонстрировал общую более низкую эффективность расщепления, при глобальной средней активности, которая составляет 60% по сравнению с диким типом (фиг. 7a и фиг. 7b). Это не соответствует предварительно опубликованным наблюдениям (Kleinstiver BP et al., *Nature.* 2016, 529(7587):490-5) и возможно предположить, что это несоответствие может быть обусловлено различными экспериментальными процедурами. Эти данные демонстрируют, что evoCas9 сохраняет уровни активности в мишени, близкие к дикому типу, в отношении панели эндогенных локусов, превосходя вариант SpCas9-HF1, о котором сообщали ранее, в тестируемых экспериментальных условиях.

#### Оценка активности evoCas9 вне мишени

Вместе с активностью в отношении участков в мишени, специфичность evoCas9 измеряли посредством верификации уровня редактирования в двух предварительно валидированных участках вне мишени, ассоциированных с редактированием в двух локусах: участок 2 FANCF 2 и CCR5 sp11. Редактированием Cas9 в этих локусах создают вставки, поскольку ассоциированное с участком 2 FANCF вне мишени представляло собой один из нескольких неповторяющихся участков, которые SpCas9-HF1 неспособен отличать от специфического участка в мишени, тогда как локус CCR5sp11, который имеет значение для его терапевтического применения в лечении СПИД, коррелирует с расщеплением высоко гомологичного гена CCR2 вне мишени. После временной трансфекции в клетках 293T, формирование инсерций-делеций в этих двух локусах вне мишени измеряли с использованием TIDE, обнаруживая значительное снижение расщепления в клетках, экспрессирующих evoCas9, по сравнению с клетками, трансфицированными SpCas9 дикого типа (фиг. 8a). Кроме того, вычисление соотношения в/вне мишени для

SpCas9 дикого типа, evoCas9 и SpCas9-HF1 подтверждало, что вариант по изобретению способен превосходить своих конкурентов в этих двух локусах (фиг. 8b). Согласованное расщепление локуса CCR5sp11 и его вне мишени в гене CCR2 создает хромосомную делецию приблизительно 16 т. о. (схематично на фиг. 8с). Частоту этой хромосомной перестройки измеряли посредством полуколичественной ПЦР в клетках, трансфицированных с использованием SpCas9 дикого типа, evoCas9 или SpCas9-HF1 вместе с sgRNA, направленной на локус CCR5. Хотя событие транслокации, в частности, было очевидно в клетках, которые редактировали с помощью SpCas9 дикого типа, сильное снижение количества делеции до едва поддающихся обнаружению уровней наблюдали как в присутствии SpCas9-HF1, так и evoCas9, причем последний дополнительно снижал измеряемую делецию почти двухкратно по отношению к первому (фиг. 8с).

Затем изучали способность evoCas9 избегать нежелательных геномных расщеплений посредством выполнения направленного глубокого секвенирования на выбранном наборе участков вне мишени, ассоциированных с редактированием в участке 3 VEGFA и геномных участках EMX1-K. Для всех выбранных участков ранее показано редактирование вместе с локусом в мишени (Kleinstiver BP, et al., Nature. 2015, 523(7561):481-5). Преимущество подхода секвенирования ампликонов заключается в возможности одновременно измерять несколько мишеней с большим покрытием для того, чтобы обнаруживать даже малораспространенные события редактирования. Анализ данных секвенирования демонстрировал, что, несмотря на сохранение высокой активности в мишени в обоих геномных локусах, evoCas9 отличался фоновыми уровнями редактирования в большинстве тестируемых участков вне мишени (фиг. 9a и 9b). Первый участок 3 VEGFA вне мишени (VEGFA3-OT1) представлял собой только локус, где возникала активность редактирования evoCas9 выше фона (фиг. 9b). Однако, следует учитывать, что тот же локус редактируют почти как участок в мишени с помощью SpCas9 дикого типа и что вариант SpCas9-HF1, о котором сообщали ранее, неспособен полностью подавлять неспецифическое расщепление этой последовательности (Kleinstiver BP et al. Nature., 2016, 529(7587):490-5). Предполагали, что эти результаты можно объяснять часто повторяющейся природой этого конкретного локуса. Для четырех участков 3 VEGFA вне мишени (VEGFA3-OT1, -OT4, -OT5, -OT7) обнаруживали значительно более высокие фоновые уровни

редактирования (фиг. 9b), вероятно из-за необычных характеристик локального хроматина, который является более хрупким и, таким образом, склонным к накоплению мутаций.

Все вместе эти данные указывают на то, что *evoCas9* значительно снижает нежелательные геномные расщепления до неподдающихся обнаружению уровней для большинства тестируемых участков вне мишени. Кроме того, параллельные сравнения с предварительно опубликованным вариантом *SpCas9-HF1* для выбранных участков вне мишени демонстрировали увеличенную способность различать несовпадающие участки.

#### Общегеномная специфичность *evoCas9*

Анализ активности *evoCas9* вне мишени на общегеномном уровне расширяли с использованием ранее созданного приема GUIDE-seq (Tsai SQ et al., Nat Biotechnol. 2015, 33(2):187-97). Этот подход основан на встраивании олигонуклеотидной метки 34 п. о. в участки, которые резали с помощью *Cas9* для того, чтобы сделать возможным их захват для получения библиотеки и секвенирования нового поколения. Таким образом, возможно идентифицировать беспристрастным образом совокупность новых геномных участков, ассоциированных с конкретной гидовой РНК, на которые неспецифически воздействовали с помощью *Cas9*. Анализ GUIDE-seq осуществляли для того, чтобы анализировать участки вне мишени, ассоциированные с редактированием локуса участка 2 *VEGFA*, который является часто повторяющимся и для которого показано, что он создает многие нежелательные расщепления в клеточном геноме. Дополнительно, недавние сообщения (Kleinstiver BP et al., Nature. 2016, 529(7587):490-5) указывают на то, что некоторое из обнаруживаемого вне мишени все еще подвержено расщеплению с помощью высокоточного варианта *SpCas9-HF1*. Библиотеки GUIDE-seq, таким образом, создавали из геномной ДНК клеток 293Т, трансфицированных с использованием *SpCas9* дикого типа или *evoCas9*, вместе с sgRNA участка 2 *VEGFA* и затравочным двухцепочечным олигонуклеотидом. Данные секвенирования анализировали с использованием программного конвейера с открытым источником (Tsai SQ et al., Nat Biotechnol. 2016, 34(5):483), обнаружив 600 различных участков вне мишени для *SpCas9* дикого типа, для которых характерно 7 или меньше несовпадений с последовательностью в мишени (фиг. 10a и 10b). Следует отметить, что приблизительно 100 этих последовательностей вне мишени редактировали более эффективно этот участок в мишени, в

соответствии с числом прочтений, полученных после анализа GUIDE-seq, о чем сообщали, что это является хорошей заменой фактической активности расщепления в каждом конкретном участке (Tsai SQ et al., Nat Biotechnol. 2015, 33(2):187-97) (фиг. 10a). Когда тот же анализ осуществляли на образцах evoCas9, обнаруживали всего только 10 участков, большинство которых имело высокое сходство в мишени и охарактеризовано меньше чем пятью несовпадениями с локусом участка 2 VEGFA (фиг. 10a). Только одно вне мишени возникало в анализе, показывающем высокую частоту расщепления (более чем в мишени) с помощью evoCas9, и это вероятно из-за специфической природы этой последовательности, которая отличалась только двумя нуклеотидами от предполагаемой мишени и содержала два непрерывающихся фрагмента из цитозинов на уровне каждого несовпадения (фиг. 10a), возможно допуская формирование выпуклых участков, чтобы вмещать несовпадающие нуклеотиды.

В общем, вышеуказанный анализ GUIDE-seq демонстрировал, что evoCas9 сохраняет очень высокую специфичность на общегеномном уровне, даже когда тестировали с использованием повторяющейся целевой последовательности, отличающейся несколькими вне мишени, в клеточном геноме.

Специфичность транскрипционного активатора на основе evo-dCas9

Альтернативное применение для Cas9, независимо от его активности нуклеазы, состоит в создании РНК-направляемых транскрипционных активаторов посредством слияния каталитически неактивной версии Cas9 (dCas9) с различными белковыми доменами, способными стимулировать транскрипцию. Транскрипционный активатор на основе VP64 конструировали с использованием каталитически неактивного мутанта evoCas9 (evo-dCas9) и тестировали параллельно с активатором dCas9-VP64 дикого типа. Транскрипционную активацию тестировали с использованием репортерной системы на основе индуцибельного EGFP-экспрессирующего вектора, регулируемого с помощью трансактиватора на основе TetR через элементы оператора Tet, встроенные в минимальный CMV промотор. Трансактиватор Tet заменяли на транскрипционный активатор на основе Cas9, направляемый с помощью sgRNA, нацеленной на повторяющиеся последовательности оператора Tet (фиг. 11a). Разрабатывали две различные sgRNA в мишени, различающиеся только присутствием или

отсутствием дополнительного 5' нуклеотида G, относительно локуса оператора Tet репортерной плазмиды, плюс три дополнительные несовпадающие гидовые на основе той же последовательности в мишени, несущие одну или две мутации в различных положениях вдоль спейсерной последовательности. Более низкие абсолютные уровни экспрессии EGFP наблюдали с использованием evo-dCas9-VP64 по сравнению с dCas9-VP64 как для совпадающих, так и несовпадающих гидовых РНК, что указывает на более сильное связывание dCas9 дикого типа с целевой ДНК (фиг. 11b). Однако, кратность активации в мишени относительно контрольной sgRNA была схожа как для dCas9 дикого типа, так и для evo-dCas9, из-за более низкой фоновой активации, наблюдаемой с evo-dCas9. Этот результат вероятно обусловлен более низкой предрасположенностью evo-Cas9 к связыванию ДНК неспецифически (фиг. 11c). Что интересно, увеличенная специфичность, демонстрируемая с помощью evo-dCas9-VP64, была умеренной, по сравнению с dCas9-VP64, что подсказывает, что evoCas9 связывается с несовпадающими мишенями, хотя и менее эффективно, но тогда неспособен конкурировать в реакции расщепления (фиг. 11c). В общем, эти результаты показывают, что evoCas9 можно использовать для создания транскрипционного регулятора, отличающегося более низкой фоновой активацией, не смотря на меньшую абсолютную силу активации, и слегка увеличенной точностью активации.

#### Долгосрочная специфичность evoCas9

Поскольку постоянная экспрессия Cas9 в клетках ассоциирована с увеличенной активностью вне мишени, важным вопросом, подлежащим исследованию, является долгосрочное поведение evoCas9 в клетках. Чтобы разобраться с этим моментом, лентивирусные частицы создавали для того, чтобы достигать стабильной экспрессии SpCas9 дикого типа, evoCas9 или SpCas9-HF1 вместе с sgRNA, представляющей интерес. Чтобы использовать клеточную модель нокаута EGFP, схожую с той, что ранее использовали для скрининга высокоспецифичных вариантов, проводили эксперименты с использованием того же набора sgRNA, направленных на кодирующую последовательность EGFP, или превосходно совпадающих с целевым локусом или содержащих одно или несколько несовпадений в различных положениях спейсерной последовательности, таким образом действуя в качестве суррогатов вне мишени. Репортерную клеточную линию трансдуцировали равными количествами различных лентивирусных векторов, и отбирали

культуры на всем протяжении всего периода экспериментов для того, чтобы выделять трансдуцированную популяцию и избегать возможного разрежения событий редактирования во времени, вызываемого утратой отредактированных клеток или сниженной годностью трансдуцированных клеток. Аналогично тому, что наблюдали в экспериментах временной трансфекции, снижение флуоресценции EGFP в различные моменты времени обнаруживало, что SpCas9 дикого типа режет целевую последовательность со схожим эффектом, с использованием или превосходно совпадающей sgRNA или несовпадающих sgRNA (см. фиг. 12a и фиг. 5c). Напротив, и evoCas9 и SpCas9-HF1 не расщепляют EGFP эффективно с использованием sgRNA sgGFP1314 и sgGFP1819, обе содержат двойные несовпадения (фиг. 12b-c). Следует отметить, что, хотя потеря флуоресценции была на фоновых уровнях для образцов evoCas9 во все моменты времени, поддающееся измерению число EGFP-отрицательных клеток, которое оставалось постоянным с течением времени, присутствовало в культурах, трансдуцированных с использованием экспрессирующих SpCas9-HF1 векторов (фиг. 12c). Кроме того, при рассмотрении сильнейшего суррогата вне мишени, sgGFP18, содержащего одно несовпадение в PAM-дистальном положении, поведение двух вариантов значительно различалось: хотя SpCas9-HF1 не способен корректно различать совпадающую и несовпадающую sgRNA и расщеплять целевой locus со схожей эффективностью, evoCas9 демонстрировал возрастающую тенденцию к утрате флуоресцентных клеток, которые достигали меньше чем половины уровня нокаута в мишени через 40 суток после трансдукции (фиг. 12b-c). Эти данные подсказывают, что при тщательном выборе гидовых РНК, высокоспецифичный вариант Cas9 по изобретению, evoCas9, делает возможной долгосрочную экспрессию нуклеазы с ограниченным или нулевым нежелательным расщеплением в клеточном геноме.

#### МАТЕРИАЛЫ И СПОСОБЫ

Плазмиды и конструкции. Плазмиду p415-Gall-Cas9-CYC1t использовали для того, чтобы экспрессировать Cas9 в дрожжах (получали из Addgene, №43804) (Di Carlo JE, et al., Nucleic Acids Res. 2013, 41(7):4336-43). Для того чтобы сделать возможным точное удаление домена Rec1-II посредством рестрикционного расщепления, синонимические мутации создавали через ПЦР для того, чтобы вводить два участка рестрикции, NcoI и NheI, выше и ниже домена Rec1-II, соответственно (праймеры см. в



таблице приложения 2). Экспрессирующую кассету для sgRNA получали из плазмиды p426-SNR52p-gRNA.CAN1.Y-SUP4t (получали из Addgene, №43803) (Di Carlo JE, et al., *Nucleic Acids Res.* 2013, 41(7):4336-43). Для того чтобы заменять исходную спейсерную последовательность на желаемую мишень, применяли стратегию сборки на основе ПЦР. 5' часть экспрессирующей sgRNA кассеты подвергали ПЦР амплификации с использованием прямого праймера T3 (отжигаемого перед промотором SNR52) и обратного праймера, отжигаемого непосредственно выше спейсерной последовательности и содержащего 5' свисающий конец, соответствующий желаемой последовательности в мишени (см. таблицу приложения 2). То же самое выполняли для 3' фрагмента sgRNA, используя обратный праймер T7 и прямой праймер, отжигаемый непосредственно после спейсерной последовательности и содержащий 5' свисающий конец, антипараллельный тому, что отмечен ранее. Реакцию сборки для получения кассеты гРНК проводили посредством смешивания обоих ПЦР ампликонов и выполнения единой стадии денатурации, отжига и достраивания, после чего следовала экспоненциальная амплификация с использованием только внешних праймеров T3 и T7. Затем получаемый фрагмент очищали в геле и клонировали с тупыми концами в pRS316 (ATCC), низкокопийную центромерную плазмиду, несущую селективный маркер дрожжей URA3, которую предварительно расщепляли с использованием SacII/XhoI и затупляли, создавая плазмиду pRS316-SNR52p-gRNA.ON-SUP4t.

Для экспрессии SpCas9 в клетках млекопитающих авторы изобретения использовали производную плазмиду pX330-U6-Chimeric\_BB-CBh-hSpCas9 (получали из Addgene, №42230) (Cong L et al., *Science.* 2013, 339(6121):819-23), где удаляли кодирующую sgRNA кассету, pX-Cas9. Проводили оптимизацию кодонов кодирующей последовательности SpCas9 для человека и регулировали ее с помощью промотора CBh. Кроме того, два сигнала ядерной локализации (NLS) добавляли на N- и C-конец SpCas9 для того, чтобы сделать возможным ядерный импорт, и тройную FLAG располагали на N-конце белка для того, чтобы содействовать обнаружению. Плазмиду, кодирующую усовершенствованные варианты Cas9, получали посредством последовательного сайт-специфического мутагенеза, начиная с плазмиды pX-Cas9. Для экспрессии предварительно опубликованных усиленных мутантов SpCas9 использовали плазмиды VP12 (получали из Addgene, №72247) (Kleinstiver BP et al., *Nature.* 2016, 529(7587):490-5) и

eSpCas9(1.1) (получали из Addgene, №71814) (Slaymaker IM et al., Science. 2016, 351(6268):84–8). Желаемые спейсерные последовательности клонировали в виде отожденных олигонуклеотидов с подходящими свисающими концами в двойном участке BbsI выше постоянной части гидовой РНК плазмиды pUC19, содержащей экспрессирующую кассету под управлением промотора U6. Для экспериментов с использованием лентивирусных векторов, вектор для переноса lentiCRISPRv1 (получали из Addgene, №49535) (Cong L et al., Science. 2013, 339(6121):819–23) использовали вместе с пакующим вектором pCMV-delta8.91 (любезный подарок от Didier Trono, EPFL, Switzerland) и pMD2.G (получали из Addgene, №12259), кодирующим гликопротеин вируса везикулярного стоматита (VSVG), чтобы получать вирусные частицы. Вектор для переноса lentiCRISPRv1 содержит экспрессирующую кассету для кодон-гуманизированной версии SpCas9 с N-концевой меткой FLAG, который слит через пептид 2A с кодирующей пуromycin последовательностью, чтобы сделать возможным отбор трансдуцированных клеток. Экспрессирующая кассета под управлением U6 транскрибирует sgRNA. Отжигаемые олиго, соответствующие желаемым спейсерам, клонировали в гидовую РНК с использованием двойного участка BsmBI. Векторы на основе lentiCRISPRv1, кодирующие усиленные варианты SpCas9, создавали посредством замены части кодирующей SpCas9 последовательности на ПЦР фрагмент, соответствующий области CDS, содержащей мутации (праймеры см. в таблице приложения 2). Плазмиду pTRE-GFP получали посредством субклонирования кодирующей последовательности EGFP из плазмиды pEGFP-N1 (Clontech) в клонирующий вектор pTRE-Tight (Clontech). Полный список целевых участков гидовой РНК доступен в приложении. Все олигонуклеотиды приобретали в Eurofins Genomics.

Культура дрожжей. Штамм дрожжей yLFM-ICORE (созданный в лаборатории одного из авторов изобретения из родительского штамма yIG397, любезный подарок от Richard Iggo) (Jegga AG et al., Proc Natl Acad Sci U S A. 2008, 105(3):944–9; Tomso DJ et al., Proc Natl Acad Sci U S A. 2005, 102(18):6431–6) использовали для того, чтобы генерировать репортерные штаммы дрожжей, используемые в этом исследовании. Синтетические минимальные среды (SD) использовали во всех экспериментах на дрожжах (азотистое основание дрожжей без аминокислот 6,7 г/л, L-изолейцин 600 мг/л, L-валин 150 мг/л, L-аденин 200 мг/л, L-аргинин 20 мг/л, L-гистидин 10 мг/л, L-лейцин 100 мг/л, L-лизин

90 мг/л, L-метионин 20 мг/л, L-фенилаланин 50 мг/л, L-треонин 200 г/л, L-триптофан 20 мг/л, L-урацил 20 мг/л, L-глутаминовая кислота 100 мг/л, L-аспарагиновая кислота 200 г/л, L-серин 400 мг/л, D-(+)-глюкоза 20 г/л). Отдельные аминокислоты не использовали в соответствии с постановкой эксперимента, когда требовалась селективная среда. Для индукции p415-GalL-Cas9-SUC1t, 20 г/л D-(+)-галактозы и 10 г/л D-(+)-раффинозы использовали вместо декстрозы. Конкретную среду для цветового скрининга мутантов ADE2 получали с использованием низких концентраций аденина (5 мг/л). Когда требовалась неселективная среда, использовали богатую YPDA среду (дрожжевой экстракт 10 г/л, пептон 20 г/л, D-(+)-глюкоза 20 г/л, L-аденин 200 мг/л). Все растворы получали с использованием ddH<sub>2</sub>O, стерилизовали фильтрованием и хранили при 4°C. Твердые среды получали посредством стерилизации в автоклаве, добавляя 20 г/л агара в раствор. Все химические соединения для получения сред дрожжей получали из Sigma-Aldrich.

Трансформация дрожжей. В сутки перед трансформацией приблизительно 1 мм<sup>3</sup> желаемого штамма дрожжей инокулировали в 5 мл богатой среды или селективной синтетической среды и растили в течение ночи при 30°C и встряхивании на 200 об./мин. В следующие сутки 3–5 мл культуры инокулировали всего в 30 мл той же среды и растили при 30°C и 200 об./мин в течение дополнительных 2–4 часов. Затем клетки собирали посредством центрифугирования на 2000×g в течение 2', промывали в 30 мл ddH<sub>2</sub>O, снова центрифугировали на 2000×g в течение 2' и ресуспендировали в 10 мл LiAc/TE 1× (ацетат лития 0,1 М и Tris 10 mM EDTA 1 mM, pH 7,5). Раствор центрифугировали снова при 2000×g в течение 2' и ресуспендировали в надлежащем объеме LiAc/TE 1× (100 мг пеллета дрожжей в 500 мкл). Смесь для трансформации содержала 500 нг плазмидной ДНК, 5 мкл носителя ДНК спермы лосося (приблизительно 1 мкг, Sigma-Aldrich), предварительно резанной посредством обработки звуком и прокипяченной при 100°C в течение 10', 50 мкл ресуспендированной культуры дрожжей и 300 мкл полиэтиленгликоля (PEG) 500 г/л с молекулярной массой ~36500 (Sigma-Aldrich), разведенного в LiAc/TE 1×. После вортекса, смесь для трансформации помещали на 30' при 30°C и затем осуществляли тепловой шок с использованием сухой ванны в течение 30' при 42°C. Затем клетки центрифугировали на 3000×g в течение 3', ресуспендировали в 5 мл подходящей селективной среды SD или

непосредственно высевали на селективные SD агарозные чашки и инкубировали при 30°C. Для оценки частоты самопроизвольных реверсных мутаций, после трансформации с использованием p415-GalL-Cas9-CYC1t клетки выращивали в селективной среде в течение 24 часов. Затем концентрацию клеток оценивали посредством измерения OD<sub>600</sub> и 1000 клеток высевали на селективные чашки, обедненные по лейцину (SDl), или 10<sup>6</sup> клеток распределяли по чашкам, дополнительно обедненным по аденину (SDla) триптофана (SDlt), чтобы оценивать число ревертантов.

ПЦР колонии дрожжей. PCR колоний осуществляли посредством ресуспендирования приблизительно 1 мм<sup>3</sup> колонии дрожжей в 49 мкл ddH<sub>2</sub>O. 1 мкл литиказы (10000 Ед/мл, Sigma-Aldrich) добавляли для расщепления клеточной стенки, и затем суспензию инкубировали при 30°C в течение 30'. Клетки осаждали, супернатант удаляли и сухой пеллет кипятили в течение 10' при 100°C. Затем пеллет ресуспендировали в 50 мкл ddH<sub>2</sub>O, и использовали 5 мкл в качестве матрицы в реакции ПЦР, используя высокоточную ДНК-полимеразу Phusion (Thermo Scientific).

Извлечение плазмидной ДНК из дрожжей. Для того чтобы выделять плазмиды мутантного Cas9 из дрожжей, отдельные колонии выращивали в течение ночи при 30°C и встряхивали на 200 об./мин в 5 мл среды SD без лейцина (SDl), чтобы осуществлять отбор на присутствие плазмиды p415-GalL-Cas9-CYC1t, при этом ослабляя отбор по плазмиде гидовой РНК, чтобы индуцировать ее разбавление и утрату. В следующие сутки клетки собирали посредством центрифугирования в течение 5' при 5000×g и ресуспендировали в 250 мкл буфера A1 (Nucleospin Plasmid, Macherey-Nagel), содержащего 0,1 мг/мл РНКазы А. Затем клетки механически лизировали посредством добавления 100 мкл стеклянных гранул, промытых кислотой, (Sigma-Aldrich) и вортекса непрерывно в течение 5'. Затем плазмидную ДНК извлекали из супернатанта с использованием стандартных колонок с диоксидом кремния Miniprep по инструкциям производителя. ДНК элюировали в 30 мкл 10 mM TrisHCl pH 8,5. Элюированную ДНК расщепляли с использованием ферментов NcoI и NheI (New England BioLabs), чтобы устранять вектор sgRNA во избежание контаминации последними плазмидами, которые также можно отбирать через устойчивость к ампициллину. После расщепления 10 мкл трансформировали в химически компетентные E. coli. Затем извлеченные плазмиды расщепляли для того, чтобы верифицировать их идентичность и затем осуществляли

секвенирование по Сэнгеру для того, чтобы идентифицировать мутации, вводимые в домен Rec1-II.

Сборка модифицированных TRP1 и ADE2 геномных кассет. ДНК кассеты, используемые для того, чтобы конструировать геномные локусы ADE2 (ADE2-Off1, ADE2-Off2, ADE2-Off3 и ADE2-Off4) и TRP1 (TRP1-On), создавали с использованием схожей стратегии. ПЦР двух различных колоний осуществляли для того, чтобы отдельно амплифицировать две половины каждого локуса дикого типа. В первой использовали прямой праймер выше CDS гена и обратный праймер свисающего конца, содержащий 1-4 последовательность в или вне мишени, после чего следуют участки рестрикции KpnI или BamHI, соответственно (см. таблицу приложения 3). Все обратные праймеры содержали терминирующий кодон перед последовательностью в/вне мишени, чтобы обеспечивать усечение белка. Вторую половину кассеты собирали с использованием обратного праймера, который отжигают ниже кодирующих последовательностей ADE2 и TRP1, и прямого праймера, который отжигают за 100 п. о. до обратного праймера, который использовали для создания первой половины кассеты. Таким образом, когда две части соединяли вместе, конечная конструкция содержала гомологичную область длиной 100 п. о. выше и ниже последовательностей в/вне мишени. Кроме того, эти прямые праймеры содержали тот же участок рестрикции, присутствующий в обратном праймере соответствующей первой половины кассеты. Фрагменты TRP1 и ADE2 собирали посредством лигирования двух половин, расщепленных с использованием KpnI или BamHI (New England BioLabs), соответственно. Продукты разделяли на агарозном геле для того, чтобы удалять фрагменты, полученные гомолигированием. Конечную кассету обогащали посредством ПЦР с использованием самых внешних праймеров и непосредственно трансформировали в дрожжи.

Создание репортерных штаммов дрожжей. Подход *delitto perfetto* делает возможным генетическое направленное воздействие на конкретные локусы с практичностью общей системы отбора через эксплуатацию механизма репарации по гомологии, который особенно эффективен у дрожжей (Stuckey S and Storici F, *Methods Enzymol.* 2013, 533:103-31). Первая стадия состоит в инсерции кассеты Counter selectable REporter I-SceI (CORE-I-SceI) в конкретный локус, представляющий интерес. Кассета содержит участок распознавания для I-SceI, а также кодирующую последовательность для самой эндонуклеазы под управлением индуцируемого галактозой

GAL1 промотора, ген устойчивости kanMX4 (G418) и негативный селективный маркер URA3 ген из *Kluyveromyces lactis* (KlURA3). Кассету CORE-I-SceI амплифицировали с использованием праймеров, содержащих специфические свисающие концы для локусов ADE2 и TRP1 (см. таблицу приложения 3). Каждый локус редактировали последовательно, придерживаясь одной и той же процедуры, начиная с локуса ADE2. 500 нг локус-специфической кассеты CORE-I-SceI трансформировали в дрожжи, и клетки высевали на чашки YPDA и инкубировали при 30°C в течение ночи. В следующие сутки колонии репликами высевали на чашки со средами YPDA, содержащими 200 мкг/мл G418 (Invivogen). Осуществляли скрининг устойчивых колоний по успешной инсерции кассеты в желаемый локус посредством ПЦР колоний с использованием праймеров, отжигаемых к геномным последовательностям, фланкирующим встраиваемый участок, и к кассете. Кассету CORE-I-SceI, встраиваемую в целевой локус, затем заменяли на конечную редактированную последовательность (TRP1-On, ADE2-Off1, ADE2-Off2, ADE2-Off3 и ADE2-Off4), создавая всего четыре различных штамма дрожжей, отличающихся одной и той же последовательностью в мишени, и четыре различных вне мишени. Подходящий промежуточный штамм дрожжей, содержащий целевую кассету CORE-I-SceI, инокулировали в течение ночи в 5 мл YPDA. В следующие сутки, перед трансформацией, инокулят ресуспендировали в 30 мл синтетической среды, содержащей галактозу и раффинозу вместо декстрозы (SRG). Эта стадия важна для того, чтобы индуцировать транскрипцию эндонуклеазы I-SceI, которая режет ее целевой участок, расположенный в кассете CORE. Этот DSB увеличивает нормальную частоту событий репарации, движимых HR, что содействует замене кассеты на желаемую новую последовательность. После 4 часов в SRG, дрожжи трансформировали с использованием 500 нг HR матрицы, содержащей желаемую последовательность, следуя стандартному протоколу трансформации. Затем трансформанты высевали на SD, содержащую 60 мг/л урацила и 1 г/л 5-фтороротовой кислоты (5-FOA) (Toronto Research Chemicals). 5-FOA в присутствии оротидин-5'-фосфатдекарбоксилазы (ее кодирует KlURA3) превращают во фторурацил, который представляет собой мощный ингибитор тимидилатсинтазы. 5-FOA-устойчивые колонии после этого репликами высевали на YPDA и YPDA с добавлением G418, чтобы дополнительно проводить отбор по утрате кассеты CORE. Посредством сравнения двух чашек с репликами можно отбирать G418-чувствительные FOA-устойчивые

колонии, которые соответствуют положительным клонам. PCR колоний, которую выполняли с использованием геномных праймеров, которые отжигают выше и ниже полного геномного локуса, анализировали посредством секвенирования по Сэнгеру для того, чтобы подтверждать последовательность редактированного локуса. Вновь созданные штаммы дрожжей, содержащие модифицированные локусы TRP1 и ADE2, которые называли  $\gamma$ ACMO-off1,  $\gamma$ ACMO-off2,  $\gamma$ ACMO-off3 и  $\gamma$ ACMO-off4, отличались отобранной последовательностью в мишени в локусе TRP1 и четырьмя различными последовательностями вне мишени в локусе ADE2, каждая содержит одно не совпадение по отношению к последовательности в мишени в положении, которое является более PAM-проксимальным для off1 и более PAM-дистальным для off4 (см. таблицу приложения 4).

Прочтение дрожжевого анализа. Гены ADE2 и TRP1 являются ключевыми ферментами в метаболических путях, ведущих к получению аденина и триптофана, и по этой причине их нокаут уничтожает способность дрожжей расти на среде, истощенной по двум связанным питательным веществам. Штаммы дрожжей  $\gamma$ ACMO, которые создавали в этом исследовании, имеют дефицит активности генов TRP1 и ADE2, пока Cas9 не разрезает целевую последовательность, нарушающую каждую кодирующую последовательность: опосредованная отжигом одной цепи рекомбинация между двумя гомологичными областями 100 п. о. по обеим сторонам целевых участков обеспечивает восстановление локуса дикого типа. Затем можно использовать скрининг на основе ауксотрофного отбора для того, чтобы оценивать активность расщепления Cas9 в двух геномных локусах, измеряя события как в мишени, так и вне мишени. После трансформации с использованием p415-Gall-Cas9-CYC1t и pRS316-SNR52p-gRNA.ON-SUP4t, клетки выращивали в синтетической среде без лейцина и урацила (SDlu) в течение 4 часов перед индукцией в течение ночи в галактозусодержащей среде (SRGlu). Затем клетки высевали равным числом на чашки SDlu, чтобы измерять общее число трансформантов, и на чашки SDlu, обедненные по триптофану и с низкой концентрацией аденина (SDluta<sub>5</sub>), чтобы различать колонии, в которых Cas9 расщеплял только последовательности в или также вне мишени. В частности, при наблюдении считываемой чашки SDluta<sub>5</sub> могут присутствовать различные фенотипы: отсутствие роста указывает на отсутствие редактирования локуса TRP1 (TRP1-/ADE2+/-); рост красной колонии (TRP1+/ADE2-) указывает на редактирование только в локусе TRP1, без обнаружения активности

вне мишени; рост белой колонии (TRP1<sup>+</sup>/ADE2<sup>+</sup>) указывает на редактирование обоих локусов TRP1 и ADE2, с обнаружением расщеплений вне мишени. Типичную красную пигментацию колонии определяют по накоплению в клеточной вакуоли промежуточного продукта биосинтетического пути аденина, образуемого путем блокирования на уровне продукта гена ADE2. Посредством сравнения общего числа колоний, получаемого на SDlu и SDluta<sub>5</sub>, можно измерять эффективность расщепления в мишени, при этом посредством определения процентной доли красных и белых колоний на чашке SDluta<sub>5</sub> можно получать оценку специфичности активности Cas9 в отношении тестируемой последовательности вне мишени.

Дрожжевой скрининг мутантов SpCas9. Библиотеку мутантов создавали посредством допускающей ошибки ПЦР (epPCR) с использованием набора GeneMorph II (Agilent). По инструкциям производителя, начальное количество матричной ДНК (p415-Gall-Cas9-CYC1t) и число циклов задавали для того, чтобы получать усредненно 5 мутаций на тысячу оснований. Праймеры длиной 50 п. о. отбирали для отжига на 150 п. о. выше и ниже кодирующей последовательности REC1-II (см. таблицу приложения 5). ПЦР библиотеку непосредственно собирали *in vivo* посредством совместной трансформации пула мутировавших ампликонов с плазмидой p415-Gall-Cas9-CYC1t, предварительно расщепленной с использованием NcoI и NheI (New England BioLabs) для того, чтобы удалять домен REC1-II, с соотношением вставка/плазмида 3:1. Две гомологичные области 150 п. о. на обоих концах ампликонов использовали с помощью дрожжей для того, чтобы репарировать расщепленную плазмиду посредством гомологичной рекомбинации, таким образом встраивая мутировавшую часть. Клоны, содержащие мутации в этих фланкирующих областях 150 п. о., вероятно отрицательно отбирали в течение этой стадии сборки *in vivo* благодаря утрате полной гомологии. Тем не менее, эти мутации лежали вне области, представляющей интерес для авторов изобретения (домен REC1-II). Скрининг мутагенной библиотеки осуществляли параллельно с ее сборкой посредством совместной трансформации фрагментов в штамме дрожжей yACMO-off4, стабильно экспрессирующем sgRNA, совпадающую с последовательностью в мишени, содержащейся в локусе TRP1. После трансформации, культуру растили в течение ночи в среде SD, содержащей урацил и лейцин (SDlu, для отбора клеток, несущих sgRNA- и Cas9-экспрессирующие плазмиды) для того, чтобы сделать возможным



восстановление и правильную рекомбинацию. В следующие сутки экспрессию Cas9 индуцировали посредством выращивания культуры в галактозосодержащей среде (SRGlu) в течение 5 часов перед высеванием на несколько селективных чашек, не содержащих триптофан и содержащих низкую концентрацию аденина (SDluta<sub>5</sub>), чтобы различать колонии в соответствии с состоянием редактирования локусов TRP1 и ADE2. После 48 часов TRP1<sup>+</sup>/ADE2<sup>-</sup> (красные) колонии штрихами высевали на селективную чашку с низким аденином и без триптофана, содержащую галактозу и раффинозу (SRGluta<sub>5</sub>), чтобы сохранять конститутивно индуцированную экспрессию Cas9 и усиливать образование расщеплений вне мишени. После дополнительных 48 часов инкубации, Cas9-экспрессирующие плазмиды извлекали из самых красных штрихов, соответствующих колониям, в которых Cas9 расщеплял только участок в мишени, и характеристики мутаций определяли секвенированием по Сэнгеру.

Цветовой анализ колоний дрожжей и количественное определение. Все изображения чашек получали с использованием Canon EOS 1100D (1/60, f/9,0 и ISO 800) и анализировали с использованием OpenCFU (Geissmann Q, PLoS One. 2013;8(2):e54072). Для всех изображений использовали инвертированный порог (значение=2) с радиусом между 8 и 50 пикселями. Вычисляли разграничение между белыми и красными колониями, вычисляя усредненный сигнал в каналах RGB и задавая ручной порог, который точно проводит различия между красными и белыми колониями в каждом эксперименте.

Клетки млекопитающих и трансфекции. Клетки 293T/17 получали из American Type Culture Collection (получали из ATCC) и культивировали в модифицированной по способу Дульбекко среде Игла (DMEM; Life Technologies) с добавлением 10% эмбриональной телячьей сыворотки (FCS; Life Technologies) и антибиотиков (PenStrep, Life Technologies). Клетки 293multiGFP создавали посредством стабильной трансфекции с использованием pEGFP-IRES-Puromycin и отбирали с использованием 1 мкг/мл пурамицина. 293blastEGFP получали посредством инфекции с низкой MOI клеток HEK293T с использованием EGFP-экспрессирующего лентивирусного вектора pAIB-GFP, после чего следовал клональный отбор с 5 мкг/мл бластицидина. Для трансфекции  $1 \times 10^5$  клеток 293multiGFP или 293T/лунка высевали в 24-луночных планшеты и осуществляли трансфекцию в следующие сутки с использованием TransIT-LT1

(Mirus Bio) в соответствии с протоколом производителя, используя 400–750 нг Cas9-экспрессирующих плазмид и 200–250 нг sgRNA-экспрессирующих плазмид. Для экспериментов по временной трансфекции, включающих экспрессию EGFP, использовали 100 нг плазмиды pEGFP-N1. Для того чтобы определять уровень понижающей регуляции EGFP с помощью Cas9 после трансфекции в 293multiGFP, клетки собирали через 7 суток после трансфекции и анализировали посредством проточной цитометрии с использованием FACSCanto (BD Biosciences).

Получение лентивирусного вектора и трансдукция. Лентивирусные частицы получали посредством высевания  $4 \times 10^6$  клеток 293T в 10 см чашку. В следующие сутки чашки трансфицировали с использованием 10 мкг каждого вектора для переноса на основе lentiCRISPR (Cong L et al., Science. 2013, 339(6121):819–23) вместе с 6,5 мкг пакующего вектора pCMV-deltaR8.91 и 3,5 мкг pMD2.G, используя способ с полиэтиленимином (PEI). После инкубации в течение ночи среду заменяли на свежую полную DMEM и через 48 часов собирали супернатант, содержащий вирусные частицы, центрифугировали на  $500 \times g$  в течение 5 минут и фильтровали через 0,45 мкм PES фильтр. Количественное определение титров вектора осуществляли с использованием способа SG-PERT (Pizzato M et al., J Virol Methods. 2009, 156(1–2):1–7). Стоковые растворы векторов сохраняли при  $-80^\circ\text{C}$  для будущего использования.

Для трансдукции  $10^5$  клеток 293blastGFP высевали в 24-луночный планшет и в следующие сутки трансдуцировали с использованием 0,4 единиц обратной транскриптазы (RTU)/луночка для каждого вектора посредством центрифугирования на  $1600 \times g$  при  $16^\circ\text{C}$  в течение 2 часов. После инкубации в течение ночи, удаляли вирусный супернатант и клетки содержали в культуре в течение всего 48 часов перед добавлением 0,5 мкг/мл пуромицина для отбора, который поддерживали на всем протяжении эксперимента. Для того чтобы определять уровень понижающей регуляции EGFP посредством Cas9 после инфекции, клетки 293blastGFP собирали в указанные моменты времени после трансдукции и анализировали посредством проточной цитометрии с использованием FACSCanto (BD Biosciences).

Обнаружение Cas9-индуцированных геномных мутаций. Геномную ДНК получали через 7 суток после трансфекции, используя раствор для экстрагирования ДНК QuickExtract (Epicentre). ПЦР реакции

для амплификации геномных локусов осуществляли с использованием высокоточной ДНК-полимеразы Phusion (Thermo Fisher). Образцы амплифицировали с использованием олиго, перечисленных в таблице приложения 8. Очищенные продукты ПЦР анализировали посредством секвенирования и применения инструмента TIDE (Brinkman EK et al., *Nucleic Acids Res.* 2014, 42(22):e168). Для того чтобы количественно определять хромосомную делецию CCR2-CCR5, использовали подход полуколичественной ПЦР с праймерами, фланкирующими участок CCR5 в мишени и локус CCR2 вне мишени (таблица приложения 8). Число циклов ПЦР модулировали для того, чтобы не достигать плато амплификации. Количественное определение выполняли посредством денситометрических анализов с использованием программного обеспечения ImageJ и геномного локуса FANCF в качестве внутреннего нормализатора.

Вестерн-блоттинг. Клетки лизировали в буфере NEHN (20 mM HEPES pH 7,5, 300 mM NaCl, 0,5% NP40, NaCl, 1 mM EDTA, 20% глицерин с добавлением 1% коктейля ингибиторов протеаз (Pierce)). Клеточные экстракты разделяли посредством SDS-PAGE с использованием PageRuler Plus Protein Standards в качестве стандартных маркеров молекулярной массы (Thermo Fisher Scientific). После электрофореза, образцы переносили на 0,22 мкм PVDF мембраны (GE Healthcare). Мембраны инкубировали с мышинным против FLAG (Sigma) для обнаружения SpCas9 и различных высокоточных вариантов, с мышинным против  $\alpha$ -тубулина (Sigma) для контроля загрузки и с подходящими конъюгированными с HRP вторичными антителами козы против мыши (KPL) для обнаружения ECL. Изображения получали с использованием системы обнаружения UVitec Alliance.

Направленное глубокое секвенирование. Выбранные участки вне мишени для геномных локусов VEGFA3 и EMX1, вместе с соответствующими им в мишени, амплифицировали с использованием высокоточной полимеразы Phusion (Thermo Scientific) или полимеразы EuroTaq (Euroclone) из геномной ДНК 293T, извлеченной на 7 сутки после трансфекции с использованием SpCas9 дикого типа или evoCas9 вместе с sgRNA, направленными на локусы EMX1 и VEGFA3, или a pUC empty вектор. Перед очисткой и индексацией ампликоны вне мишени объединяли в концентрациях, близких к эквимольным. Библиотеки индексировали посредством ПЦР с использованием индексов Nextera (Illumina), количественно определяли с использованием набора Qubit dsDNA High Sensitivity

Assay (Invitrogen), объединяли в соответствии с числом мишеней и секвенировали в системе Illumina Miseq с использованием набора реактивов Illumina Miseq V3-150 циклов (одно прочтение 150 п. о.). Полный список праймеров, используемых для того, чтобы генерировать ампликоны, приведен в таблице приложения 7.

Эталонный геном создавали с использованием Picard (<http://broadinstitute.github.io/picard>) и samtools (Li H et al., *Bioinformatics*. 2009, 25(16):2078-9) из последовательностей ДНК рассматриваемых областей в/вне мишени. Необработанные данные секвенирования (файлы FASTQ) картировали с привязкой к созданному эталонному геному с использованием BWA-MEM (Li H and Durbin R, *Bioinformatics*. 2010, 26(5):589-95) и стандартных параметров, а получаемые файлы выравниваний сортировали с использованием samtools. Сохраняли только прочтения с качеством картирования выше или равным 30. Присутствие инсерций-делеций в каждом прочтении для каждой рассматриваемой области определяли посредством поиска инсерций-делеций размером 1 п. о., непосредственно смежных с предсказанным сайтом расщепления, или инсерций-делеций размером  $\geq 2$  п. о., перекрывающих фланкирующие области размером 5 п. о. около предсказанного сайта расщепления.

Эксперименты GUIDE-seq и анализ данных. GUIDE-seq осуществляли как описано ранее (Bolukbasi MF et al., *Nat Methods*. 2015, 12(12):1150-6) с некоторыми модификациями. В кратком изложении,  $2 \times 10^5$  клеток 293T трансфицировали с использованием 750 нг экспрессирующей плазмиды Cas9, вместе с 250 нг sgRNA-кодирующей плазмиды или пустой плазмиды pUC19 и 10 пмоль затравки dsODN, содержащей фосфотиоатные связи на обоих концах (разработано в соответствии с исходным протоколом GUIDE-seq), используя реактив для трансфекции липофектамин 3000 (Invitrogen). Через 3 суток после трансфекции геномную ДНК экстрагировали с использованием набора DNeasy Blood and Tissue (Qiagen) по инструкциям производителя и резали до усредненной длины 500 п. о. с использованием устройства для обработки звуком Bioruptor Pico (Diagenode). Получение библиотек осуществляли с использованием исходных адаптеров и праймеров в соответствии с предыдущей работой (Tsai SQ et al., *Nat Biotechnol*. 2015, 33(2):187-97). Библиотеки количественно определяли с использованием набора Qubit dsDNA High Sensitivity Assay (Invitrogen) и секвенировали с использованием системы секвенирования MiSeq (Illumina) и набора реактивов Illumina

Miseq V2-300 циклов (2×150 п. о. спаренный конец).

Необработанные данные секвенирования (файлы FASTQ) анализировали с использованием вычислительного конвейера GUIDE-seq (Tsai SQ et al., Nat Biotechnol. 2015, 33(2):187-97). После демультиплексирования, предполагаемые ПЦР дубликаты объединяли в единые прочтения. Объединенные прочтения картировали на эталонном геноме человека GrCh37 с использованием BWA-MEM (Li H and Durbin R, Bioinformatics. 2010, 26(5):589-95); отфильтровывали прочтения с качеством картирования ниже 50. При идентификации геномных областей, интегрирующих двухцепочечных олигодезоксинуклеотид (dsODN) в выровненные данные, сохраняли участки RGN в случае присутствия самое большее 8 несовпадений с мишенью и в случае отсутствия в фоновых контролях. Визуализация выровненных участков вне мишени доступна в виде сетки последовательностей с цветовым кодированием.

#### Приложение

Таблица 1. Мутанты, получаемые при дрожжевом скрининге, относительно фиг. 4.

<b>Мутант</b>	<b>Аминокислотные замены</b>
A0	R403H, N612Y, L651P, K652E, G715S
A1	R403H, N612Y, L651P, K652E, G715S
A2	W659R, R661L
A3	L540Q, L607P
A4	R400H, Y450S
A5	T472A, P475H, A488V
A6	N407P, F498I, P509L
A7	D406Y
A8	A488V, D605V, R629G, T657A
A9	S487Y, N504S, E573D
A10	K377E, L598P, L651H
A11	T496A, N609D, A728G
A12	Y450H, F553L, Q716H
B1	K484M, Q695H, Q712R
B2	L683P
B3	P449S, F704S
B4	S675C, Q695L
B5	W464L

B6	I473F, D550N, Q739E
B7	A421V, R661W
B8	M495V, K526N, S541P, K562E
B9	D406V, E523K
B10	N407H, K637N, N690I
B11	M495V, K526N, S541P, K562E
B12	I679V, H723L
B13	N522K, G658E
B14	L423P, M465R, Y515N, K673M
B15	R654H, R691Q, H698Q
C1	L683P
C2	T474A
C3	L683P
C4	L683P
C5	W476R, L738P
C6	W476R, L738P
C7	W476R, L738P
C8	W476R
C9	L683P
C10	L683P
C11	L683P
C12	L683P
C13	K526E
C14	F405L, F518L, L651P, I724V
C15	K526E
C16	E470D, I548V, A589T, Q695H
C17	E470D, I548V, A589T, Q695H
C18	E470D, I548V, A589T, Q695H
C19	Y450N, H698P, Q739K
C20	K526E
C21	F405L, F518L, L651P, I724V
C22	F405L, F518L, L651P, I724V
C23	D397E, Y430C, L666P
C24	L380*, I473V

C25	Сдвиг рамки, N522S, K646R, D686V
C26	Сдвиг рамки, N522S, K646R, D686V
C27	Сдвиг рамки, N522S, K646R, D686V
C28	Сдвиг рамки, N522S, K646R, D686V
C29	R691L, H721R, I733V
C30	Q402R, V561M, Q695H
C31	Q402R, V561M, Q695H
C32	Q402R, V561M, Q695H
C33	F478Y, N522I, L727H
C34	Сдвиг рамки, A640T, Q709L

\* = терминирующий кодон

Таблица 2. Праймеры, используемые для клонирования плазмиды

Название праймера	Последовательность (5' - ... -3')
RecI-II- NheI-F	CCAGAAAGCACAAAGTTGCTAGCCAGGGGGACAGTC (SEQ ID NO: 4)
RecI-II- NheI-R	GACTGTCCCCCTGGCTAGCAACTTGTGCTTTCTGG (SEQ ID NO: 5)
RecI-II- NcoI-F	CAGCGCACTTTTCGACCATGGAAGCATCCCCCA (SEQ ID NO: 6)
RecI-II- NcoI-R	TGGGGGATGCTTCCATGGTCGAAAGTGCGCTG (SEQ ID NO: 7)
T3-Forward	AATTAACCCTCACTAAAGGG (SEQ ID NO: 8)
T7-Reverse	TAATACGACTCACTATAGGG (SEQ ID NO: 9)
sgRNA- Ontarget-F	CTCGTGACCACCCTGACCTAGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO: 10)
sgRNA- Ontarget-R	TAGGTCAGGGTGGTCACGAGGATCATTTATCTTTCACTGCG (SEQ ID NO: 11)
Apa- ZhangCas-F	ACGTGGGCCCTCTGGCCAG (SEQ ID NO: 12)
Nhe- ZhangCas-R	TACGCTAGCTCCCTTTTTCTTTTTTGCTGG (SEQ ID NO: 13)
Apa- JoungCas-F	ATTAGGGCCCCCTGGCCCGAGGGAAC (SEQ ID NO: 14)
Spe-	TAATACTAGTGACTTTCCTCTTCTTCTGGG (SEQ ID NO: 15)

JoungCas-R	
------------	--

Таблица 3. Праймеры, используемые в конструкции репортерной кассеты дрожжей

Название праймера	Последовательность (5' - ... -3')
TRP1- genomic-F	ССААGAGGGAGGGCATTGG (SEQ ID NO: 16)
TRP1Pt1- ON-Kpn-R	TGCGGTACCGTAGGTCAGGGTGGTCACGAGTTAGAGGAACTCTTGGTATTCT TGC (SEQ ID NO: 17)
TRP1Pt2- Kpn-F	TTAGGTACCGTAATCAACCTAAGGAGGATGTTT (SEQ ID NO: 18)
TRP1- genomic-R	TGCTTGCTTTTCAAAGGCCTG (SEQ ID NO: 19)
ADE2- genomic-F	TGCCTAGTTTCATGAAATTTTAAAGC (SEQ ID NO: 20)
ADE2Pt1- OFF1-Bam- R	CCAGGATCCGGAGGTCAGGGTGGTCACGAGTTAGACGCAAGCATCAATGGTA T (SEQ ID NO: 21)
ADE2Pt1- OFF2-Bam- R	CCAGGATCCGTAGGTAAGGGTGGTCACGAGTTAGACGCAAGCATCAATGGTA T (SEQ ID NO: 22)
ADE2Pt1- OFF3-Bam- R	CCAGGATCCGTAGGTCAGGGCGGTCACGAGTTAGACGCAAGCATCAATGGTA T (SEQ ID NO: 23)
ADE2Pt1- OFF4-Bam- R	CCAGGATCCGTAGGTCAGGGTGGTAACGAGTTAGACGCAAGCATCAATGGTA T (SEQ ID NO: 24)
ADE2Pt2- Bam-F	ATTAGGATCCTGGTGTGGAAATGTTCTATTTAG (SEQ ID NO: 25)
ADE2- genomic-R	GTAATCATAACAAAGCCTAAAAAATAG (SEQ ID NO: 26)
TRP1- CORE-F	TATTGAGCACGTGAGTATACGTGATTAAGCACACAAAGGCAGCTTGGAGTTA GGGATAACAGGGTAATTTGGATGGACGCAAAGAAGT (SEQ ID NO: 27)
TRP1- CORE-R	TGCAGGCAAGTGCACAAACAATACTTAAATAAATACTACTCAGTAATAACTT CGTACGCTGCAGGTCGAC (SEQ ID NO: 28)



ADE2 <sup>-</sup> CORE- F	CCTACTATAACAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAGTTA GGGATAACAGGGTAATTTGGATGGACGCAAAGAAGT (SEQ ID NO: 29)
ADE2 <sup>-</sup> CORE- R	ATATCATTTTTATAATTATTTGCTGTACAAGTATATCAATAAACTTATATATT CGTACGCTGCAGGTCGAC (SEQ ID NO: 30)

Таблица 4. Участки в и вне мишени, используемые для того, чтобы генерировать штаммы уАСМО

<b>Название мишени</b>	<b>Последовательность (5' – ... –3', РАМ в нижнем регистре), несовпадения полужирным</b>
TRP1-on	STCGTGACCACCCTGACCTAcgg (SEQ ID NO: 31)
ADE2-off1	STCGTGACCACCCTGACCT <b>C</b> cgg (SEQ ID NO: 32)
ADE2-off2	STCGTGACCACCCT <b>T</b> ACCTAcgg (SEQ ID NO: 33)
ADE2-off3	STCGTGACC <b>G</b> CCCTGACCTAcgg (SEQ ID NO: 34)
ADE2-off4	STCGT <b>T</b> ACCACCCTGACCTAcgg (SEQ ID NO: 35)

Таблица 5. Праймеры для подверженной ошибкам ПЦР

<b>Название праймера</b>	<b>Последовательность (5' – ... –3')</b>
epPCR-F	GTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAG G (SEQ ID NO: 36)
epPCR-R	TCTCGGGCCATCTCGATAACGATATTCTCGGGCTTATGCCTTCCCATTA C (SEQ ID NO: 37)

Таблица 6. Спейсерные последовательности, используемые для того, чтобы получать sgRNA для репортерного анализа

<b>Название мишени</b>	<b>Спейсерная последовательность (5' – ... –3')</b>
GFPon	GGGCACGGGCAGCTTGCCGG (SEQ ID NO: 38)
GFP1314	GGGCAC <b>CC</b> GCAGCTTGCCGG (SEQ ID NO: 39)
GFP1819	<b>GCC</b> CACGGGCAGCTTGCCGG (SEQ ID NO: 40)
GFP18	GG <b>CC</b> CACGGGCAGCTTGCCGG (SEQ ID NO: 41)
Участок 2 GFP	GTCGCCCTCGAACTTCACCT (SEQ ID NO: 42)
Участок 14 GFP	GAAGGGCATCGACTTCAAGG (SEQ ID NO: 43)
Участок 16 GFP	GCTGAAGCACTGCACGCCGT (SEQ ID NO: 44)
Участок 18 GFP	GACCAGGATGGGCACCACCC (SEQ ID NO: 45)
Участок 20 GFP	GAAGTTCGAGGGCGACACCC (SEQ ID NO: 46)
GFPon-19nt	GGCACGGGCAGCTTGCCGG (SEQ ID NO: 47)
GFPB-18nt	GGCAAGCTGCCCGTGCCC (SEQ ID NO: 48)
GFPW-17nt	GTGACCACCCTGACCTA (SEQ ID NO: 49)

TetO-on	GTGATAGAGAACGTATGTCTG (SEQ ID NO: 50)
TetO-on+G	gGTGATAGAGAACGTATGTCTG (SEQ ID NO: 51)
TetO-off6	GTGATAGAGAACGT <b>CT</b> GTCTG (SEQ ID NO: 52)
TetO-off1314	GTGATA <b>CT</b> GAAACGTATGTCTG (SEQ ID NO: 53)
TetO-off1819	<b>GAC</b> ATAGAGAACGTATGTCTG (SEQ ID NO: 54)

5' – несовпадающие нуклеотиды показаны в нижнем регистре. Мутации показаны полужирным.

Таблица 7. Олиго для направленного глубокого секвенирования

Общий прямой свисающий конец: 5' –  
TCGTCCGGCAGCGTCAAGATGTGTATAAGAGACAG-3' (SEQ ID NO: 55)

Общий обратный свисающий конец: 5' –  
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3' (SEQ ID NO: 56)

Локус	Мишень	Прямой (5' – ... – 3')	Обратный (5' – ... – 3')
EMX1-on	GAGTCCGAGCAGAAGA AGAAggg (SEQ ID NO: 57)	CCGGAGGACAAAGTAC AAACGGC (SEQ ID NO: 58)	AAGCAGCACTCTGCCC TCGTG (SEQ ID NO: 59)
EMX1-ot1	GAGTT <b>A</b> GAGCAGAAGA AGAAagg (SEQ ID NO: 60)	CTTTTATACCATCTTG GGGTACAG (SEQ ID NO: 61)	CTAGGAAAGATTAACA GAGAGTCTGAC (SEQ ID NO: 62)
EMX1-ot2	GAGT <b>CTA</b> GCAGAAGA AGAAgag (SEQ ID NO: 63)	CAATGTGCTTCAACCC ATCACGGC (SEQ ID NO: 64)	CCTCTACTTCATTGTA CTCAAGGTAAG (SEQ ID NO: 65)
EMX1-ot3	<b>AAGTCTG</b> AGCACAAGA AGAAtgg (SEQ ID NO: 66)	TAGTTCTGACATTCCT CCTGAGGG (SEQ ID NO: 67)	CTCTGTTGTTATTTTT TGGTCAATATCTG (SEQ ID NO: 68)
EMX1-ot4	GAGTCC <b>TAGCAGG</b> AGA AGAAgag (SEQ ID NO: 69)	AAAGCCTGGAGGCTGC CAGGT (SEQ ID NO: 70)	ATCTAGCTGTCCTGTC TCATTGGC (SEQ ID NO: 71)
EMX1-ot5	GAG <b>G</b> CCGAGCAGAAGA <b>AAGAc</b> gg (SEQ ID NO: 72)	CAGGAGCCGGGTGGGA G (SEQ ID NO: 73)	CCTCAGCCTTCCCTCA GCCAC (SEQ ID NO: 74)
VEGFA3-on	GGTGAGTGAGTGTGTG CGTGtgg (SEQ ID NO: 75)	CTGGGTGAATGGAGCG AGCAG (SEQ ID NO: 76)	GGAAGGCGGAGAGCCG GACA (SEQ ID NO: 77)
VEGFA3-	<b>AGT</b> GAGTGAGTGTGTG	GAAGGGGAGGGGGAAG	CGTGCCTGCCCGCTT

ot1	<b>TGTG</b> ggg (SEQ ID NO: 78)	TCACC (SEQ ID NO: 79)	GATC (SEQ ID NO: 80)
VEGFA3- ot2	<b>TGTGGGTGAGTGTGTG</b> CGT <b>G</b> agg (SEQ ID NO: 81)	TCTGTCACCACACAGT TACCACC (SEQ ID NO: 82)	GTAGTTGCCTGGGGAT GGGGTATG (SEQ ID NO: 83)
VEGFA3- ot3	<b>GCTGAGT</b> GAGTGT <b>ATG</b> CGT <b>Gt</b> gg (SEQ ID NO: 84)	CACCTGGCCCATTTCT CCTTTGAGG (SEQ ID NO: 85)	TGGGGACAGCATGTGC AAGCCACA (SEQ ID NO: 86)
VEGFA3- ot4	GGTGAGT <b>GAGTGTGTG</b> <b>TGTG</b> agg (SEQ ID NO: 87)	GGACCCCTCTGACAGA CTGCA (SEQ ID NO: 88)	CACACACCCTCACATA CCCTCAC (SEQ ID NO: 89)
VEGFA3- ot5	<b>AGAGAGT</b> GAGTGT <b>GTG</b> C <b>AT</b> Gagg (SEQ ID NO: 90)	GGAAGAATGCAAAGGA GAAGCAAGTAC (SEQ ID NO: 91)	GACCTGGTGGGAGTTG ATTGGATC (SEQ ID NO: 92)
VEGFA3- ot6	<b>AGTGTGTGAGTGTGTG</b> CGT <b>Gt</b> gg (SEQ ID NO: 93)	CCTTGGAATCTATCT TGAATAGGCCT (SEQ ID NO: 94)	GACACCCACACACTC TCATGC (SEQ ID NO: 95)
VEGFA3- ot7	<b>TGTGAGT</b> <b>AAGTGTGTG</b> <b>TGTG</b> tgg (SEQ ID NO: 96)	CCTAAGCTGTATGTGA GTCCCTGA (SEQ ID NO: 97)	CTGTTTTGCTAAGAGA TGATTAGATGGTC (SEQ ID NO: 98)
VEGFA3- ot8	<b>GTTGAGTGA</b> <b>ATGTGTG</b> CGT <b>G</b> agg (SEQ ID NO: 99)	GCCCTCTCCGGAAGTG CCTTG (SEQ ID NO: 100)	GAAGGGTTGGTTTGGGA AGGCTGTC (SEQ ID №101)
VEGFA3- ot9	GGTGAGT <b>GAGTGC</b> GTG CG <b>GG</b> tgg (SEQ ID NO: 102)	CCACAGGAATTTGAAG TCCGTGCT (SEQ ID NO: 103)	CCCCACGTCCACCCAT ACACAC (SEQ ID NO: 104)
VEGFA3- ot10	<b>AGCGAGTGGTGTGTG</b> CGT <b>G</b> gg (SEQ ID NO: 105)	GACGTCTGGGTCCCGA GCAGT (SEQ ID NO: 106)	GGCCGTCAGTCGGTCC CGA (SEQ ID NO: 107)
VEGFA3- ot11	<b>TGTGAGT</b> GAGTGT <b>GTG</b> CGT <b>Gtga</b> (SEQ ID NO: 108)	GGAGGGTTGAACTGTG ACAGAACTG (SEQ ID NO: 109)	TGAGTATGTGTGAGTG AGAGTGTGCA (SEQ ID NO: 110)
VEGFA3- ot12	<b>ACTGTGTGAGTGTGTG</b> CGT <b>G</b> agg (SEQ ID NO: 111)	GATCCTTAGGCGTGCG TGTGC (SEQ ID NO: 112)	CACCGGCACAGTGACA CTCAC (SEQ ID NO: 113)

VEGFA3- ot13	<b>T</b> GTGAGTGAGTGTGTG <b>TAT</b> Gggg (SEQ ID NO: 114)	AGACCTTCAATGTGGA TGTGCGTG (SEQ ID NO: 115)	CATAGAGTGTAGCAGA TTCCATAACTTC (SEQ ID NO: 116)
-----------------	---	--	--

Мутации показаны полужирным. РАМ показаны в нижнем регистре.

Таблица 8. Спейсерные последовательности для геномных мишеней и олиго для амплификации геномных локусов

Локус	Мишень	Прямой (5'-. . .- 3')	Обратный (5'-. . .- 3')
VEGFA3	GGTGAGTGAGTGTGTG CGTGTGG (SEQ ID NO: 75)	GCATACGTGGGCTCCA ACAGGT (SEQ ID NO: 117)	CCGCAATGAAGGGGAAG CTCGA (SEQ ID NO: 118)
ZSCAN2	GTGCGGCAAGAGCTTC AGCCGGG (SEQ ID NO: 119)	GACTGTGGGCAGAGGT TCAGC (SEQ ID NO: 120)	TGTATACGGGACTTGAC TCAGACC (SEQ ID NO: 121)
EMX1-K	GAGTCCGAGCAGAAGA AGAAGGG (SEQ ID NO: 57)	CTGCCATCCCCTTCTG TGAATGT (SEQ ID NO: 122)	GGAATCTACCACCCAG GCTCT (SEQ ID NO: 123)
VEGFA2	GACCCCTCCACCCCG CCTCCGG (SEQ ID NO: 124)	TCAGCGGACTCACCGG CCAG (SEQ ID NO: 125)	GCGCCGAGTCGCCACTG CGG (SEQ ID NO: 126)
FANCF2	GCTGCAGAAGGGATTC CATGAGG (SEQ ID NO: 127)	GCCAGGCTCTCTTGGA GTGTC (SEQ ID NO: 128)	AGCATAGCGCCTGGCAT TAATAGG (SEQ ID NO: 129)
FANCF2- OT1	GCTGCAGAAGGGATTC <b>CA</b> AGAGG (SEQ ID NO: 130)	CCCCTGAGGTG CTGAGATTTGAAC (SEQ ID NO: 131)	CACATGGAGGAG GTGACGCTG (SEQ ID NO: 132)
CCR5sp1 1	GGTACCTATCGATTGT CAGGAGG (SEQ ID NO: 133)	ATGCACAGGGTGGAAC AAGATGGA (SEQ ID NO: 134)	CTAAGCCATGTGCACAA CTCTGAC (SEQ ID NO: 135)
CCR2	GGT <b>A</b> TCTATCGATTGT CAGGAGG (SEQ ID NO: 136)	CATTGTGGGCTCACTC TGCTGCA (SEQ ID NO: 137)	CTGAGATGAGCTTTCTG GAGAGTCA (SEQ ID NO: 138)

Несовпадающие 5'-G показаны в нижнем регистре. Мутации показаны полужирным. РАМ включены в целевые последовательности.

**ФОРМУЛА ИЗОБРЕТЕНИЯ**

1. Выделенная модифицированная молекула Cas9, содержащая по меньшей мере одну мутацию, расположенную в положении аминокислотного остатка, выбранном в группе, состоящей из K526, K377, E387, D397, R400, D406, A421, L423, R424, Q426, Y430, K442, P449, V452, A456, R457, W464, M465, K468, E470, T474, P475, W476, F478, K484, S487, A488, T496, F498, L502, N504, K506, P509, F518, N522, E523, L540, S541, I548, D550, F553, V561, K562, E573, A589, L598, D605, L607, N609, N612, E617, D618, D628, R629, R635, K637, L651, K652, R654, T657, G658, L666, K673, S675, I679, L680, L683, N690, R691, N692, F693, S701, F704, Q712, G715, Q716, H723, I724, L727, I733, L738, Q739; где положение модифицированной аминокислотной последовательности идентифицируют посредством обращения к нумерации аминокислот в соответствующем положении немодифицированного зрелого Cas9 *Streptococcus pyogenes* (SpCas9), как идентифицируют с помощью SEQ ID NO: 1.

2. Модифицированный Cas9 по п. 1, который содержит по меньшей мере одну мутацию в положении K526.

3. Модифицированный Cas9 по п. 2, в котором мутацию в положении K526 выбирают в группе, состоящей из K526N и K526E.

4. Модифицированный Cas9 по любому одному из пп. 2-3, который содержит по меньшей мере одну дополнительную мутацию, расположенную в положении аминокислотного остатка, выбранном в группе, состоящей из:

K377, E387, D397, R400, Q402, R403, F405, D406, N407, A421, L423, R424, Q426, Y430, K442, P449, Y450, V452, A456, R457, W464, M465, K468, E470, T472, I473, T474, P475, W476, F478, K484, S487, A488, M495, T496, N497, F498, L502, N504, K506, P509, Y515, F518, N522, E523, L540, S541, I548, D550, F553, V561, K562, E573, A589, L598, D605, L607, N609, N612, E617, D618, D628, R629, R635, K637, L651, K652, R654, T657, G658, W659, R661, L666, K673, S675, I679, L680, L683, N690, R691, N692, F693, Q695, H698, S701, F704, Q712, G715, Q716, H721, H723, I724, L727, A728, I733, L738, Q739.

5. Модифицированный Cas9 по п. 4, в котором по меньшей мере одна дополнительная мутация находится в положении, выбранном в группе, состоящей из Y450, M495, Y515, R661, N690, R691, Q695, H698.

6. Модифицированный Cas9 по любому из п. 4-5, в котором по

меньшей мере одну дополнительную мутацию выбирают в группе, состоящей из:

K377E, E387V, D397E, R400H, Q402R, R403H, F405L, D406Y, D406V, N407P, N407H, A421V, L423P, R424G, Q426R, Y430C, K442N, P449S, Y450A, Y450S, Y450H, Y450N, V452I, A456T, R457P, R457Q, W464L, M465R, K468N, E470D, T472A, I473F, I473V, T474A, P475H, W476R, F478Y, F478V, K484M, S487Y, A488V, M495V, M495T, T496A, N497A, F498I, F498Y, L502P, N504S, K506N, P509L, Y515N, F518L, F518I, N522K, N522I, E523K, E523D, L540Q, S541P, I548V, D550N, F553L, V561M, V561A, K562E, E573D, A589T, L598P, D605V, L607P, N609D, N609S, N612Y, N612K, E617K, D618N, D628G, R629G, R635G, K637N, L651P, L651H, K652E, R654H, T657A, G658E, W659R, R661A, R661W, R661L, R661Q, R661S, L666P, K673M, S675C, I679V, L680P, L683P, N690I, R691Q, R691L, N692I, F693Y, Q695A, Q695H, Q695L, H698Q, H698P, S701F, F704S, Q712R, G715S, Q716H, H721R, H723L, I724V, L727H, A728G, A728T, I733V, L738P, Q739E, Q739P и Q739K.

7. Модифицированный Cas9 по любому из пп. 1–6, содержащий четверную мутацию, выбранную в группе, состоящей из M495V+Y515N+K526E+R661X и M495V+K526E+R661X+H698Q; где X представляет собой аминокислоту, выбранную в группе, состоящей из L, Q и S; предпочтительно X представляет собой Q или S.

8. Модифицированный Cas9 по п. 1, содержащий по меньшей мере одну мутацию, выбранную в группе, состоящей из K377E, E387V, D397E, R400H, Q402R, R403H, F405L, D406Y, D406V, N407P, N407H, A421V, L423P, R424G, Q426R, Y430C, K442N, P449S, Y450S, Y450H, Y450N, V452I, A456T, R457P, R457Q, W464L, M465R, K468N, E470D, T472A, I473F, I473V, T474A, P475H, W476R, F478Y, F478V, K484M, S487Y, A488V, M495V, M495T, T496A, F498I, F498Y, L502P, N504S, K506N, P509L, Y515N, F518L, F518I, N522K, N522I, E523K, E523D, K526E, K526N, L540Q, S541P, I548V, D550N, F553L, V561M, V561A, K562E, E573D, A589T, L598P, D605V, L607P, N609D, N609S, N612Y, N612K, E617K, D618N, D628G, R629G, R635G, K637N, L651P, L651H, K652E, R654H, T657A, G658E, W659R, R661W, R661L, R661Q, R661S, L666P, K673M, S675C, I679V, L680P, L683P, N690I, R691Q, R691L, N692I, F693Y, Q695H, Q695L, H698Q, H698P, S701F, F704S, Q712R, G715S, Q716H, H721R, H723L, I724V, L727H, A728G, A728T, I733V, L738P, Q739E, Q739P и Q739K.

9. Модифицированный Cas9 по любому из пп. 4–8, который содержит общее число указанных мутаций, составляющее от 1 до 9.

10. Модифицированный Cas9 по п. 8 или 9, который содержит:

одну мутацию, выбранную в группе, состоящей из D406Y, W464L, T474A, N612K, L683P; или

двойную мутацию, выбранную в группе, состоящей из R400H+Y450S, D406V+E523K, A421V+R661W, R424G+Q739P, W476R+L738P, P449S+F704S, N522K+G658E, E523D+E617K, L540Q+L607P, W659R+R661W, S675C+Q695L и I679V+H723L; или

тройную мутацию, выбранную в группе, состоящей из K377E+L598P+L651H, D397E+Y430C+L666P, Q402R+V561M+Q695L, N407P+F498I+P509L, N407H+K637N+N690I, Y450H+F553L+Q716H, Y450N+H698P+Q739K, T472A+P475H+A488V, I473F+D550N+Q739E, F478Y+N522I+L727H, K484M+Q695H+Q712R, S487Y+N504S+E573D, T496A+N609D+A728G, R654H+R691Q+H698Q и R691L+H721R+I733V;

четверную мутацию, выбранную в группе, состоящей из F405L+F518L+L651P+I724V, L423P+M465R+Y515N+K673M, R457P+K468N+R661W+G715S, E470D+I548V+A589T+Q695H, A488V+D605V+R629G+T657A и M495V+K526N+S541P+K562E; или

пять мутаций, выбранных в группе, состоящей из R403H+N612Y+L651P+K652E+G715S;

шесть мутаций, выбранных в группе, состоящей из E387V+V561A+D618N+D628G+L680P+S701F, R403H+K526E+N612Y+L651P+K652E+G715S, R403H+M495T+N612Y+L651P+K652E+G715S, R403H+L502P+N612Y+L651P+K652E+G715S, R403H+K506N+N612Y+L651P+K652E+G715S, R403H+N612Y+L651P+K652E+N692I+G715S; или

семь мутаций, выбранных в группе, состоящей из R403H+A456T+N612Y+L651P+K652E+G715S+G728T, R403H+F498Y+N612Y+L651P+K652E+R661L+G715S,

R403H+Q426R+F478V+N612Y+L651P+K652E+G715S; или

восемь мутаций, выбранных в группе, состоящей из R403H+R442N+V452I+N609S+N612Y+R635G+L651P+K652E+F693Y+G715S; или

девять мутаций, выбранных в группе, состоящей из R403H+R457Q+F518I+N612Y+R635G+L651P+K652E+F693Y+G715S.

11. Модифицированный Cas9 по любому из пп. 1–10, дополнительно содержащий по меньшей мере одну дополнительную мутацию, снижающую активность нуклеазы, в остатке, выбранном в группе, состоящей из D10, E762, D839, H840, N863, H983 и D986.

12. Модифицированный Cas9 по любому из пп. 1–10, дополнительно содержащий по меньшей мере одну дополнительную мутацию, меняющую специфичность распознавания PAM, в остатке,

выбранном в группе, состоящей из D1135, G1218, R1335, T1337.

13. Модифицированный Cas9 по любому из пп. 1-12, который представляет собой *S. pyogenes* или его ортологи.

14. Модифицированный Cas9 по любому из пп. 1-12, который сливают с одной или несколькими другими полипептидными последовательностями.

15. Способ получения модифицированного Cas9 по любому из пп. 1-14, указанный способ содержит восстановление модифицированного Cas9 из одного или нескольких его фрагментов.

16. Белок или рибонуклеопротеиновые комплексы или смешанные белковые рибонуклеопротеиновые или липидные комплексы, содержащие модифицированный Cas9 по любому из пп. 1-14.

17. Слитый белок, содержащий модифицированный Cas9 по любому из пп. 1-14.

18. Последовательность нуклеотидов, кодирующая модифицированный Cas9 по любому из пп. 1-14 и его фрагменты.

19. Нуклеиновая кислота, содержащая последовательность нуклеотидов по п. 18.

20. Вектор, содержащий нуклеиновую кислоту по п. 19.

21. Способ получения модифицированного Cas9 по любому из пп. 1-14, в соответствии с чем модифицированный Cas9 экспрессируют с помощью нуклеиновой кислоты по п. 19.

22. Клетка, животное или растение, сконструированные с использованием нуклеиновой кислоты по п. 19 или вектора по п. 20.

23. Последовательность нуклеотидов по п. 18 или нуклеиновая кислота по п. 19 или вектор по п. 20 для применения в качестве лекарственного средства в генной терапии.

24. Фармацевтическая композиция, содержащая последовательность нуклеотидов по п. 18 или нуклеиновую кислоту по п. 19 или вектор по п. 20 и по меньшей мере один фармацевтически приемлемый эксципиент.

25. Фармацевтическая композиция, содержащая рекомбинантный модифицированный Cas9 по любому из пп. 1-14 и по меньшей мере один фармацевтически приемлемый эксципиент.

26. Использование *in vitro* рекомбинантного модифицированного Cas9 по пп. 1-14 вместе с гРНК или последовательностью нуклеотидов по п. 18 или нуклеиновой кислотой по п. 19 или вектором по п. 20 для конструирования генома, клеточной инженерии, экспрессии белка или других



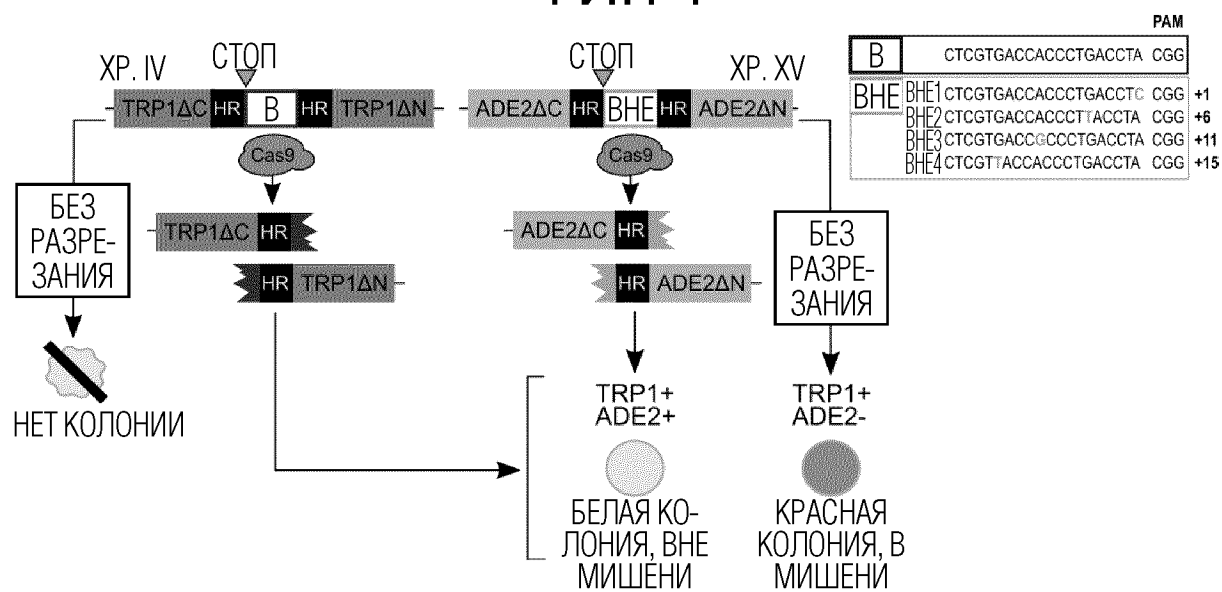
биотехнологических применений.

27. Набор частей для одновременного, отдельного или последовательного использования, который содержит рекомбинантный модифицированный Cas9 по пп. 1-14 вместе с гРНК или последовательность нуклеотидов по п. 18 или нуклеиновую кислоту по п. 19 или вектор по п. 20.

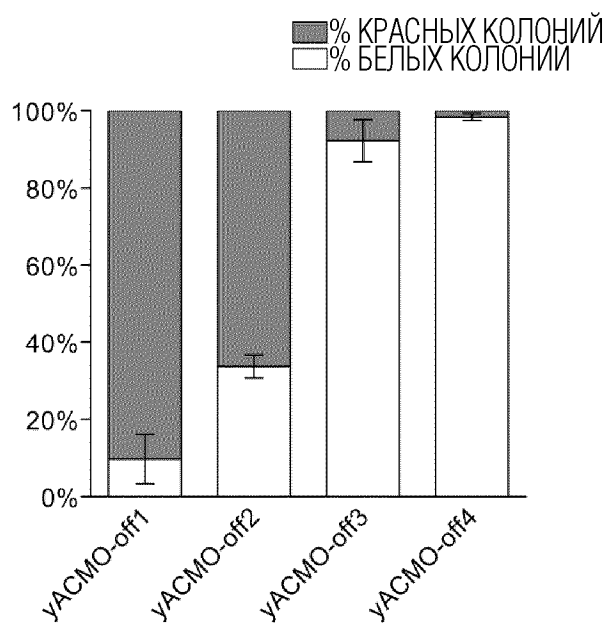
28. Способ изменения генома клетки *in vitro*, способ включает экспрессию в клетке модифицированного Cas9 по пп. 1-14 вместе с гидовой РНК, направленной на конкретную геномную последовательность.

1/9

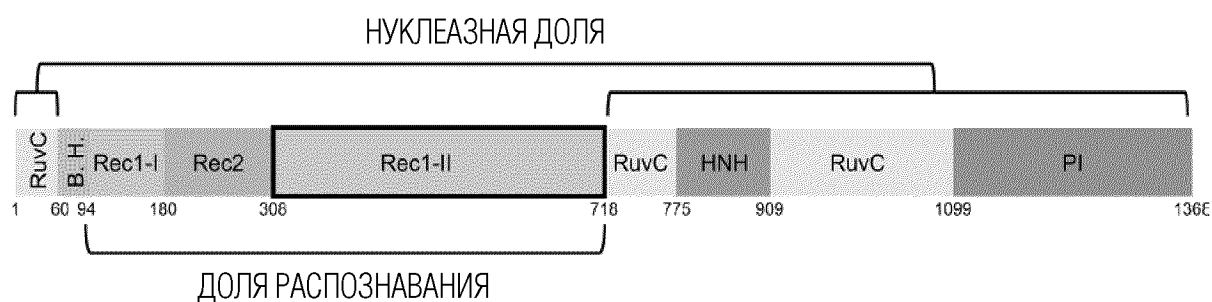
ФИГ. 1



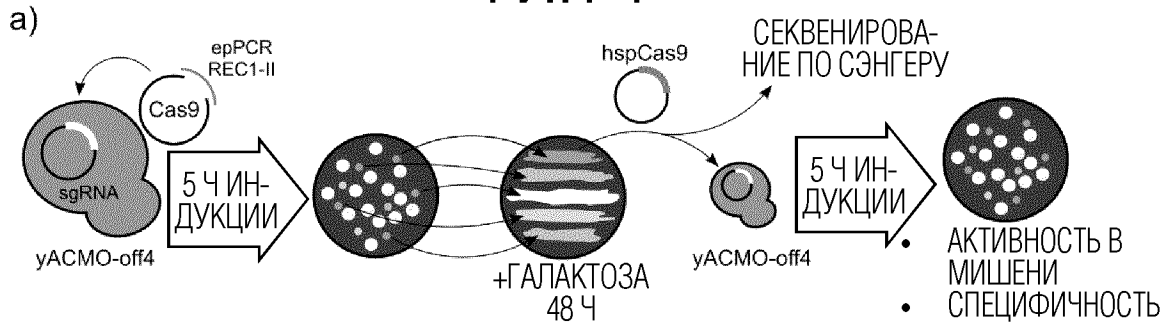
ФИГ. 2



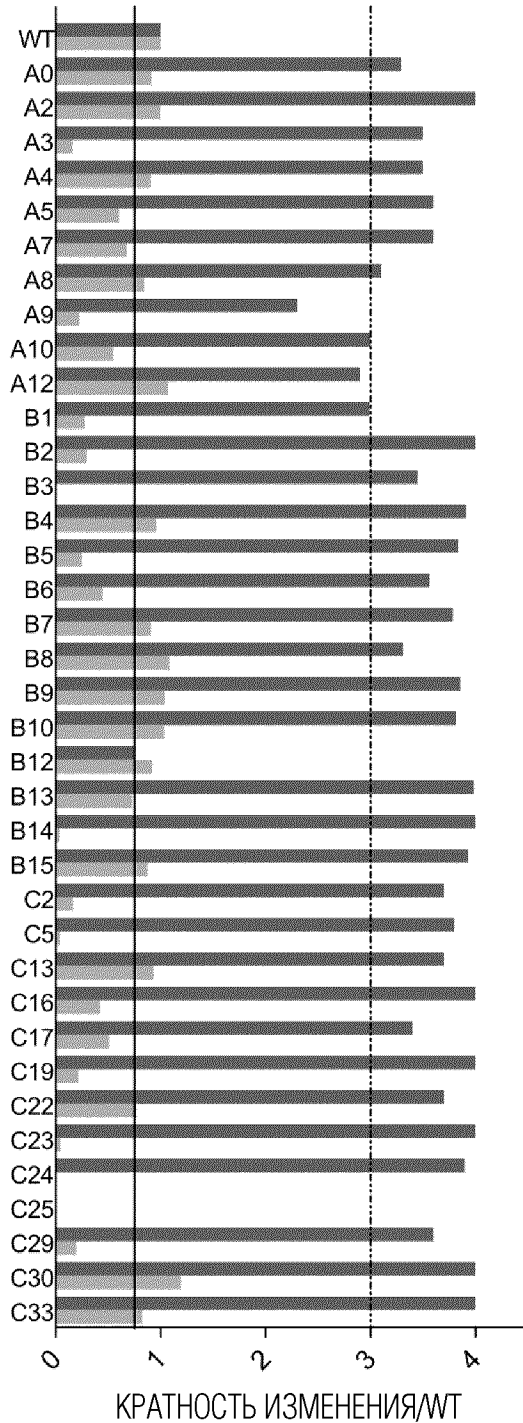
ФИГ. 3



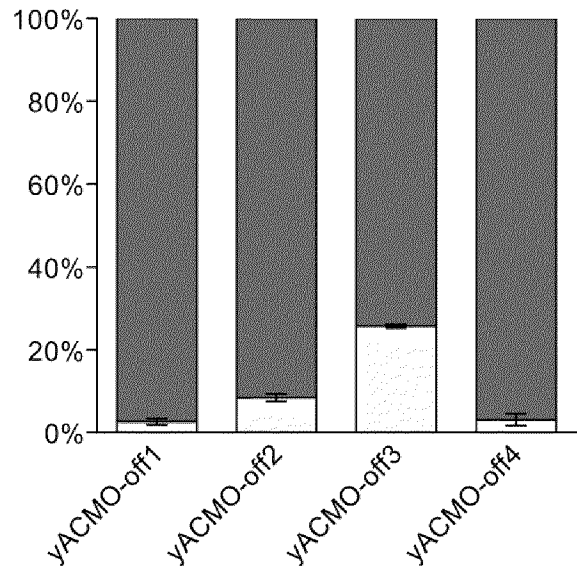
ФИГ. 4



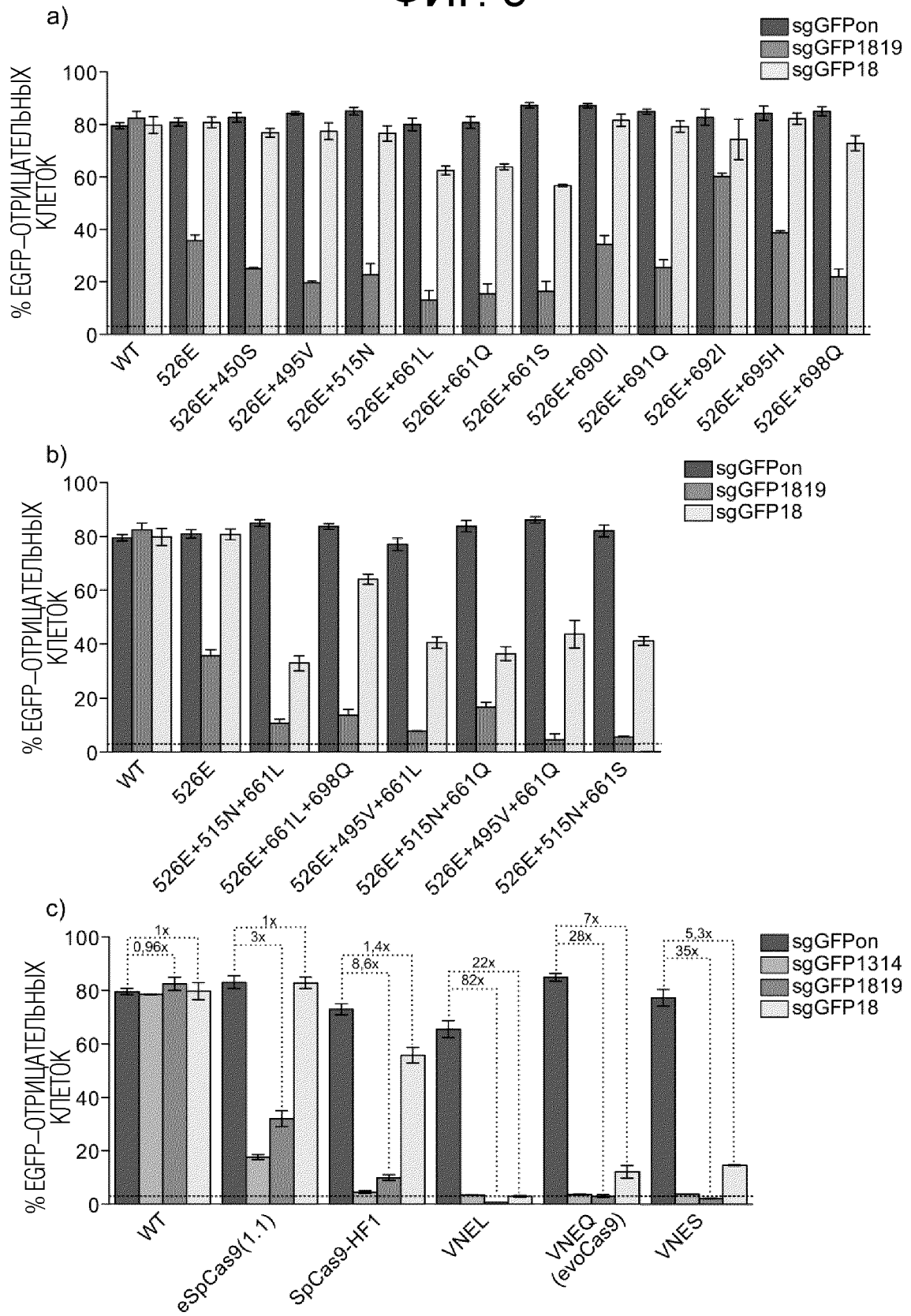
b) АКТИВНОСТЬ В МИШЕНИ  
СПЕЦИФИЧНОСТЬ



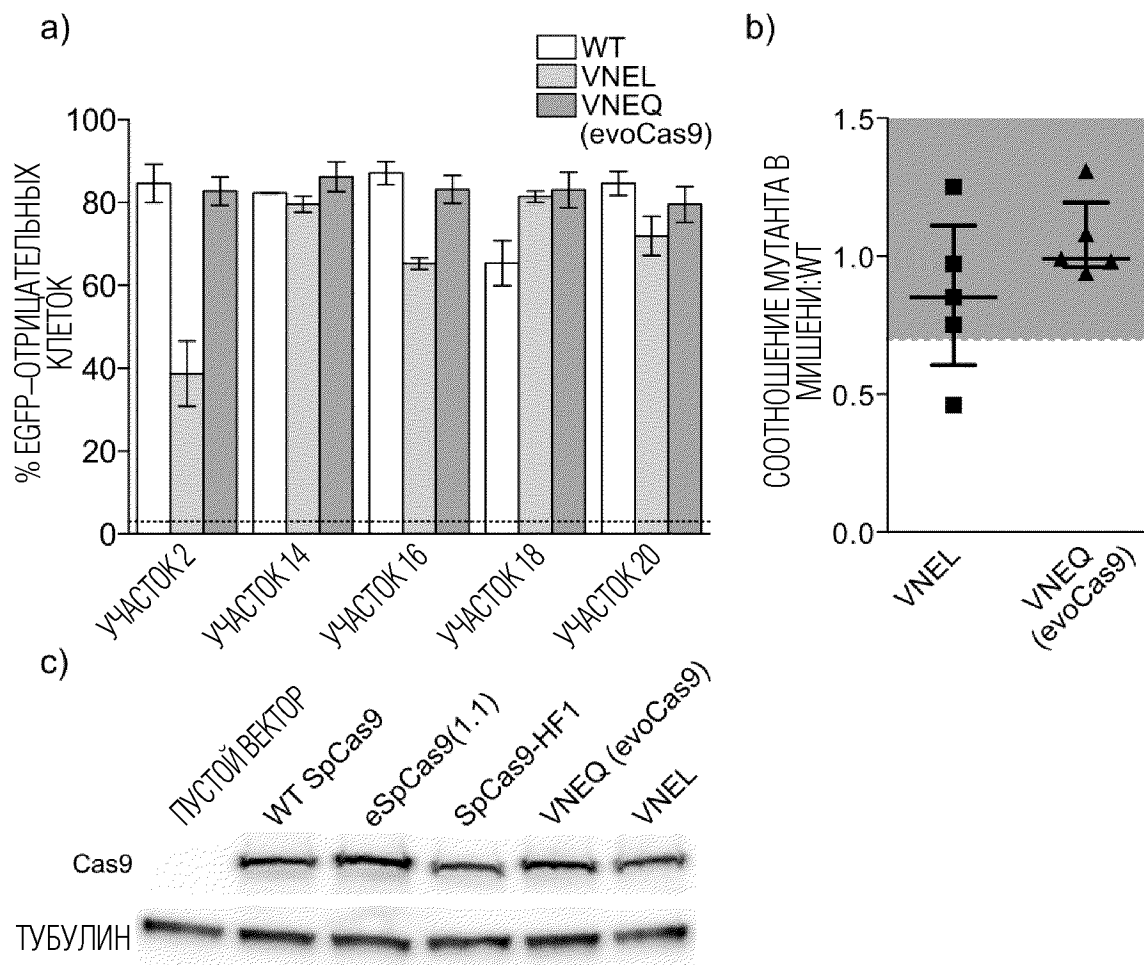
c) % КРАСНЫХ КОЛОНИЙ  
% БЕЛЫХ КОЛОНИЙ



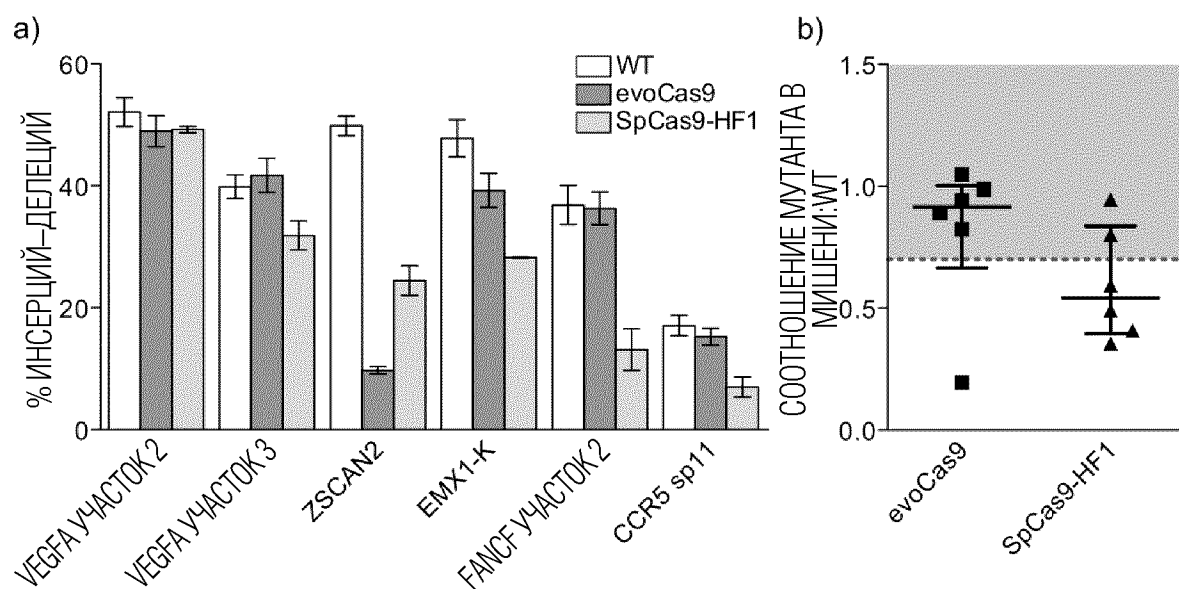
## ФИГ. 5



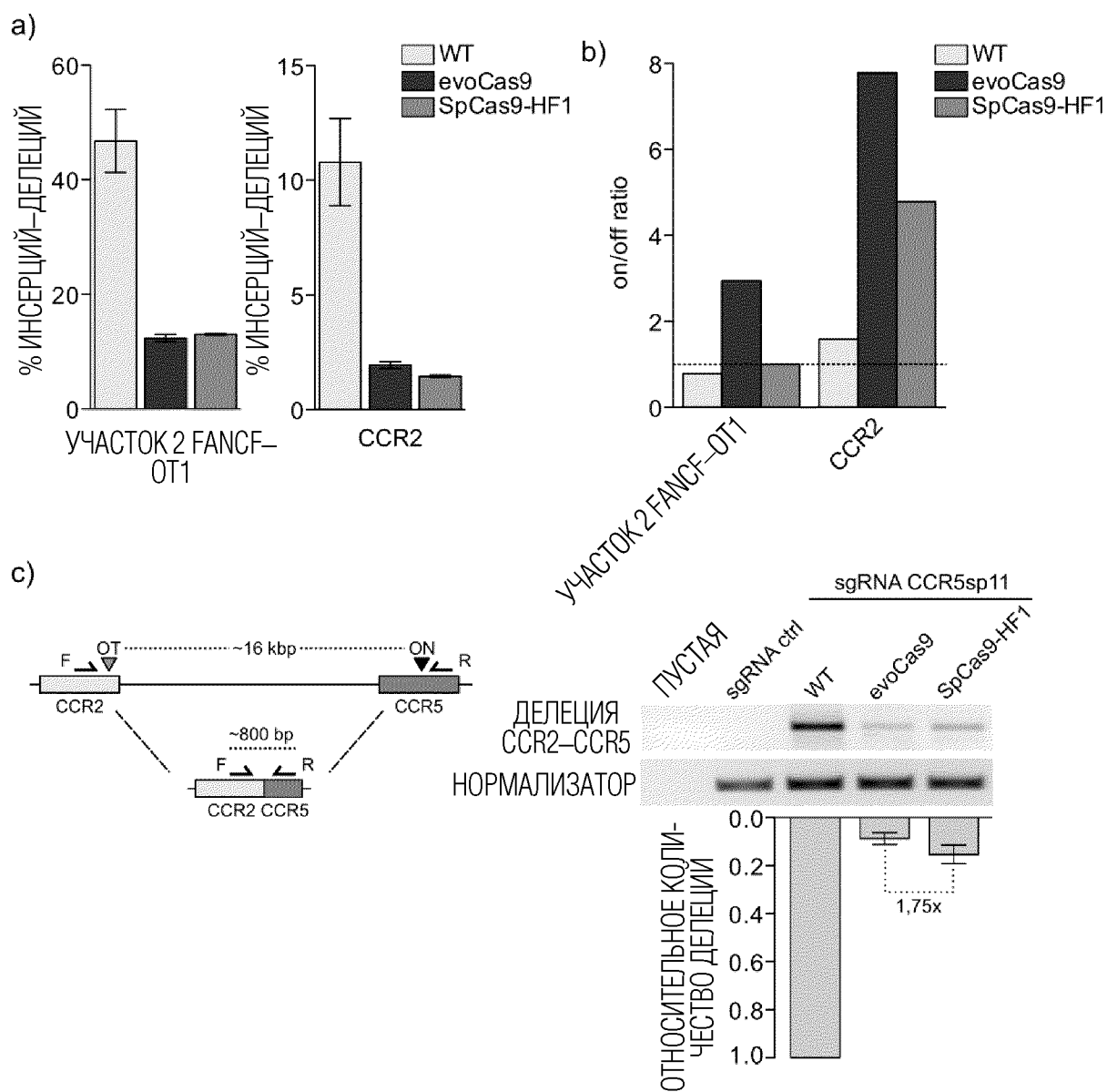
## ФИГ. 6



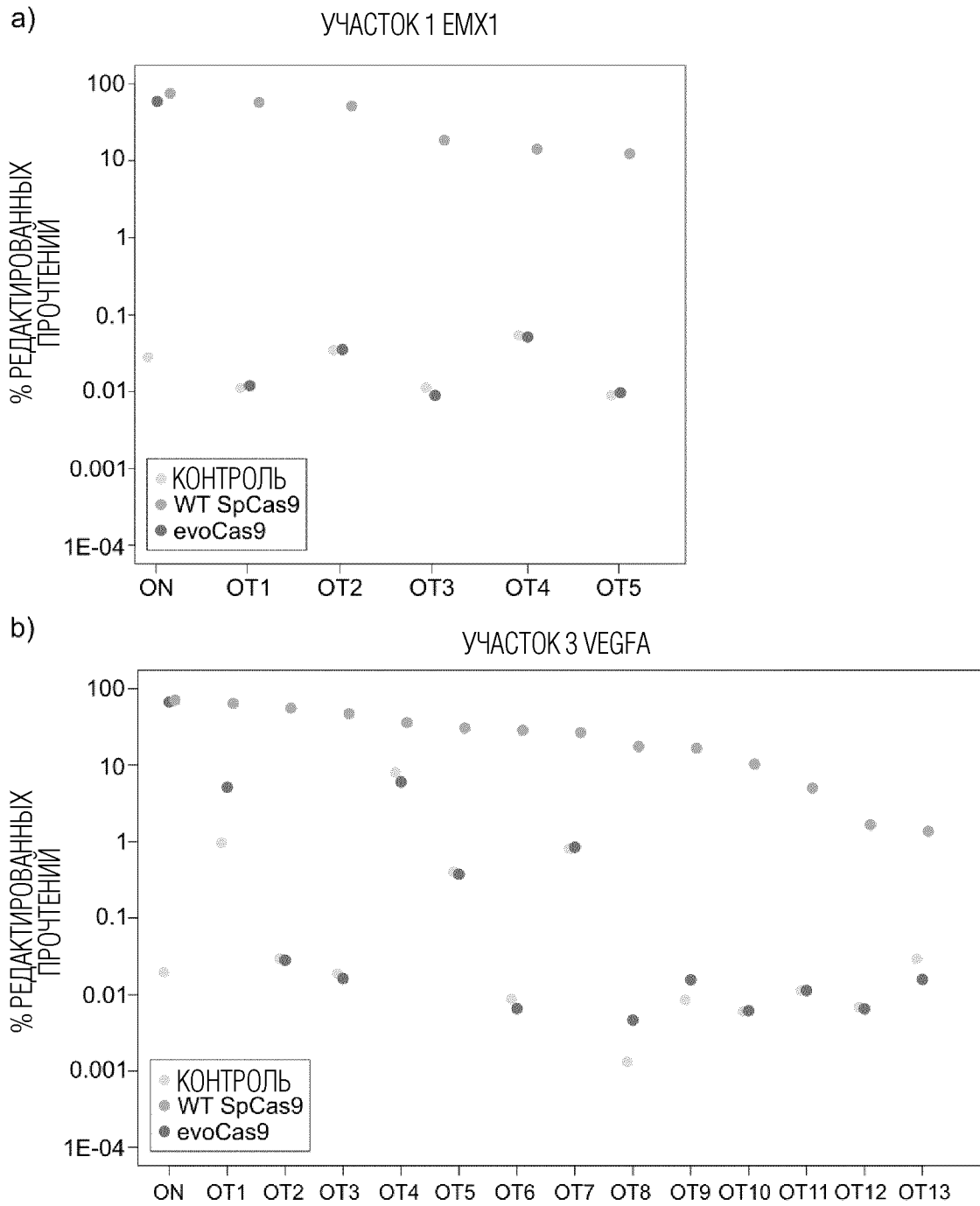
## ФИГ. 7



## ФИГ. 8



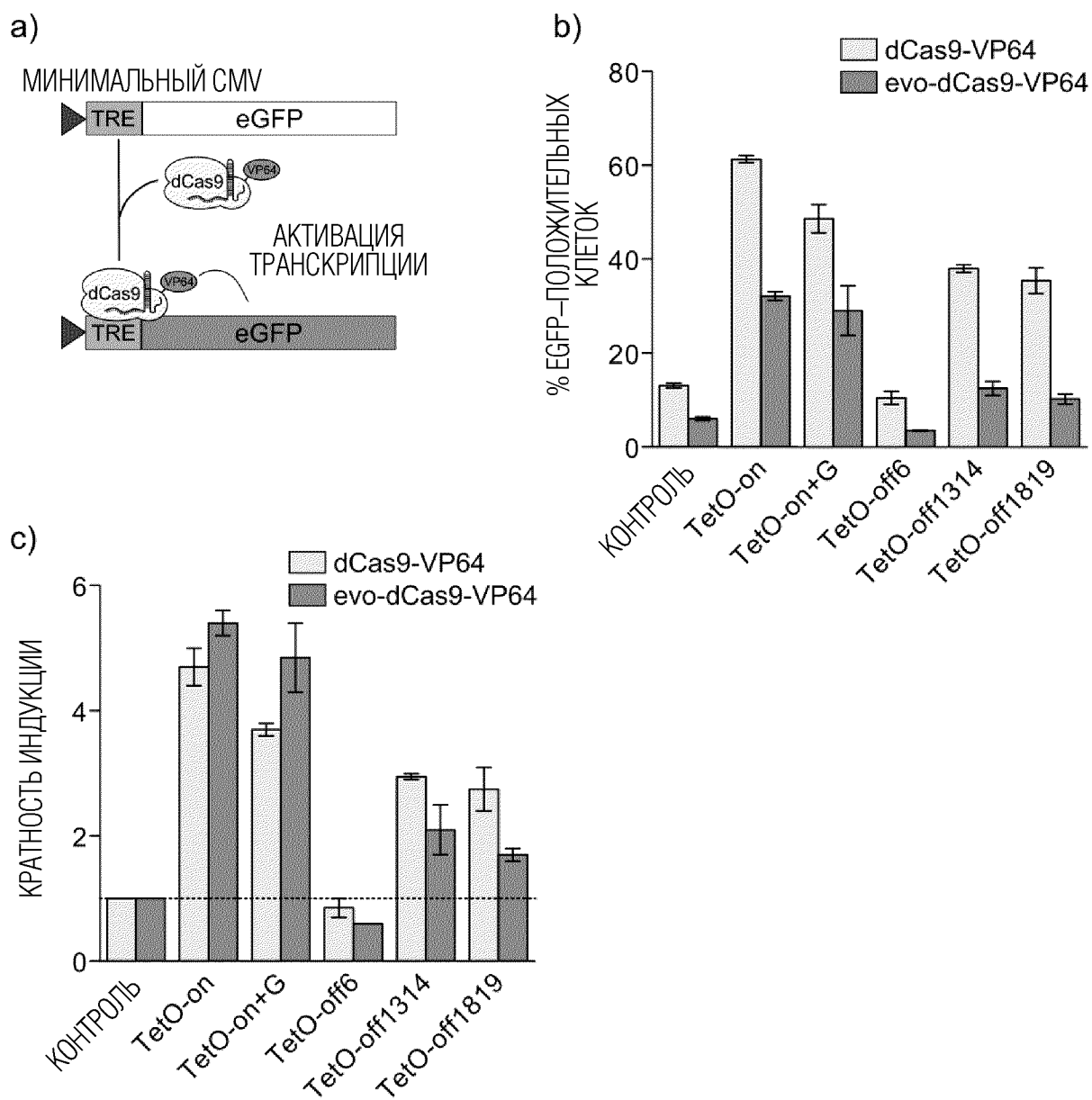
ФИГ. 9







## ФИГ. 11



## ФИГ. 12

