

БИОХИМИЯ, БИОФИЗИКА,  
МОЛЕКУЛЯРНАЯ БИОЛОГИЯ

УДК 577.29

КОНСТРУИРОВАНИЕ ГИДОВЫХ РНК ДЛЯ РЕДАКТИРОВАНИЯ  
ГЕНОМА КАРТОФЕЛЯ С ИСПОЛЬЗОВАНИЕМ СИСТЕМЫ CRISPR/Cas9

© 2018 г. А. В. Хромов<sup>1,2</sup>, В. А. Гуцин<sup>1,2</sup>, В. И. Тимербаев<sup>1,3</sup>, Н. О. Калинина<sup>1,2</sup>,  
М. Э. Тальянский<sup>1,4</sup>, В. В. Макаров<sup>1,2,\*</sup>

Представлено академиком РАН Л.П. Овчинниковым 11.07.2017 г.

Поступило 27.09.2017 г.

Провели сравнение *in vitro* активности пула молекул sgРНК, сконструированных к разным участкам генов коилина и фитоендесатуразы картофеля. Присутствие неспаренных с ДНК нуклеотидов в составе sgРНК может не только ингибировать, но и стимулировать активность комплекса Cas9–sgРНК *in vitro*. Первые 6 нуклеотидов, находящиеся в субстрате ДНК проксимально к сайту РАМ на 3'-конце, хотя и являются участками связывания Cas9, не оказывают значимого влияния на активность комплекса Cas9–sgРНК.

DOI: 10.7868/S0869565218090244

Картофель является одной из самых значимых сельскохозяйственных культур в мире. Однако глобальные изменения климата, появление новых патогенов и потребности в улучшении качества и лёжкости картофеля при хранении диктуют необходимость создания новых сортов картофеля с заданными свойствами. Методы геномного редактирования открывают принципиально новые возможности получения таких сортов в сжатые сроки и позволяют исследователям осуществлять манипуляции с практически любым геном самых разных клеток, тканей и организмов. Наиболее популярной является простая и удобная система редактирования генов, основанная на РНК-направляемых ДНК-нуклеазах и известная как CRISPR/Cas9 [1]. В природе CRISPR (clustered regularly interspaced short palindromic repeats) – короткие палиндромные повторы, регулярно расположенные группами, которые разделены уникальными последовательностями (спейсерами), были обнаружены в геноме прокариот и происходили от чужеродных генетических элементов (плазмид и бактериофагов). Молекулы РНК, транскрибирующиеся с локусов CRISPR, совместно с ассоциированной нуклеазой

Cas комплементарно связываются с ДНК чужеродных элементов с последующим их разрезанием белком Cas [2]. Упрощённая версия этой системы была разработана для геномного редактирования и включает два основных компонента: эндонуклеазу Cas9 и одиночную гидовую РНК (sgРНК), которая связывает комплементарный ей участок целевой ДНК длиной 20 нуклеотидов (так называемый протоспейсер), обеспечивая тем самым направленную доставку Cas9 к мотиву, прилегающему к протоспейсеру (сайт РАМ, для Cas9 – это обычно NGG), взаимодействие которого с эндонуклеазой приводит к эффективному разрезанию целевой последовательности ДНК.

Эта технология обеспечивает эффективную и точную модификацию генома, вызывая целевые разрывы ДНК, которые активируют клеточные механизмы репарации ДНК по механизмам нехомологичного (NHEJ) или гомологичного (HDR) соединения концов. Система CRISPR/Cas9 активно используется в качестве мощной платформы для редактирования генома, для влияния на транскриптом, для эпигенетической модуляции и визуализации генома [3–5]. Ключевую роль в успешном создании приложений на основе этой технологии играет рациональный дизайн sgРНК. Обычно дизайн потенциально возможных sgРНК осуществляется с помощью биоинформатических сервисов [9]. Однако на практике далеко не все из предсказанных sgРНК обладают высокой активностью. Таким образом, подбор sgРНК к уникальной последовательности в геноме имеет важное значение и зависит от не до конца изученных факторов. Кроме

<sup>1</sup>ООО “Дока Генные Технологии”, с. Рогачёво Московской обл.

<sup>2</sup>Московский государственный университет им. М.В. Ломоносова

<sup>3</sup>Институт биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова Российской Академии наук, Москва

<sup>4</sup>The James Hutton Institute, Dundee, UK

\*E-mail: makarovvalentine@gmail.com

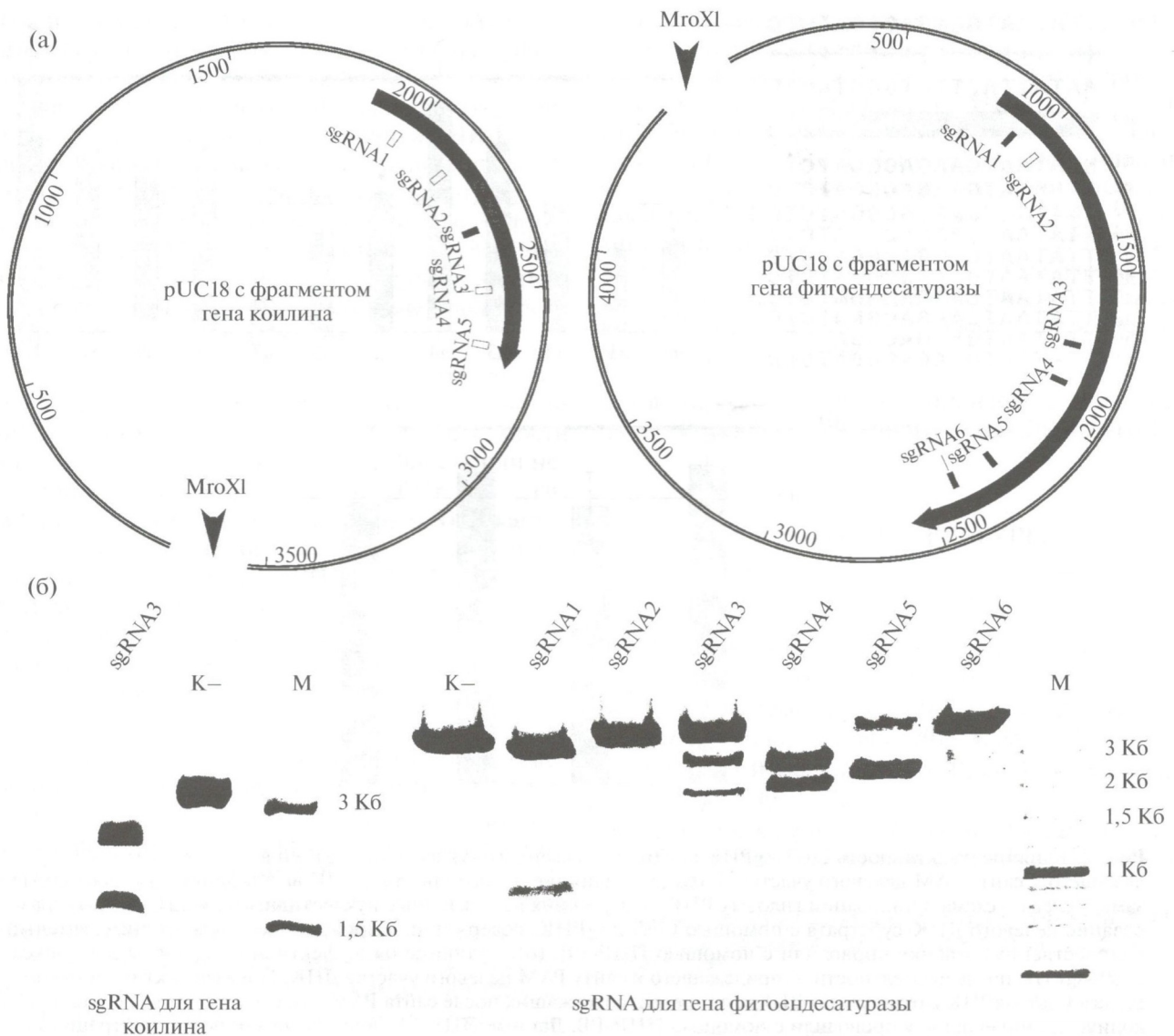
того, при подборе sgРНК следует исключить возможность нецелевого редактирования генома (так называемый “эффект off-target”), которая может являться следствием взаимодействия sgРНК с другими нуклеотидными последовательностями в геноме данного вида, имеющими некоторую гомологию с целевой последовательностью.

В настоящей работе мы проанализировали особенности функциональной активности sgРНК при разрывах нуклеазой Cas9 разных участков ДНК двух генов картофеля: гена фитоендесатуразы и гена основного белка субъядерных телец Кахалы – коилина. Фитоендесатураза – это фермент, вовлечённый в процессы биосинтеза каротиноидов, дефицит которого приводит к депигментации растений, и, таким образом, он может служить удобной модельной системой для визуализации и оценки эффективности методов геномного редактирования растений, в том числе и картофеля [8]. Дефицит коилина в растениях может повышать устойчивость растений к некоторым вирусным инфекциям и абиотическому (солевому) стрессу [6, 7], что делает ген коилина привлекательной мишенью для геномного редактирования.

Поскольку ключевыми параметрами для дизайна sgРНК являются положение сайта PAM, комплементарность спейсерной последовательности sgРНК протоспейсеру и возможность спаривания с нецелевыми участками ДНК в геноме, мы провели сравнительный анализ активности молекул sgРНК, сконструированных для получения разрывов в разных участках генов фитоендесатуразы и коилина картофеля. Кроме того, мы исследовали влияние неспаренных оснований в составе sgРНК и изучили роль 3'-прилежащих к сайту PAM нуклеотидных последовательностей, вовлечённых во взаимодействие между Cas9 и ДНК, на активность sgРНК *in vitro*.

Для сравнения эффективности индукции разрывов в разных участках ДНК генов коилина и фитоендесатуразы с помощью Cas9 к каждому из этих генов с помощью биоинформатического веб-сервиса (<http://crispr.mit.edu/>) мы отобрали последовательности (5 для коилина и 6 для фитоендесатуразы), комплементарные потенциальному сайту редактирования. Затем выбранные последовательности были синтезированы в виде олигонуклеотидов, клонированы в экспрессионные генно-инженерные конструкции, содержащие остальные необходимые элементы sgРНК, накоплены с помощью транскрипции *in vitro* и очищены с помощью Guide-it IVT RNA Clean-Up Kit (каталожный номер 632638, “New England BioLabs”, США). Контроль качества полученных sgРНК осуществляли путём анализа их подвижности и отсутствия фрагментации в агарозном геле. Активность синтезированных

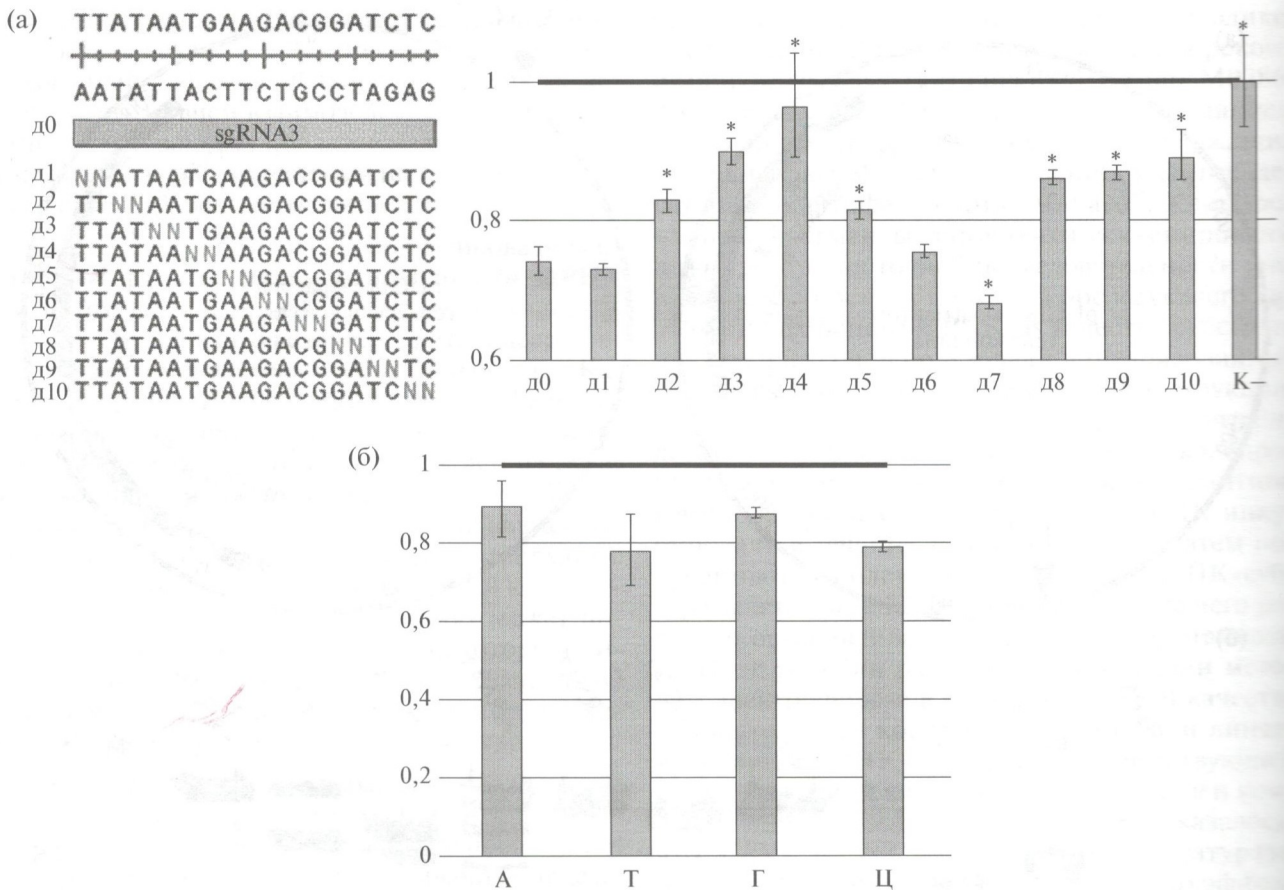
sgРНК определяли *in vitro* согласно методике, представленной на сайте производителя рекомбинантного Cas9 (каталожный номер M0386, “New England BioLabs”) [10] с использованием генно-инженерных конструкций, содержащих вставки фрагментов генов коилина и фитоендесатуразы картофеля сорта “Чикаго”. Клонированные фрагменты генов были секвенированы для определения точной последовательности гена и его целостности, а также для последующего дизайна на их основе sgРНК. В качестве субстрата для комплексов использовали линейаризованные по сайту MroXI генно-инженерные конструкции, содержащие фрагменты генов коилина и фитоендесатуразы. Для формирования рибонуклеопротеидного комплекса 30 нмоль рекомбинантной Cas9 и 30 нмоль синтезированной sgРНК инкубировали в течение 10 мин при 25°C. Затем полученным комплексом обрабатывали ДНК-субстраты в течение 15 мин при 37°C, после чего реакцию останавливали добавлением протеиназы К. Итог реакции разрезания определяли методом электрофореза в агарозном геле. В качестве отрицательного контроля использовали линейаризованный ДНК-субстрат в соответствующей концентрации, к которому не был добавлен комплекс эндонуклеазы и гидовой РНК. Оказалось, что 5 из 6 sgРНК для гена фитоендесатуразы были активны и могли в условиях *in vitro* эффективно направлять Cas9 для расщепления целевого участка ДНК, тогда как в случае коилина подобную активность мы зарегистрировали только у одной из 5 sgРНК. На рис. 16 представлен результат анализа фрагментов для 6 sgРНК фитоендесатуразы и единственной активной sgРНК для гена коилина. Размеры получившихся после разрезания фрагментов соответствовали ожидаемому, что указывало на специфичность разрезания целевой ДНК. Специфический характер разрезания также был подтверждён с помощью секвенирования образовавшихся фрагментов по Сэнгеру (данные не представлены). Следует отметить, что биоинформатический анализ не выявил каких-либо закономерностей, объясняющих присутствие или отсутствие активности sgРНК в зависимости от последовательности целевого участка генома картофеля, что указывает на возможность существования и других факторов, контролирующих активность sgРНК. В частности, мы не нашли корреляций между предсказанной с помощью веб-сервиса (<https://crispr.cdk.uni-heidelberg.de/>) эффективностью и реальной активностью sgРНК. Так, например, наивысшей предсказанной эффективностью должна была обладать sgРНК4 для гена коилина, однако в исследованиях *in vitro* она не проявила никакой активности.



**Рис. 1.** Активность *in vitro* комплекса Cas9–sgРНК при разрезании целевого ДНК-субстрата – генов коилина и фитоендесатуразы *S. tuberosum*. (а) – схема расположения sgРНК на целевых фрагментах генов коилина и фитоендесатуразы в составе генно-инженерных конструкций. Прямоугольниками чёрного цвета отмечены активные sgРНК. Стрелками – сайты линейаризации. (б) – визуализация реакции разрезания с помощью фрагментного анализа в агарозном геле. В качестве отрицательного контроля использовали линейаризованный ДНК-субстрат в соответствующей концентрации, к которому не был добавлен комплекс эндонуклеазы и гидовой РНК. Сопоставление (а) и (б) показывает, что размеры получившихся после разрезания фрагментов соответствуют ожидаемым. В случае фрагмента гена коилина приведены данные только для активной sgРНК.

Для дальнейшей работы мы использовали активную sgРНК гена коилина, поскольку она не обеспечивала полного разрезания матрицы, и поэтому имелась возможность повышения эффективности разрезания ДНК посредством оптимизации последовательности sgРНК. В последовательность взаимодействующего с протоспейсером участка sgРНК вставляли вырожденные нуклеотидные замены. При синтезе олигонуклеотида нуклеотиды попарно в последовательном порядке заменяли на случайный. Синтез sgРНК для оценки влияния

вырожденных нуклеотидных замен на её активность осуществляли так, как было описано выше. Активность синтезированных sgРНК при разрезании *in vitro* фрагмента гена коилина оценивали количественно методом ПЦР в реальном времени (ПЦР-РВ) с использованием пары праймеров, прилежащих к участку разрезания. В случае, когда матрица была разрезана с помощью комплекса Cas9–sgРНК, эффективность амплификации соответствующего фрагмента снижалась и ПЦР блокировалась. Мы проанализировали активность



**Рис. 2.** Влияние на активность Cas9–sgРНК *in vitro* вырожденных нуклеотидных замен в составе sgРНК и 3'-прилежащего к сайту РАМ целевого участка ДНК. (а) – влияние на активность sgРНК вырожденных нуклеотидных замен; слева – схема организации гидовых РНК, содержащих вырожденные нуклеотидные замены, справа – разрезание целевого ДНК-субстрата с помощью Cas9 и sgРНК, содержащих вырожденные нуклеотидные замены; количественную оценку проводили с помощью ПЦР-РВ. (б) – влияние на эффективность работы комплекса Cas9–sgРНК последовательности 3'-прилежащего к сайту РАМ целевого участка ДНК. Показана активность комплекса Cas9–sgРНК в отношении ДНК-субстратов, содержащих после сайта РАМ нуклеотиды 6А, 6Т, 6Г или 6С; количественную оценку проводили с помощью ПЦР-РВ. Данные ПЦР-РВ нормализовали по концентрации исходной матрицы (исходного ДНК-субстрата, горизонтальная линия) и представляли в виде отношения количества “нерасщеплённой” после реакции ДНК к общему количеству ДНК, внесённой в реакционную смесь.  $M \pm m$ ,  $n = 5$ ,  $*p < 0,05$  при сравнении с положительным контролем (исходная гидовая РНК).

10 разных sgРНК, содержащих вырожденные нуклеотидные замены (рис. 2а). Ряд вырожденных замен – нуклеотиды 3–4 (д2), 5–6 (д3), 9–10 (д5), 15–16 (д8), 17–18 (д9) и 19–20 (д10) – привели к снижению активности, а замена нуклеотидов 7–8 (д4) – к полному подавлению активности соответствующей sgРНК. Замены 1–2 (д1) и 11–12 (д6) не влияли на активность sgРНК, а замена 13–14 (д7) достоверно повысила эффективность разрезания.

Таким образом, полная комплементарность между sgРНК и целевым участком ДНК не является фактором, определяющим эффективность индукции разрывов ДНК нуклеазой. Более того, отсутствие полного спаривания между sgРНК и ДНК

может приводить к повышению эффективности разрезания.

Известно [11], что помимо взаимодействия комплекса Cas9–sgРНК с целевым участком ДНК имеются также прямые взаимодействия между белком Cas9 и 6 нуклеотидами ДНК-субстрата, прилегающими к сайту РАМ и не вовлечёнными во взаимодействие с sgРНК. Поэтому мы предположили, что эта последовательность ДНК-субстрата может быть одним из факторов, влияющих на активность комплекса Cas9–sgРНК. Для экспериментальной проверки этого предположения были получены 4 мутантных варианта потенциального участка редактирования гена коидина, в которых шесть 3'-прилежащих к сайту РАМ нуклеотидов были заменены на 6 цитозинов (мутант 6С).

6 гуанинов (мутант 6Г), 6 аденинов (мутант 6А) и 6 тиминов (мутант 6Т) соответственно. Анализ эффективности разрезания целевых ДНК, проведённый с помощью ПЦР-РВ, не выявил достоверных различий между эффективностью разрезания ДНК-мишени как дикого типа, так и всех остальных мутантных вариантов (рис. 2б). Эти данные позволили нам сделать предположение, что хотя 6 нуклеотидов, 3'-проксимальных к РАМ, и участвуют в непосредственном взаимодействии с Cas9, их состав и первичная структура заметной роли в определении специфичности этого взаимодействия не играют.

В заключение следует подчеркнуть, что данные, полученные нами при тестировании активности sgРНК *in vitro*, не позволили выявить особенности их структуры, определяющие способность или неспособность ассоциированной с sgРНК нуклеазы Cas9 индуцировать разрывы геномной ДНК картофеля. С помощью тестов *in vitro* мы отобрали и оптимизировали активные sgРНК для последующего редактирования генов коилина и фитоендесатуразы картофеля. Следующим этапом нашей работы будет оценка эффективности геномного редактирования с применением этих sgРНК в условиях *in vivo*.

Используя вырожденные нуклеотидные замены, при которых в определенной позиции равновероятно может находиться любой из четырех нуклеотидов, нами было показано, что неспаренные нуклеотиды в спейсерном участке sgРНК могут не только снижать, но и повышать эффективность действия комплекса Cas9–sgРНК. Данное обстоятельство расширяет возможности неспецифической доставки Cas9 к частично комплементарным участкам генома, повышая риск его нецелевой модификации, и, следовательно, должно учитываться при предварительном биоинформатическом поиске целевых сайтов для геномного редактирования.

Рациональный подбор sgРНК к уникальной последовательности в геноме обеспечивает высокую специфичность редактирования. В настоящий момент существует ряд принципов рационального дизайна sgРНК [12]. Известно, что наиболее значимым для проявления специфической активности sgРНК является так называемый регион “seed” гидовой РНК (10–12 нуклеотидов, проксимальных к сайту РАМ) [12], причём sgРНК, имеющие низкое содержание G/C обычно существенно менее активны [12, 13]. Однако данные наблюдения справедливы только для больших выборок и не могут предсказать, будет ли активной конкретная sgРНК. Кроме того, большинство литературных данных свидетельствует о роли отдельных нуклеотидов, а не их сочетаний, что и обусловило попарную замену нуклеотидов в составе sgРНК в нашей работе.

Таким образом, мы показали, что присутствие неспаренных с ДНК нуклеотидов в составе sgРНК может не только ингибировать, но и стимулировать активность комплекса Cas9–sgРНК *in vitro*. Кроме того, было показано, что первые 6 нуклеотидов, находящиеся в ДНК-субстрате 3'-области проксимально к сайту РАМ, хотя и являются участками связывания Cas9, не оказывают значимого влияния на активность комплекса Cas9–sgРНК. Полученные результаты имеют важное значение для прогнозирования специфичных сайтов целевого редактирования генома и предварительного тестирования sgРНК *in vitro*.

Работа выполнена при финансовой поддержке гранта Российского научного фонда 16–16–04019.

## СПИСОК ЛИТЕРАТУРЫ

1. Mali P., Yang L., Esvelt A., Aach J., Guell M., DiCarlo J.E., et al. // *Science* 2013. V. 339. P. 823–826.
2. Al-Attar S., Westra E.R., van der Oost J., Brouns S.J. // *Biol. Chem.* 2011. V. 392. 277–289.
3. Cong L., Ran F.A., Cox D., Lin S., Barretto R., Habib N., et al. // *Science* 2013. V. 339. P. 819–823.
4. Taylor J., Woodcock S. // *J. Biomol. Screen* 2015. V. 20. P. 1040–1051.
5. Nakamura K., Fujii M., Tsuboi M., Tanihata J., Teramoto N., Takeuchi S., et al. // *Sci. Rept.* 2014. V. 4. P. 5635.
6. Love A.J., Yu C., Petukhova N.V., Kalinina N.O., Chen J., Taliany M.E. // *RNA Biol.* 2016. V. 11. P. 1–12.
7. Shaw J., Love A.J., Makarova S.S., Kalinina N.O., Harrison B.D., Taliany M.E. // *Nucleus*. 2014. V. 5. P. 85–94.
8. Tian S., Jiang L., Gao Q., Zhang J., Zong M., Zhang H., Ren Y., Guo S., Gong G., Liu F., Xu Y. // *Plant Cell Rept.* 2017. V. 36. P. 399–406.
9. Электронный ресурс <http://crispr.mit.edu/>
10. Larson M.N., Gilbert L.A., Wang X., Lim W.A., Weissman J.S., Qi L.S. // *Nat. Protoc.* 2013. V. 8. P. 2180–2196.
11. Hirano H., Gootenberg J.S., Horii T., Abudayyeh O.O., Kimura M., Hsu P.D., Nakane T., Ishitani R., Hata-da I., Zhang F., Nishimasu H., Nureki O. // *Cell*. 2016. V. 164. P. 950–961.
12. Doench J.G., Hartenian E., Graham D.B., Tothova Z., Hegde M., Smith I., Sullender M., Ebert B.L., Xavier R.J., Root D.E. // *Nat. Biotechnol.* 2014. V. 32. P. 1262–1267.
13. Doench J.G., Fusi N., Sullender M., Hegde M., Vaimberg E.W., Donovan K.F., Smith I., Tothova Z., Wilen C., Orchard R., Virgin H.W., Listgarten J., Root D.E. // *Nat. Biotechnol.* 2016. V. 34. P. 184–191.