



БИОМИКА/BIOMICS

<http://biomics.ru>



БИОИНФОРМАТИЧЕСКИЕ РЕСУРСЫ ДЛЯ CRISPR/Cas РЕДАКТИРОВАНИЯ ГЕНОМОВ

Чемерис Д.А.¹, Кирьянова О.Ю.^{2,3}, Геращенко Г.А.¹, Кулуев Б.Р.¹, Рожнова Н.А.¹,
Матниязов Р.Т.¹, Баймиев Ан.Х.¹, Баймиев Ал.Х.¹, Губайдуллин И.М.^{2,3}, Чемерис А.В.¹

¹Институт биохимии и генетики Уфимского научного центра Российской академии наук, Уфа, chemeris@anrb.ru

²Институт нефтехимии и катализа Российской академии наук, Уфа

³Уфимский государственный нефтяной технический университет, Уфа

Резюме

Кратко рассмотрены особенности многочисленных программ дизайна гидРНК с указанием их актуальных web-страниц и описаны этапы работы с ними. Для ряда программ приведена более подробная характеристика, отраженная в специальном Приложении. Перечислены программы поиска произведенных в результате CRISPR/Cas редактирования мутаций в геномах различных организмов. Отмечены базы данных по предварительно подобранным гидРНК для ряда геномов растительных и животных организмов, включая человека. Проведен анализ практически всей имеющейся литературы по компьютерному дизайну гидРНК и соответствующих интернет-ресурсов.

Ключевые слова: CRISPR, Cas, PAM, гидРНК, спейсер, протоспейсер, компьютерная программа, web-ресурс, база данных

Содержание

	Стр.
Введение	203
Основные термины и определения в CRISPR/Cas редактировании	205
Особенности РНК-компонентов некоторых Cas нуклеазных комплексов в CRISPR/Cas редактировании	206
Основные требования, предъявляемые к дизайну гидРНК	207
Компьютерные программы и web-ресурсы для дизайна гидРНК	208
Основные принципы работы программ дизайна гидРНК	219
Базы данных по подобранным гидРНК	220
Программы для анализа отредактированных геномов	220
Заключение	221
Благодарности	221
Литература	221
Приложение	226

Введение

Успех конкретного эксперимента по геномному CRISPR/Cas редактированию во многом зависит от дизайна единой гидовой (направляющей) РНК (направляющего домена егидРНК или иначе варибельной части егидРНК), который в первую очередь должен учитывать максимально возможное недопущение возникновения нежелательных мутаций при уверенном производстве целевых. Так как поиск целевых и нецелевых сайтов геномного

редактирования ведется путем анализа совпадения в геномах разных организмов последовательностей лишь варибельной части егидРНК, а константная ее часть, зависящая лишь от типа используемой Cas нуклеазы, остается, что называется, «за кадром» ввиду того, что в подавляющем большинстве программ для дизайна егидРНК экспериментатор ее не видит, то для краткости в данной статье будем использовать словосочетание «направляющая или гидовая РНК» и аббревиатуру «гидРНК»,

подразумевая под ними вариабельный домен егидРНК или иначе спейсер.

После того, как в CRISPR/Cas9 геномном редактировании, благодаря объединению тракрРНК и крРНК в единой молекуле направляющей РНК [Jinek et al., 2012], произошел серьезнейший прорыв, стали массово появляться программы для дизайна гидРНК, в том числе в виде web-ресурсов и облачных сервисов. К настоящему времени программ для дизайна молекул гидРНК написано несколько десятков и уже имеется немало обзорных статей, посвященных таким программным продуктам для проведения CRISPR/Cas редактирования [Zhu, 2015; Brazelton, Jr. et al., 2015; Graham, Root, 2015; Kanchiswamy, 2016; Peng et al., 2016; Chuai et al., 2017; Periwal, 2017; Yan et al., 2017]. В двух последних обзорах одного коллектива авторов рассмотренные ими программы дизайна гидРНК (более трех десятков) разделены на три группы – написанные на основе выравнивания данных, управляемые гипотезой и самообучающиеся, представленные в их подборке в соотношении приблизительно 2:1:1. Исходя из ряда критериев, лучшими программами для поиска гидРНК по версии этих авторов [Yan et al., 2017] были признаны sgRNA-designer и E-CRISPR. Забегая вперед, скажем, что первая из них позволяет работать лишь с двумя геномами – мыши и человека. В еще одном недавнем кратком обзоре индийских авторов [Yennmalli et al., 2017] приведен перечень компьютерных программ, имеющих отношение к CRISPR/Cas системам, где они были сгруппированы, исходя из неких вариантов их использования, и пронумерованы. Так, в табличной форме с указанием их URL-адресов перечислены 65 компьютерных программ, позволяющих (как указано в заголовке таблицы) вести дизайн гидРНК. Однако при детальном рассмотрении становится очевидным, что в этот список попало семь программ, нацеленных на анализ последствий геномного редактирования, а также пять программ, позволяющих вести поиск CRISPR локусов в секвенированных геномах микроорганизмов. Таким образом, непосредственно программ, осуществляющих дизайн гидРНК, в их подборке содержится 53.

Насколько можем судить из известной нам литературы и интернет-ресурсов (<https://omictools.com>, <https://github.com>, <https://sourceforge.net>, <https://www.hsls.pitt.edu/obrc/>) в данной обзорной статье приведен наиболее полный перечень таких программ дизайна гидРНК, количество которых составило около семи десятков, не считая программ анализа результатов произведенного редактирования геномов, которых упомянуто в данной статье более 10. К тому же здесь нами даны их как краткие, так и

довольно подробные характеристики, а не только URL-адреса как в статье Yennmalli и соавт. [2017].

Несмотря на то, что акцент здесь будет делаться на программы, которые позволяют анализировать геномы растений на предмет выявления в них возможных нецелевых сайтов редактирования, прочие существующие программы дизайна гидРНК также будут упомянуты. В том числе потому, что сам по себе дизайн гидРНК может проводиться с любыми нуклеотидными последовательностями, не взирая на их принадлежность к тем или иным организмам. Некоторым недостатком этих прочих программ данной группы можно считать невозможность с их помощью оценивать наличие в геномах растений нецелевых мест редактирования. Однако у экспериментатора всегда имеется возможность проверить как полное, так и неполное совпадение предлагаемых спейсерных последовательностей гидРНК по всему GenBank с помощью сервиса Nucleotide BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). При этом можно заметить, что полных геномов растений, представленных в этой и других подобных базах данных до сих пор не так много. И это в основном модельные или широко возделываемые сельскохозяйственные растения, тогда как перед исследователем может стоять задача редактирования генома какого-либо не очень популярного или даже редкого вида растения, полный геном которого или не секвенирован совсем или находится только лишь на стадии завершения секвенирования/сборки контигов и потому весь не доступен. Для таких объектов вполне пригодны и программы дизайна гидРНК, не предназначенные для выявления off-target мишеней в полных геномах растений.

Учитывая растущий объем подобных исследований по редактированию геномов, стали создаваться, поддерживаться и пополняться специализированные базы данных, содержащие заранее найденные последовательности протоспейсеров, способные служить основой для составления гидРНК. Отдельную группу компьютерных программ формируют позволяющие вести поиск мест произошедшего CRISPR/Cas редактирования путем анализа результатов секвенирования, включая полногеномное, отредактированных организмов.

Здесь за пределами нашего рассмотрения окажутся компьютерные программы, нацеленные на поиск в геномах микроорганизмов CRISPR-локусов, а также имеющиеся базы данных по таковым, поскольку этим вопросам посвящена самостоятельная статья в этом же номере журнала [Баймиев и др., 2017].

Основные термины и определения в CRISPR/Cas редактировании

В другой нашей статье в данном номере журнала [Кулуев и др., 2017] приведено большое

число терминов и пояснений к ним и поэтому здесь ограничимся лишь теми, которые преимущественно используются в биоинформационном обеспечении CRISPR/Cas редактировании геномов.

Таблица

Основные термины и определения в CRISPR/Cas редактировании геномов

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)	Короткие палиндромные повторы, регулярно расположенные группами
Cas (CRISPR associated)	Ассоциированные с CRISPR кассетами нуклеазы
Cas9	Наиболее широко используемая CRISPR нуклеаза, генерирующая образование тупых концов
<i>Sp</i> Cas9	Cas9 нуклеаза, где <i>Sp</i> означает, что данный фермент обнаружен в микроорганизме <i>Streptococcus pyogenes</i> . Аналогичные обозначения (обычно двухбуквенные) делаются и для других подобных Cas нуклеаз, когда важно подчеркнуть принадлежность фермента
Cpf1	CRISPR нуклеаза, относящаяся ко II классу и V-A типу CRISPR/Cas систем (Cas12), генерирующая 5'-выступающие концы
C2c2	CRISPR нуклеаза, относящаяся ко II классу и VI типу CRISPR/Cas систем (Cas13a), разрушающая молекулы РНК
DSB (Double-Stranded Breaks)	Разрывы обеих цепей молекул ДНК, образующиеся под действием различных ферментов или физическими воздействиями напротив друг друга с образованием тупых концов или на некотором (небольшом) расстоянии друг от друга
nCas9 (nickase) (никаза)	Мутантная форма Cas9 нуклеазы, ставшая никазой, поскольку способна разрезать только одну из цепей ДНК
PAM (Protospacer Adjacent Motif)	Мотив, смежный с протоспейсером (для нуклеаз разного типа может располагаться как с его 5'-, так и с 3'-конца), служащий неким указателем мишени в составе редактируемой геномной ДНК
Protospacer	Специфичная последовательность, служащая мишенью в составе редактируемой геномной ДНК, связывающаяся со спейсером или вариательной частью гидРНК
crRNA (крРНК)	РНК, включающая спейсер из CRISPR-кассеты
tracrRNA (тракрРНК)	Транс-активирующая РНК, участвующая в образовании комплекса Cas9 с молекулами РНК
sgRNA (гидРНК)	Единая гидовая (направляющая) РНК, состоящая из вариательной (спейсер) и константной (крРНК и тракрРНК) частей
Spacer	Специфичная вариательная последовательность в составе единой направляющей РНК, связывающаяся с протоспейсером, а также участок CRISPR-кассеты, перемежающий палиндромные квазитандемные повторы
«Seed sequence»	Якорная или затравочная последовательность обычно из 8-12 нуклеотидов на 3'-конце вариательной последовательности в составе направляющей РНК, связывающаяся с протоспейсером и определяющая специфичность узнавания
On-target site	Специфичная последовательность в геноме, с которой должна связываться направляющая РНК
Off-target site	Неспецифичная последовательность в геноме, с которой может, но не должна связываться направляющая РНК
Knock-out (нокаут)	Нарушение работы какого-либо гена, вызываемого обычно репарацией негомлогичных концов (NHEJ)
Knock-in (нокин)	Внедрение в геномную ДНК какого-либо гена, вызываемого гомологичной репарацией (HDR)

**Особенности РНК-компонентов
некоторых Cas нуклеазных комплексов
в CRISPR/Cas редактировании**

Прежде чем перейти к описанию программ дизайна гидРНК следует уделить определенное внимание организации РНК-компонентов CRISPR/Cas систем. Так, объединение в единой молекуле направляющей РНК двух молекул – тракрРНК и крРНК [Jinek et al., 2012] произвело настоящую революцию в геномном редактировании, заметно упростив подготовку к нему и повысив его эффективность. Поскольку протяженность комплекса крРНК:тракрРНК довольно велика и для природных молекул составляет для Cas9 нуклеазы из *Streptococcus pyogenes* 125 нуклеотидов (42 и 83 нуклеотида для крРНК и тракрРНК соответственно), то вполне оправданным выглядело намерение установить минимальные последовательности этих РНК, достаточные для формирования активного комплекса с Cas9 нуклеазой. Так, был создан ряд укороченных вариантов крРНК и тракрРНК,

сохраняющих функциональность, и показано, что можно обойтись без последних 10 нуклеотидов на 3'-конце крРНК, а тракрРНК может быть укорочена с обоих концов весьма существенно. Для большего удобства было произведено слияние этих молекул РНК в единую химерную последовательность с помощью введения дополнительных четырех нуклеотидов, сформировавших так называемую тетра-GAAA-петлю, между 3'-концом крРНК и 5'-концом тракрРНК. В результате были созданы максимально укороченные константные части гидРНК длиной 34 и 42 нуклеотида без учета варибельной последовательности / спейсера гидРНК – длиной 20 нуклеотидов [Jinek et al., 2012].

Однако, такие излишне короткие молекулы гидРНК проявляли активность, уступающую природным формам, и поэтому в настоящее время наиболее часто используемые гидРНК для нуклеазы Cas9 имеют протяженность около 90 нуклеотидов [Haussler, Concordet, 2016]. На рис. 1 приведена вторичная структура гидРНК для Cas9 нуклеазы.

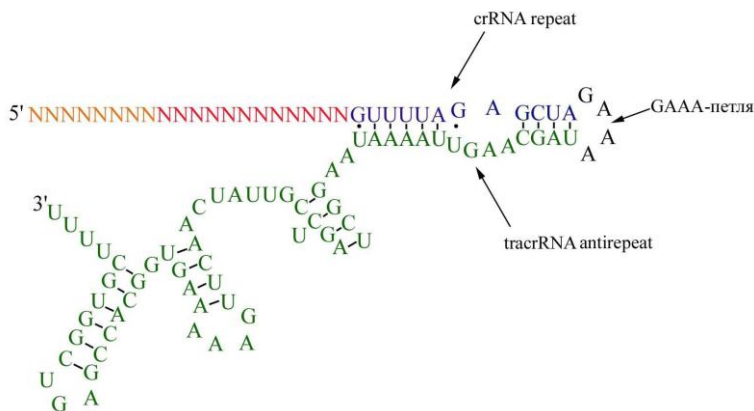


Рис. 1. Вторичная структура константной части гидРНК для нуклеазы Cas9.

Варибельная часть гидРНК (спейсер) показана оранжевым и красным цветами. Красным цветом выделен наиболее важный якорный или затравочный участок, прилегающий к РАМ. Синим цветом показана константная часть крРНК, зеленым – константный фрагмент тракрРНК, а черным – искусственная тетра-нуклеотидная GAAA петля, соединяющая в единую гидРНК две природные (но слегка укороченные) молекулы крРНК и тракрРНК.



Рис. 2. Вторичная структура константной части гидРНК для нуклеазы AsCpf1 («минус»-область) с «линейной» варибельной частью гидРНК (спейсером) и якорным или затравочным участком, прилегающим к РАМ последовательности.

Как оказалось, для проявления ферментативной активности нуклеазы Cas12 (Cpf1) не требуется вторичная структура РНК [Zetsche et al., 2015]. На рис. 2 приведена вторичная структура одной из таких Cas12 нуклеаз из *Acidaminococcus sp.*

Длина такой молекулы гидРНК для Cpf1 ферментов из разных микроорганизмов обычно составляет около 45 нуклеотидов, что еще легче синтезировать химическим способом, хотя и синтез 100 звеньев рибоолигонуклеотидов не представляет сейчас особой проблемы и для CRISPR/Cas редактирования довольно широко используется [Kelley et al., 2016]. В настоящее время стоимость химического синтеза молекул РНК обходится приблизительно раз в пять дороже синтеза ДНК, но можно предположить, что с ростом популярности таких заказов цена за звено будет быстро снижаться.

Основные требования, предъявляемые к дизайну гидРНК

При дизайне гидРНК для проведения геномного редактирования необходимо обеспечить исключение возникновения нецелевых мутаций при эффективном произведении целевых. Для достижения первой цели требуется анализ всего генома (в случае, если вся его последовательность известна) на предмет наличия в нем совпадающих с редактируемым участком (протоспейсером) нуклеотидных последовательностей (полностью или частично). Так, некая обычная (рандомная) 20-звенная последовательность (наиболее часто используемая длина протоспейсера) в случайной ДНК теоретически может встречаться только через каждый триллион нуклеотидов (4^{20}). Тем самым практически исключается ее нахождение в каком-либо другом месте редактируемого генома (если она только не является частью каких-либо повторяющихся элементов), поскольку самые крупные из известных геномов составляют около сотни миллиардов оснований и таких очень немного. Однако с учетом того, что при редактировании возможно неполное спаривание, крайне желательно исследовать весь геном конкретного организма, допуская при этом не 100%-ное совпадение нуклеотидных последовательностей спейсерного участка направляющей РНК с потенциальными протоспейсерами. Причем следует уделять особое внимание якорной части протоспейсера, где замены нуклеотидов наиболее критичны. Без использования компьютера и соответствующего программного обеспечения подобный анализ провести невозможно. Причем многие программы ранжируют подобранные протоспейсеры по их пригодности для редактирования и исключения нецелевых мутаций, давая им количественную оценку.

Другая решаемая задача при дизайне гидРНК заключается в обеспечении наиболее эффективного процесса редактирования и также зависит от нуклеотидной последовательности выбранного участка, который должен исключать возникновение нежелательной (неправильной) вторичной структуры молекулы РНК, а также отсутствия подобных мест в донорной ДНК в случаях, когда предполагается провести нокин CRISPR-Cas редактирование.

Существует уже довольно большое число написанных программ дизайна гидРНК, в основе которых лежат отличающиеся алгоритмы, разработанные несколькими авторами [Hsu et al., 2013; Doench et al., 2014; 2016] и нашедшие применение в ряде последующих программ дизайна гидРНК. Однако прежде чем перейти к их рассмотрению следует заметить, что в связи с бурным развитием технологий геномного редактирования на основе CRISPR/Cas систем довольно быстро меняются воззрения экспериментаторов на этот процесс, и получается, что ранее написанные программы уже не в полной мере соответствуют реалиям дня сегодняшнего. Так, подавляющее большинство программ рассчитано на то, что гидРНК будет синтезироваться энзиматически *in vitro* или даже *in vivo* в живых организмах, будь то бактерия или эукариотический организм. При этом некоторые программы даже предлагают дизайн олигонуклеотидов для составления правильной конструкции при ее клонировании в подходящем векторе для наработки соответствующей гидРНК. Однако в этом случае из состава гидРНК исключаются последовательности, которые могут служить сигналом для преждевременного завершения транскрипции. Например, разработчики программы WU-CRISPR [Wong et al., 2015] отмечают, что специально заложили в программу исключение в гидРНК подряд четырех урацилов (UUUU), поскольку, встретив их, РНК полимеразы III, управляемая с промотора U6, с высокой вероятностью завершит транскрипцию. Тем самым сужается выбор протоспейсеров, которые по остальным параметрам могли бы быть вполне удачными при том, что химический синтез гидРНК и доставка тем или иным путем готового (образованного *in vitro*) комплекса Cas нуклеаза/гидРНК в клетку снимает эти проблемы ферментативной наработки нужной гидРНК. При этом ряд недавно созданных программ дизайна гидРНК (например, Breaking Cas) позволяют использовать подобные участки, но обращают на них внимание экспериментаторов, например, восклицательными знаками в соответствующем месте. Также при ферментативном синтезе гидРНК для его эффективности часто рекомендуется делать

так, что первым нуклеотидом (или даже двумя нуклеотидами) должны быть остатки гуанина. При химическом синтезе эта проблема также отпадает сама собой. Однако остается главная проблема исключения нежелательной вторичной структуры всей гидРНК, которая может возникнуть вследствие формирования вторичной структуры спейсерной (вариабельной) области гидРНК с остальной константной частью, искажив вторичную структуру последней и не дав образоваться полноценному комплексу с используемой нуклеазой. При этом в некоторых статьях отмечается, что успешная редакция была ими осуществлена с помощью гидРНК, содержащей от 30 до 80% GC-оснований (имеется в виду в спейсере) [Liang et al., 2016]. Здесь можно напомнить, что гуанины и цитозины образуют по три водородных связи и формируют более прочные вторичные структуры, которые с большей вероятностью могут нарушить ту правильную структуру всей гидРНК, которая требуется для взаимодействия с Cas нуклеазой. И высокое GC-содержание в спейсерной области гидРНК чревато почти непредсказуемым спариванием. Так, показано, что из многих гидРНК, содержащих в своей последовательности участки GGGG, лишь около 5% оказались функционально активными [Wong et al., 2015]. Также более высокое содержание GC-оснований будет способствовать возникновению большего числа нецелевых сайтов редактирования из-за того, что в условиях неполного спаривания такие гидРНК смогут связаться с нецелевым участком ДНК и тем самым дадут возможность Cas нуклеазе произвести двуцепочный разрыв ДНК в ненужном месте. Помимо GC-состава гидРНК, был проанализирован вклад типов азотистых оснований при их нахождении на всех позициях гидРНК в эффективность процесса редактирования геномов [Liu et al., 2015; Xu et al., 2015].

Компьютерные программы и web-ресурсы для дизайна гидРНК

Большинство описанных ниже программ представлено web-ресурсами, включая облачные сервисы, часть может быть также установлена на компьютеры пользователей и совсем немного программ не рассчитаны на работу в сети Интернет и требуют установки. Ниже с соблюдением насколько это возможно хронологического порядка приведены сведения о таких программах, их возможностях и порядке работы с ними, что представляет интерес в первую очередь для конечных пользователей данных программ для дизайна соответствующих гидРНК при выполнении ими конкретных экспериментов по редактированию геномов тех или иных организмов.

В своей статье [Ma et al., 2013], опубликованной epub в октябре 2013 г., китайские авторы отметили, что на тот момент не имелось доступного программного обеспечения для дизайна гидРНК, в связи с чем они разработали специальную on-line платформу **Guide RNA Sequence Design Platform**, указав ее адрес в сети Интернет - <http://cas9.cbi.pku.edu.cn>. Эта программа учитывала наличие однонуклеотидных замен, нуклеотидный состав вариабельной области гидРНК, вторичную структуру гидРНК и ранжировала найденные последовательности по этим характеристикам. Однако надо заметить, что еще в июле 2013 появилась epub публикация [Hsu et al., 2013] (вышедшая в журнальном варианте в сентябре того же года), в которой упоминалось о разработке соответствующего алгоритма и написании работающей on-line программы дизайна гидРНК. В конце октября 2013 г. вышла статья той же группы авторов [Ran et al., 2013], где был приведен подробный протокол для проведения геномного редактирования, включая дизайн гидРНК, который как было сообщено производился с помощью on-line ресурса **CRISPR Design Tool**, адрес которого был указан как <http://tools.genome-engineering.org>. В настоящее время программа **Optimized CRISPR Design** (<http://crispr.mit.edu:8079>) предлагает два варианта подбора гидРНК – для единичной или для множественных последовательностей ДНК. Далее предстоит выбрать один из 16 геномов различных организмов, среди которых только один принадлежит растению – арабидопсису. После этого необходимо подтвердить согласие о том, что экспериментатор не будет использовать предоставленную ему этой программой информацию для клинических целей, и запустить процесс дизайна. Также требуется указать свой email адрес. Размещенная в другом месте (<https://www.atum.bio/eCommerce/cas9/input>) аналогичная программа **crispr grna design tool** имеет некоторые отличия в организации дизайна гидРНК. Так, на этом сайте исследователю предлагается меньший выбор геномов (всего пять, в том числе арабидопсис), два участка PAM (NGG и NAG), при этом предоставляется выбор ферментов из обычной Cas9 нуклеазы или никаз на ее основе. Далее на усмотрение пользователя необходимо ввести или наименование конкретного гена, или участок генома или ввести через буфер обмена интересующую его последовательность ДНК (не более 10 тысяч нуклеотидов) и запустить поиск. Также имеется вариант анализа уже подобранной ранее гидРНК на предмет обнаружения с ней потенциальных мест редактирования (как целевых, так и нецелевых) только в двух геномах – человека и мыши.

При использовании программы **sgRNA Designer (CRISPRko)** (<http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>) [Doench et al., 2014; 2016] предлагается сначала выбрать тип нуклеазы, например Cas9 из *S.pyogenes* (PAM - NGG) или *S.aureus* (PAM - NNGRR), затем необходимо перейти к выбору генома всего из двух предлагаемых вариантов - мышь или человек. Нуклеотидные последовательности (до десяти), в которых необходимо найти участки, пригодные быть протоспейсерами, могут быть загружены через буфер обмена или в виде файлов формата FASTA (размер файла не должен превышать 10 Кбайт) или с помощью наименований генов/транскриптов, после чего начинается поиск гидПНК, заканчивающийся их ранжированием, лучшие располагая выше. Данная программа может быть установлена на компьютер пользователя путем загрузки дистрибутива с адреса <http://portals.broadinstitute.org/gpp/public/software/sgrna-scoring>.

CasOT – CRISPR/Cas system (Cas9/gRNA) Off-Targeter (<http://casot.cbi.pku.edu.cn>) [Xiao et al., 2014] также оказалась среди первых программ, направленных на выбор участков для CRISPR/Cas9 редактирования геномов, также учитывающей нецелевое связывание спейсерных последовательностей гидПНК, разграничивая при этом затравочную часть этой молекулы от менее значимой 5'-области. Программа рассчитана на автономное использование и доступна для скачивания. Программный код реализован на языке Perl. В программе предусмотрены три режима поиска: single-gRNA, paired-gRNA и target-and-off-target. Пользователи могут указывать несколько параметров поиска, в зависимости от поставленных целей.

Усилиями двух университетов и одного исследовательского института из Южной Кореи создан специализированный портал **CRISPR RGEN Tools** (<http://www.rgenome.net>), где представлены несколько компьютерных программ, используемых в экспериментах по CRISPR/Cas технологии - **Cas-OFFfinder** [Bae et al., 2014], **Microhomology-Predictor** [Bae et al., 2014a], **Cas-Designer** [Park et al., 2014], **Cas-Database** [Park et al., 2016], **Cas-Analyzer** [Park et al., 2017], **Digenome-Seq** [Kim et al., 2015]. Кроме того, данный web-ресурс содержит в себе две базы данных **Cas-Database** и **Cpf1-Database** [Park et al., 2016]. В базах данных пользователи могут выбирать оптимальные целевые последовательности из тысяч генов сразу, просто изменяя условия фильтрации. Программа **Cas-Designer** (<http://www.rgenome.net/cas-designer/>) [Park et al., 2014] в анализируемой последовательности находит

все возможные протоспейсеры, служащие местами связывания гидПНК и при этом принимает во внимание микрогомологию, которая может оказывать влияние на объединение концов при репарации по типу NHEJ. При этом предлагается широкий выбор нуклеаз в том числе их мутантных форм с отличающимися PAM последовательностями (*SpCas9*, *StCas9*, *NmCas*, *SaCas9*, *CjCas9*, *AsCpf1*, *FnCpf1*). Затем вводится анализируемая последовательность (не более 1000 нуклеотидов) и задается размер спейсера гидПНК (от 15 до 25), после чего необходимо выбрать тот или иной геном из предлагаемого списка, разделенного на группы. Состав группы «Plant» довольно внушителен и содержит 59 видов: *Arabidopsis thaliana*, *Oryza sativa*, *Solanum lycopersicum*, *Zea mays*, *Chlamydomonas reinhardtii*, *Solanum tuberosum*, *Glycine max*, *Vitis vinifera*, *Manihot esculenta*, *Malus domestica*, *Hordeum vulgare*, *Nicotiana benthamiana*, *Fragaria vesca*, *Citrus sinensis*, *Theobroma cacao*, *Musa acuminata*, *Arachis ipaensis*, *A.duranensis*, *Actinidia chinensis*, *Brassica napus*, *Glycine max*, *Sorghum bicolor*, *Panicum virgatum*, *Gracilaria lemaneiformis*, *Cannabis sativa*, *Medicago truncatula*, *Setosphaeria turcica*, *Capsicum annum*, *Daucus carota*, *Petunia axillaris*, *P.inflata*, *Phaseolus vulgaris*, *Eragrostis tef*, *Eucalyptus grandis*, *Cajanus cajan*, *Fragilariopsis cylindrus*, *Brassica rapa*, *Gossypium hirsutum*, *Brassica oleracea*, *Nicotiana tabacum*, *Coffea canephora*, *Citrus clementina*, *Beta vulgaris*, *Kalanchoe fedtschenkoi*, *K.laxiflora*, *Sesamum indicum*, *Cucumis sativus*, *Citrullus lanatus*, *Cucumis melo*, *Populus trichocarpa*, *Eucalyptus grandis*, *Solanum lycopersicum*, *Populus tremula*, *Populus tremula x tremuloides*, *Populus tremuloides*, *Physcomitrella patens*, *Brachypodium distachyon*, *Nicotiana obtusifolia* и *Triticum aestivum*. Допускается обратиться к разработчикам с просьбой добавления еще какого-нибудь генома. Программа также может быть скачана с сайта и установлена на компьютер пользователя и тогда появляется возможность менять дополнительные параметры самостоятельно, расширяя возможности Cas-Designer, поскольку в on-line версии ряд установок выполняются по умолчанию. Для обнаружения возможных нецелевых мест в исследуемых геномах к работе присоединяется программа **Cas-OFFfinder** (<http://www.rgenome.net/cas-offfinder/>) [Bae et al., 2014], имеющая конкретное предназначение. При этом данная программа может работать самостоятельно, ведя поиск нецелевых мест в геномах разных организмов для уже подобранных гидПНК. В этом случае через буфер обмена в соответствующее окошко вводится последовательность гидПНК (от 15 до 25 нуклеотидов) и задается допустимое количество

неспариваний нуклеотидов и выбирается анализируемый геном из предлагаемого списка, разделенного на те же группы, что и при использовании Cas-Designer.

Программа дизайна гидПНК **CRISPRseek** рассчитана на ее установку на компьютер пользователя [Zhu et al., 2014] и может быть загружена с сайта <http://bioconductor.org/packages/release/bioc/html/CRISPRseek.html>, где имеются ее версии для операционных систем Windows и Mac. Возможности программы CRISPRseek довольно широки. Она позволяет искать парные места редактирования для Cas9 нуклеаз, сравнивать несколько последовательностей, предсказывать вторичную структуру молекул гидПНК, отмечать наличие сайтов рестрикции в редактируемых областях, а также искать нецелевые места редактирования во многих геномах, при этом позволяя менять РАМ последовательности, например вместо канонической NGG искать также off-target участки с NAG.

Написанная в 2014 г. программа **E-CRISP** [Heigwer et al., 2014] сейчас существует уже в виде Version 5.3 (<http://www.e-crisp.org/E-CRISP/>). В первоначальной версии для анализа предлагалось 12 геномов, сейчас их количество выросло до 55, из них более десятка растительных – арабидопсис, мягкая пшеница, ячмень, кукуруза, виноград, тополь, три вида риса и др. После выбора генома требуется ввести анализируемую последовательность, указав ее идентификационный номер или воспользовавшись буфером обмена компьютера. Допустимый уровень неспариваний задается исходя из трех режимов – смягченный, средний или строгий. Также необходимо указать требуется ли единичная гидПНК или ее парные варианты для использования нуклеазы или для нокин-экспериментов. Прежде чем начать поиск рекомендуется воспользоваться дополнительными опциями. В варианте **MultiCRISP** можно провести выравнивание нескольких родственных последовательностей.

Программа **CRISPR Optimal Target Finder** (<http://tools.flycrispr.molbio.wisc.edu/targetFinder/>) ориентирована на поиск и валидацию целевых и нецелевых сайтов для гидПНК в геномах преимущественно насекомых, среди которых много видов дрозофил, имеются геномы комара, пчелы и некоторых других видов [Gratz et al., 2014].

Программа **GT-Scan** (<http://gt-scan.csiro.au>) [O'Brien, Bailey, 2014] позволяет искать гидПНК как в off-line режиме после ее установки на компьютер пользователя, так и в on-line версии при работе в сети интернет. Поиск начинается с загрузки анализируемой последовательности, затем выбирается геном, который будет анализироваться на

наличие целевых и нецелевых последовательностей для выбранных гидПНК. Предлагается довольно большое количество геномов, среди которых есть и растительные – ячмень, щетинник, кукуруза, рис, соя, томат, арабидопсис и мох *Physcomitrella patens*. Как и в ряде других программ можно связаться с разработчиками на предмет введения в программу какого-то другого отсутствующего генома. По умолчанию ведется поиск гидПНК под *SpCas9* нуклеазу. Для повышения эффективности поиска программа позволяет учитывать опционально некоторые другие параметры.

Рассчитанная на работу off-line программа **sgRNACas9** (<http://www.biooools.com>) [Xie et al., 2014] может быть скачана с сайта <https://sourceforge.net/projects/sgrnacas9>. Она позволяет вести дизайн гидПНК с минимальным количеством образующихся нецелевых сайтов связывания. При этом в анализ берутся обе цепи ДНК и находятся парные гидПНК, требующиеся при использовании Cas9 нуклеазы или при нокин-редактировании геномов. По завершению поиска и подбора гидПНК эта программа осуществляет дизайн олигонуклеотидов, которые необходимы для клонирования в подходящем векторе участка, кодирующего выбранные гидПНК.

Web-ресурс **COSMID (CRISPR Off-target Sites with Mismatches, Insertions, and Deletions)** [Cradick et al., 2014], доступный по адресу <https://crispr.bme.gatech.edu>, ориентирован на дизайн гидПНК с исключением нецелевых мест редактирования в геномах ряда млекопитающих, включая человека, а также в геноме нематоды. При этом важным отличием данной программы от других аналогичных стал учет возможных инделов (до двух) в спейсерах/протоспейсерах, помимо обычно допускаемых неспариваний, что увеличило возможности подбора гидПНК. Другая особенность ресурса COSMID заключается в дизайне праймеров для последующей амплификации с целью детекции произошедшего редактирования с рекомендуемым средним размером ампликона 275 п.н. Работа с программой начинается с выбора того или иного генома из предлагаемого списка и введения анализируемой нуклеотидной последовательности длиной до 55 нуклеотидов, после чего предлагается ввести с клавиатуры РАМ последовательность (можно оставить строку пустой), указать неспаривания или инделы и запустить поиск.

Web-ресурс **CRISPR-P 2.0** (<http://cbi.hzau.edu.cn/crispr/>) стал первой специализированной программой для дизайна гидПНК, рассчитанной на редактирование геномов растений [Lei et al., 2014]. По сравнению с некоторыми имевшимися тогда программами

CRISPR-Design, Cas-OFFinder, E-CRISP, Cas9 Design помимо того, что CRISPR-P была нацелена на анализ растительных геномов, она выдавала еще полную визуальную информацию о всех целевых и нецелевых сайтах, включая подсчеты эффективности воздействия на них подобранными гидПНК, заодно ведя поиск сайтов узнавания рестрикционных эндонуклеаз, причем на все это уходило секунды. Недавно написана новая версия данной программы **CRISPR-P-2.0** (<http://cbi.hzau.edu.cn/CRISPR2/>) [Liu et al., 2017]. Программа CRISPR-P-2.0 сохранила интерфейс прежней версии, но приобрела много новых важных черт. Так, список геномов растений, по которым проводится поиск мест, способных служить протоспейсерами, пополнился 23 геномами и теперь состоит из 49 геномов, относящихся к 45 видам из разных семейств - *Arabidopsis lyrata*, *A.thaliana*, *Arachis duranensis*, *A.ipsaensis*, *Brachypodium distachyon*, *Brassica napus*, *B.oleracea*, *B.rapa*, *Capsella rubella*, *Chlamydomonas reinhardtii*, *Citrullus lanatus*, *Citrus sinensis*, *Coffea canephora*, *Cucumis melo*, *C.sativus*, *Cyanidioschyzon merolae*, *Fragaria vesca*, *Glycine max*, *Gossypium hirsutum*, *G.raimondii*, *Lentinula edodes*, *L.edodes*, *Lotus japonicus*, *Manihot esculenta*, *Marchantia polymorpha*, *Medicago truncatula*, *Musa acuminata*, *Nicotiana benthamiana*, *Oryza brachyantha*, *O.glaberrima*, *O.indica*, *O.sativa*, *Panicum virgatum*, *Physcomitrella patens*, *Populus trichocarpa*, *Ricinus communis*, *Selaginella moellendorffii*, *Setaria italica*, *Solanum lycopersicum*, *S.tuberosum*, *Sorghum bicolor*, *Utricularia gibba*, *Vitis vinifera*, *Zea mays*. Оценка целевых и нецелевых сайтов в геномах теперь проводится, исходя из последних воззрений на их взаимодействия с *SpCas9* и другими нуклеазами, число которых заметно возросло. Программа CRISPR-P-2.0 предполагает подбор протоспейсеров для их расщепления помимо *SpCas9* следующими нуклеазами – *StCas9*, *NmCas9*, *SaCas9*, а также нуклеазами из группы *Cpf1* – *AsCpf1*, *LbCpf1*, *FnCpf1* и целым рядом других, что сильно расширяет возможности экспериментаторов по подбору наиболее оптимальных мишеней в редактируемых геномах растений. В дополнение к этому программа CRISPR-P-2.0 позволяет устанавливать длину спейсерной части гидПНК в пределах от 15 до 22 нуклеотидов. По завершению поиска данная программа выдает, в том числе визуально хорошо организованные результаты в виде расположения на линейном графике мест взаимодействия спейсеров и протоспейсеров, включая информацию о GC-составах, сайтах рестрикции, оценочной характеристики подобранных гидПНК для on- и off-target сайтов, а также изображение предполагаемой вторичной структуры гидПНК. В качестве

дополнительной информации пользователю предоставляются сведения о микргомологии соседних участков с местами разреза ДНК нуклеазами. Все это в целом делает программу CRISPR-P-2.0 одной из наиболее удобных и обладающей широкими возможностями.

Web-ресурс **CHOPCHOP v2** (<http://chopchop.cbu.uib.no>) [Labun et al., 2016] позволяет искать гидПНК, работающие в комплексе с Cas9 нуклеазой, ее никазными вариантами, а также с нуклеазой *Cpf1*. Последних двух ферментов в предыдущей версии этой программы **CHOPCHOP** (<https://chopchop.rc.fas.harvard.edu>) [Montague et al., 2014] не было. Поиск начинается с введения анализируемого гена или координат геномной последовательности, затем проводится выбор генома из довольно большого списка, содержащего по два вида арахисов и нута, арабидопсис, томат и теф (*Eragrostis tef*) из злаковых.

Web-ресурс **CROP IT (CRISPR/Cas9 Off-target Prediction and Identification Tool)** [Singh et al., 2015] (<http://www.adlilab.org/CROP-IT/cas9tool.html>) рассчитан на дизайн гидПНК и выявление потенциальных нецелевых сайтов редактирования всего в двух геномах (человека и мыши). Но при этом заявлено, что программа учитывает состояние хроматина в этих местах, что выгодно отличает ее от аналогичных ресурсов. Работа начинается с введения имени экспериментатора, организации, где он работает, и электронной почты, затем предлагается выбрать геном мыши или человека и участок PAM из двух на выбор. Перед запуском программы надо решить все ли найденные участки должны быть показаны или только верхние тысяча или пятьсот.

Программа **CRISPR Primer Designer** [Yan et al., 2015] может быть скачана с сайта http://plantsignal.cn/CRISPR/crispr_primer_designer.html. Возможность нецелевых сайтов редактирования отслеживается в геномах человека, мыши, дрозофилы, а также риса и арабидопсиса. При этом данная программа автоматически обращается к BLAST, генерируя наиболее оптимальные спейсеры для дизайна гидПНК.

Программа **CRISPRdirect** (<http://crispr.dbcls.jp>) [Naito et al., 2015] является простым и удобным web-сервером, обеспечивающим дизайн гидПНК. Экспериментатор может вручную или из файла в формате FASTA ввести в соответствующее окно анализируемую последовательность и затем ввести с клавиатуры PAM участок, однако его ориентация фиксирована. В ниспадающем меню необходимо выбрать геном, в котором следует искать целевые и нецелевые сайты связывания. Можно отметить, что к данной программе «привязано» немало растительных геномов – разные виды риса, банан, папайя, соя,

ячмень, мягкая пшеница, сорго, какао, капуста, зеленые и красные водоросли, рапс, виноград и некоторые другие. Количество целевых сайтов в геноме отображается с использованием k -мерного счетчика Jellyfish.

Программа **ССТop** (CRISPR/Cas9 Target online predictor) (<http://crispr.cos.uni-heidelberg.de>) [Stemmer et al., 2015] позволяет осуществлять дизайн гидПНК и искать для них целевые и нецелевые сайты в геномах различных организмов, среди которых около полутора десятков - растения. Допускается искать подходящие для гидПНК участки только в загружаемой через буфер обмена весьма небольшой последовательности из 500 нуклеотидов. При этом предоставляется довольно широкий выбор нуклеаз Cas9 из разных видов бактерий, а также Cpf1 из *Acadaminoccus*.

Длина спейсерной последовательности гидПНК может варьировать в пределах от 15 до 23 нуклеотидов. Можно указать допустимое количество неспариваний. После этого предстоит выбрать геном из ниспадающего окошка и начать поиск.

Программа **CRISPR MultiTargeter** (<http://www.multicrispr.net>) [Prykhozhiy et al., 2015] интересна тем, что позволяет проводить множественное выравнивание гомологичных последовательностей для последующего выбора в качестве протоспейсера одинаковых участков. В соответствующее окно загружаются через буфер обмена или в виде файлов в FASTA формате последовательности, длина которых в общей сложности не должна превышать 50 т.п.н. На сайте приведен демо-вариант подбора гидПНК для двух гомологичных генов аквариумной рыбки *Dario rerio*, позволяющий оценить удобство данной программы. При этом пользователю дозволяется самостоятельно определять тип нуклеазы, указывая конкретные последовательности, прилегающие к протоспейсеру, его расположение, длину спейсера, наличие неспариваний, что делает программу довольно гибкой.

Программа **COD** (Cas9 Online Designer) (<http://cas9.wicp.net>) после того, как в нее был добавлен модуль, ищущий нецелевые сайты в анализируемых геномах, превратилась в **COD2** [Guo et al., 2015]. Дизайн гидПНК начинается с введения анализируемой последовательности длиной до 400 нуклеотидов, затем необходимо выбрать конкретный геном из относительно небольшого списка, где имеются и растительные объекты (рис, кукуруза, арабидопсис) и выбрать длину вариабельной части гидПНК (17-20 нуклеотидов), задать барьер выявления off-target сайтов и запустить поиск. На этой же web-странице приведен образец выдаваемого программой результата и рекомендации по поиску

гидПНК в случае использования никазы nCas9 (D10A).

Требующая установки на компьютер пользователя программа **Protospacer Workbench** позволяет искать гидПНК для фактически любых геномов, включая секвенированные неполностью [MacPherson, Scherf, 2015]. Проведенное авторами сравнение возможностей программы Protospacer Workbench с некоторыми другими программами дизайна гидПНК (E-CRISPR, Optimized CRISPR Design, ZiFiT, SSFinder, sgRNAscas9, Cas-Offinder), как работающими в виде web-ресурсов, так и установленными на компьютеры пользователей показало по многим параметрам заметное превосходство Protospacer Workbench. Программа Protospacer Workbench работает в операционной системе Mac (версия не позднее Mac OS X 10.7) и может быть загружена с сайта <http://www.protospacer.com>. На данный момент ведется разработка (на этапе бета-тестирования) данной программы для операционной системы Windows 7/8. Пользователь может зарегистрироваться на вышеуказанном сайте и получить доступ к тестированию программы. Для Linux систем разработка программы не планируется по причине малой востребованности среди пользователей (по мнению разработчиков).

Работа в программе **CRISPR Genome Analysis Tool – CGAT** (<http://cbc.gdcb.iastate.edu/cgat/>) [Brazelton, Jr. et al., 2015] состоит из двух этапов – на первом выбирается конкретный ген для редактирования или последовательность нуклеотидов вводится через буфер обмена и после установления некоторых параметров (длина гидПНК, желательный GC-состав) запускается анализ, после завершения которого опционально можно, выбрав из пяти видов растений (ячмень, кукуруза, арахис (два вида), соя и рис) определить потенциальные места нецелевого редактирования их геномов.

Программа **sgRNA Scorer 2.0** [Chari et al., 2017] (<https://crispr.med.harvard.edu/sgRNAScorer/>) является усовершенствованной версией **sgRNA Scorer 1.0**, которая может быть скачана и установлена на компьютере пользователя для автономной работы (<https://crispr.med.harvard.edu/sgRNAScorerV1/>) [Chari et al., 2015]. Работа с sgRNA Scorer 2.0 начинается с ввода анализируемой нуклеотидной последовательности длиной до 1 млн.п.н. через буфер обмена в FASTA формате или загрузкой нужного файла из компьютера пользователя. Затем предлагается решить - требуется ли проводить поиск возможных мест нецелевого редактирования в геномах на выбор, который ограничивается всего 14 предложениями, преимущественно человеком и

другими млекопитающими (позвоночными) и дрожжами или обойтись без него. Результаты анализа можно получить на введенный в специальном окне адрес электронной почты в течение 24 часов, причем найденные спейсеры будут ранжированы по эффективности их действия, оцененной *in silico*. Данная программа дает возможность искать протоспейсеры, которые могут редактироваться разными нуклеазами. По умолчанию программа использует нуклеазу *SpCas9*. Помимо нее предлагается выбор нуклеаз из *Staphylococcus aureus*, *Neisseria meningitidis*, *Streptococcus thermophilus* 1 и *S.thermophilus* 3, а также нуклеазы *AsCpf1*. Кроме этого допускается вести анализ для неких других («Other») нуклеаз, PAM которых следует вводить с клавиатуры. Далее предлагается выбрать длину спейсера (от 14 до 24 нуклеотидов) и 5'- или 3'-расположение PAM. После чего начинается поиск целевых и нецелевых (в случае необходимости off-target анализа) мест во введенной последовательности и в конкретном геноме. Кроме того, на сайте размещена программа **CRISPR GA** (Genome Analyzer) для оценки качества эксперимента по редактированию генома с учетом вставок, делеций и гомологичной рекомбинации. Genome Analyzer предоставляет отчет для выбранного локуса, который включает количественную оценку отредактированного сайта и анализ обнаруженных различных изменений. Платформа отображает данные, оценивает и находит вставки и удаления, вычисляет эффективность замены аллеля и предоставляет отчет, объединяющий всю информацию. Другим программным продуктом лаборатории J.Church для анализа CRISPR/Cas систем редактирования геномов является программа **CasFinder** (<http://arep.med.harvard.edu/CasFinder/>) [Aach et al., 2014]. Отличительной чертой CasFinder явилась простота и быстрота использования при соответственно меньшем числе контролируемых параметров по сравнению с программой sgRNA Scorer 2.0.

Web-ресурс **CRISPRScan** (<http://www.crisprscan.org>) [Moreno-Mateos et al., 2015] помимо того, что находит гидПНК в конкретных генах ряда животных организмов или в любой введенной экспериментатором последовательности, содержит еще базу данных по предварительно подобранным для тех же видов гидПНК, о которой речь пойдет в соответствующем разделе. Работа с программой дизайна гидПНК заключается в выборе организма из двух видов рыб (трех геномов), земноводного, двух видов млекопитающих (мышь и человек) и насекомого дрозофилы, после чего необходимо выбрать должен ли это быть ген или его транскрипт, затем тот,

который конкретно нужен и запустить поиск. При обнаружении подходящих гидПНК во введенной экспериментатором последовательности также предлагается выбрать один из имеющихся во встроенной базе данных восьми геномов для выявления в них нецелевых сайтов редактирования или отказаться от такой возможности, выбрав в этом окне вариант «No search». Облегчить работу с этой программой призваны имеющиеся примеры выдачи информации после соответствующих запросов.

WU-CRISPR - WashU gRNA Designer for CRISPR/Cas9 Knockout (<http://crispr.wustl.edu>) [Wong et al., 2015]. Программа WU-CRISPR может использоваться как в сети, так и автономно, ведя анализ предварительно подобранных гидПНК в геномах только человека и мыши, а также осуществлять *de novo* дизайн гидПНК по введенной через буфер обмена последовательности нуклеотидов длиной до 30 т.п.н. Также небольшой web-ресурс **Off-Spotter** (<https://cm.jefferson.edu/Off-Spotter/>) [Pliatsika, Rigoutsos, 2015] позволяет вести поиск гидПНК (допускается максимально 5 неспариваний) для нескольких Cas9 нуклеаз (выбор из четырех PAM) во вводимой последовательности (не более 1000 нуклеотидов) и нецелевых сайтов связывания в нескольких геномах (два человеческих, мышь и дрожжи).

Специфическая программа дизайна гидПНК для редактирования геномов патогенных микроорганизмов **EuPaGDT** (Eukaryotic Pathogen gRNA Design Tool) [Peng, Tarleton, 2015] находится по адресу <http://grna.ctegd.uga.edu>. При ее использовании первым делом предлагается назвать свою работу, затем выбрать тип нуклеазы (*SpCas9* или *SaCas9*, также имеется возможность ввести PAM последовательность с клавиатуры), после чего необходимо определиться с различными опциями в виде - guide RNA search parameters; on-target search parameters; off-target search parameters; HDR repair template parameters. Далее предстоит выбрать группу патогенных организмов (*Amoeba*, *Cryptosporidium*, *Fungi*, *Giardia*, *Microsporidia*, *Oomycetes*, *Piroplasma*, *Plasmodium*, *Toxoplasma*, *Trichomonas*, *Trypanosomatid*) и в ниспадающем меню (для каждой группы свое) выбрать геном конкретного вида и начать поиск. При этом есть опции - Courtesy Uploads и Custom genome, благодаря которым можно загрузить соответствующий файл. При редактировании генома в режиме нокин предлагается подобрать гомологичные участки, фланкирующие места редактирования (по умолчанию длиной в 30 нуклеотидов), содержащие внутри, например, последовательность из трех терминирующих кодонов в трех рамках считывания (например, TAGATAGATAG - всего 11 нуклеотидов).

Web-сервер **Breaking Cas**
[\(http://bioinfogp.cnb.csic.es/tools/breakingcas/\)](http://bioinfogp.cnb.csic.es/tools/breakingcas/)

[Oliveros et al., 2016] выгодно отличается от других программ дизайна гидПНК тем, что предоставляется возможность искать нецелевые сайты во всех известных к настоящему времени эукариотических геномах, которых по последним релизам базы данных Ensembl (<http://ensemblgenomes.org>) уже более 700. В том числе доступны многочисленные геномы растений (http://bioinfogp.cnb.csic.es/tools/breakingcas/?gset=4x2_GENOMES_EnsemblGenomes_35). Дизайн гидПНК с помощью web-сервера Breaking Cas начинается с выбора конкретного генома для чего необходимо или открыть алфавитный список или в имеющейся строке ввести несколько первых букв тривиального названия организма или его латинского обозначения, причем неважно будут ли они принадлежать роду или виду. Далее следует загрузить одну или несколько последовательностей определенных генов через буфер обмена или выбрав соответствующие файлы, не превышая в общей сложности 20 тысяч нуклеотидов. Затем экспериментатору предлагается выбрать тип нуклеазы из трех вариантов Cas9 и трех вариантов Cpf1. Помимо них допускается введение с клавиатуры и каких-либо других ПAM последовательностей для других нуклеаз, что дает широкие возможности целевого редактирования конкретных мест в выбранном геноме. Есть еще ряд полезных опций.

Данная программа также позволяет вести поиск мест редактирования геномов с помощью нуклеазы другого типа NgAgo Argonaute, для которой не требуется участок ПAM, однако эта опция уже стала неактуальной ввиду того, что для данного фермента не получила подтверждения его каталитическая активность, о чем подробно говорится в другой статье в этом номере журнала [Кулуев и др., 2017].

Работающая off-line программа **CRISPR Library Designer (CLD)** [Heiwer et al., 2016] доступна для скачивания с этого сайта - <https://github.com/boutrosfab/cld>. Она позволяет создавать библиотеки гидПНК для выбранных генов конкретного организма, геном которого может быть также взят из базы данных Ensembl (<http://ensemblgenomes.org>), для чего составляется специальный лист с перечнем таких генов. После сканирования выбранных генов на возможность их редактирования отбираются наиболее приемлемые варианты гидПНК и осуществляется аннотирование мест их связывания.

Компьютерная программа **PhytoCRISP-Ex** (<http://www.phytoCRISPex.biologie.ens.fr/CRISP-Ex/>), работающая как в сети, так и допускающая

возможность установки на компьютер пользователя (на платформе Linux), предназначена для дизайна гидПНК для редактирования геномов фитопланктона более десятка видов (*Phaeodactylum tricornutum*, *Chlamydomonas reinhardtii*, *Chlorella vulgaris*, *Fistulifera*, *Fragilariopsis cylindrus*, *Nannochloropsis gaditana*, *Ostreococcus lucimarinus*, *O.tauri*, *Thalassiosira oceanica*, *T.pseudonana*, *Emiliania huxleyi*, *Psuedo-nitzschia multiseriis* и *Volvox carteri*) [Rastogi et al., 2016]. Работа с программой начинается с загрузки интересующих экспериментатора последовательностей ДНК (или в виде файла либо через буфер обмена), затем требуется выбрать один из вышеперечисленных геномов и тот или иной участок ПAM (NGG или NAG), а также указать стартовый нуклеотид транскрипции (G или N). После завершения поиска его результаты могут быть скачаны в виде архивированного файла в отдельном окне.

Программа **CRISPOR v3.0** (<http://crispor.tefor.net>) [Haeussler et al., 2016] позволяет быстро осуществлять дизайн гидПНК для CRISPR/Cas систем редактирования геномов. В апреле 2017 г. заработала бета-версия данной программы **V4.3** с помощью которой в три шага предлагается выбрать оптимальные протоспейсерные последовательности. На первом шаге через буфер обмена вводится анализируемая последовательность (до 1000 нуклеотидов), на втором - предлагается выбрать один из геномов, в котором будут искажаться нецелевые сайты, при этом какой-либо геном можно не выбирать (вариант - No Genome). На третьем шаге необходимо выбрать нуклеазу из предлагаемых 10 вариантов, включающих кроме стандартной *SpCas9* ее различные мутантные формы с измененными ПAM-участками, а также другие ортологичные нуклеазы этого типа. В дополнении к ним можно выбрать нуклеазу *AsCpf1* из бактерии *Acidaminococcus sp.* Также весьма важным является то, что среди предлагаемых почти 90 геномов немало растительных. Так, среди них фигурируют *Arabidopsis lyrata*, *A.thaliana*, соя, томат, картофель, дыня, папайя, виноград, фасоль, эвкалипт, земляника, лен, один из видов хлопчатника *Gossypium raimondii*, яблоня, рис, кукуруза, персик, тополь, какао, клещевина, люцерна *Medicago truncatula*, петуния, рапс, два вида табака *Nicotiana tabacum* и *N.benthaliana*, а также еще целый ряд видов злаковых, цитрусовых и др. Отдельно упомянем, что можно выбрать и геном пчелы. При этом помимо, появляющегося окна с этими геномами, на сайте программы говорится, что есть возможность вести анализ по 177 геномам и для этого надо связаться с разработчиками.

Web-ресурс **CRISPR-DO** (Design and Optimization) [Ma et al., 2016], расположенный по адресу <http://cistrome.org/crispr/>, после завершения поиска подходящих гидРНК для нокаутных экспериментов с нуклеазой Cas9, сопровождает их довольно подробной аннотацией, указывающей экзоны, сайты, гиперчувствительные к ДНКазе I, однонуклеотидные замены. Работа с программой начинается с введения адреса электронной почты производящего анализ, затем предлагается выбрать какой-либо геном из небольшого списка (растений нет), длину спейсера (19 или 20 нуклеотидов) и интересующий регион генома, после чего предстоит указать ожидаемую специфичность гидРНК (от 0 до 100) и эффективность ее действия (от 0,3 до 1,4) и задать поиск.

Воспользоваться web-ресурсом **CrisprGold** [Chu et al., 2016] академические учреждения могут бесплатно, при этом авторы сайта (<http://crisprgold.mdc-berlin.de>) отмечают, что он еще продолжает совершенствоваться. Сейчас с помощью CrisprGold можно подобрать гидРНК и проверить возможное нецелевое редактирование с ними по трем геномам – человека, мыши и нематоды. Анализируемая последовательность ДНК вводится или через буфер обмена или, используя название гена. Затем предлагается выбрать число желательных гидРНК для намеченного гена, которое может быть ограничено 10 вариантами или допускается выбрать все возможные - вариант «All». Аналогично ограничивается число нецелевых сайтов редактирования – в диапазоне от 5 до 100 или все возможные. Имеется такая опция как выбор вектора для клонирования – предлагается из трех возможных.

Web-ресурс **CT-Finder** [Zhu et al., 2016] (<http://bioinfolab.miamioh.edu/ct-finder/>) позволяет вести поиск гидРНК в трех вариациях. При использовании обычной нуклеазы Cas9 программа находит единичные гидРНК, при использовании по отдельности нуклеазы nCas9 (D10A) и нуклеазы Cas9, сшитой с ферментом FokI, в результате поиска предлагаются парные гидРНК. Поиск во всех случаях начинается с введения анализируемой последовательности, на втором шаге идет выбор генома из небольшого списка предлагаемых (из растительных – арабидопсис, рис и кукуруза). На третьем шаге вводится информация о РАМ, длине гидРНК, протяженности спейсерного участка. На четвертом опционально вводятся допустимые количества неспаренных нуклеотидов в разных областях спейсерной последовательности гидРНК, после чего дается команда начать поиск. В случае использования нуклеазы и нуклеазы с FokI доменом на третьем шаге задаются расстояния между парными гидРНК - от 0 до 10 нуклеотидов и от 14 до 17

нуклеотидов для первого и второго фермента соответственно. Также указывается количество анализируемых цепей ДНК – одна или две.

Web-ресурс **Benchling** (<https://benchling.com>) разграничивает его применение для коммерческих и академических организаций и для последних он бесплатен (<https://benchling.com/academic>), но при этом необходимо зарегистрироваться. Данный ресурс позволяет планировать эксперименты по редактированию геномов (нокаутированию отдельных генов) и осуществлять дизайн гидРНК под широкий спектр Cas нуклеаз.

Портал CRISPR/Cas9 редактирования геномов растений **CRISPR-Plant** (<http://genome.arizona.edu/crispr/>) позволяет подбирать гидРНК для следующих растений – *Arabidopsis thaliana*, *Brachypodium distachyon*, *Glycine max*, *Medicago truncatula*, *Oryza sativa*, *Sorghum bicolor*, *Solanum lycopersicum*, *Zea mays*. Можно выбрать какой-либо из этих геномов, конкретную хромосому и локализацию последовательностей в ней, после чего запустить поиск, который при анализе 10 тысяч нуклеотидов длится секунды и программа выдает все возможные гидРНК для этой последовательности, которые попадают в две группы – class 0.0 и class 1.0, отличающиеся тем, что первая группа – хорошие гидРНК с минимальной возможностью нецелевого спаривания, а вторая – менее хорошие и их лучше не использовать. Аналогичный анализ для этих же геномов растений может быть проведен для конкретных генов, которые должны выбираться с помощью их идентификатора.

Облачный сервис **DESKGEN** (<https://www.deskgen.com/landing/cloud>) бесплатно доступен для некоммерческих организаций, в связи с чем требуется регистрация, занимающая менее минуты, после чего становится возможна работа с такими разделами программы как Knockout; Knockin; Guide Picker [Hough et al., 2017] и Genome Editor. В настоящее время для анализа с ними доступно 15 геномов, включая из растений - ячмень, мягкую пшеницу, кукурузу и арабидопсис, однако программа Guide Picker работает только с геномами человека и мыши. Последовательность действий при этих анализах довольно стандартна – предстоит сначала выбрать референсный геном, затем конкретный ген, тип нуклеазы и длину варибельного участка гидРНК - от 17 до 23 нуклеотидов. Возможен дизайн гидРНК для образования комплексов с 10 различными нуклеазами, включая три из группы Cpf1, однако появляется предупреждение, что программа настроена на использование классической Cas9 нуклеазы и выдаваемые результаты могут быть не совсем корректны. При этом возможности DESKGEN сервиса довольно широки и сопровождаются

подробными руководствами. Данный облачный сервис регулярно расширяется, пополняются обучающие материалы и обновляются приложения.

Небольшая программа **GuideScan** (<http://guidescan.com>) [Perez et al., 2017] позволяет вести поиск гидРНК, которые будут образовывать комплексы или с нуклеазой *SpCas9* или *AsCpf1/LbCpf1*. Программа **Green Listed** (<http://greenlisted.cmm.ki.se>) [Panda et al., 2017], если судить по названию, возможно, должна была быть нацелена на анализ растительных геномов, но на самом деле с ее помощью предлагается искать нецелевые сайты только в геномах человека и мыши.

Такой web-ресурс как **CRISPEta** (**CRISPR Paired Excision Tool**) (<http://crispeta.crg.eu>) [Pulido-Quetglas et al., 2017] рассчитан на дизайн парных гидРНК для обычной *Cas9* нуклеазы для редактирования геномов человека, мыши, аквариумной рыбки *Dario rerio*, дрозофилы и нематоды. Причем разработчики утверждают, что это первая программа такого рода с подобными возможностями. CRISPEta предполагает проведение анализа намеченного места редактирования, охватывающего прилежащие регионы, выбор протоспейсеров и дизайн парных гидРНК, после чего следует определение их приоритетности. Благодаря тому, что упомянутые выше геномы заранее проанализированы на предмет нахождения в них подходящих протоспейсеров, то поиск интересующих экспериментатора ведется в ускоренном варианте. Так, например, для генома человека hg19 размером около 3,14 млрд.п.н. существует почти 300 миллионов потенциальных протоспейсеров (по числу РАМ участков NGG), но в результате их предварительного анализа на предмет присутствия потенциальных нецелевых сайтов, отличающихся друг от друга двумя неспаренными нуклеотидами, и оцененной потенциальной эффективности редактирования выбраковке подверглось 96% этих протоспейсеров, составив в среднем итоговую «плотность» в виде 6,6 протоспейсеров на тысячу пару нуклеотидов. Для небольших и проще организованных геномов нематоды и дрозофилы эти показатели составили – 10,6 и 19,8 протоспейсеров на килобазу соответственно. Работа с программой начинается с введения в соответствующее окно последовательности ДНК, которую предстоит отредактировать, при этом задаются выше- и нижележащие участки вокруг мишени, к которым следует и наоборот не следует подбирать гидРНК. Имеется довольно много опций, выбор которых позволяет конкретизировать предъявляемые требования к гидРНК.

Web-ресурс **CRISPR Knockout Guide Designer** (<https://design.synthego.com/#/>) предоставляет возможность дизайна гидРНК для нокаутирования генов с помощью нуклеазы *SpCas9*. Таких мишеней для редактирования на середину мая 2017 г. этой программой можно выбрать из более чем 61 млрд. последовательностей, принадлежащих 103 тысячам 166 геномам организмов всех уровней генетической сложности. При этом пользователь должен сначала указать организм, вводя в соответствующую строку его или латинское или тривиальное название и из ниспадающего меню выбрать именно тот вид, который необходимо редактировать, после чего в соседнее окно надо ввести название гена или его идентификационный номер в GenBank и запустить поиск. Программа позволяет не только искать новые мишени, но и валидировать найденные ранее протоспейсеры путем введения в строку поиска последовательности из 20 нуклеотидов.

Web-ресурс **CRISPR4P** [Rodríguez-López et al., 2017] (<http://bahlerweb.cs.ucl.ac.uk/cgi-bin/crispr4p/webapp.py>) предполагает выбор в геноме дрожжей *Shizosaccharomyces pombe* протоспейсеров, для синтеза гидРНК к которым подбираются праймеры для клонирования в подходящем векторе, праймеры для гомологичной рекомбинации и праймеры для контроля произошедшего редактирования. Для работы off-line CRISPR4P может быть скачана с этого адреса - <https://github.com/Bahler-Lab/crispr4p>. Программа **predictSGRNA** [Kuan et al., 2017] может быть скачана с сайта <http://www.ams.sunysb.edu/~pfkuan/software.html#predictsgn>. Авторы отмечают, что при дизайне гидРНК они постарались учесть термодинамику нуклеотидных последовательностей этих молекул и формирование возможных вторичных структур.

Программа **CRISPR-FOCUS** [Cao et al., 2017] является web-ресурсом (<http://cistrome.org/crispr-focus/>) с помощью которого можно подобрать гидРНК с учетом их возможного нецелевого связывания в геномах двух организмов – человека и мыши. При этом можно ввести анализируемую последовательность (до 1000 нуклеотидов) или через буфер обмена, или указать идентификационный номер, либо официальное название гена после чего предстоит выбрать длину спейсера (19 или 20 нуклеотидов) и задать желаемое число гидРНК на редактируемый ген (до 30) и запустить поиск. Имеется еще ряд опций в виде, например, дизайна контрольных гидРНК (как положительных, так и отрицательных) и отображения всей молекулы гидРНК, включая ее константную часть.

Целый пакет прикладных программ **CRISPR-GE** (**G**enome **E**diting) [Xie X. et al., 2017], состоящий из следующих инструментов - **seqDownload**, **targetDesign**, **offTarget**, **primerDesign-V**, **primerDesign-A**, **DSDecodeM**, находится по адресу <http://skl.scau.edu.cn>. Этот программный комплекс нацелен на редактирование геномов растений, кои в нем представлены 25 видами (*Oryza sativa japonica*, *O.sativa indica*, *O.indica*, *O.japonica*, *Arabidopsis thaliana*, *A.lyrata*, *Brachypodium distachyon*, *Brassica napus*, *B.oleracea var. oleracea*, *B.rapa*, *Camelina sativa*, *Cyanidioschyzon merolae*, *Glycine max*, *Hordeum vulgare*, *Medicago truncatula*, *Musa acuminata*, *Populus trichocarpa*, *Prunus persica*, *Setaria italica*, *Solanum lycopersicum*, *S.tuberosum*, *Sorghum bicolor*, *Theobroma cacao*, *Vitis vinifera*, *Zea mays*) и в общей сложности 27 геномами, а также геномами четырех других организмов - *Homo sapiens*, *Mus musculus*, *Danio rerio* и *Caenorhabditis elegans*. Дизайн гидРНК с помощью программы **targetDesign** начинается с выбора нуклеазы и характерной для нее РАМ последовательности, среди которых *SpCas9* (NGG), *FnCpf1* (TTN) и *AsCpf1* (TTTN). Также пользователю допускается указать иной тип РАМ. Далее необходимо выбрать конкретный геном из ниспадающего списка видов, приведенного выше, при этом допускается не указать никакого генома, выбрав «None». Далее в соответствующее окно через буфер обмена вносится анализируемая последовательность (не более 10 тысяч нуклеотидов) или конкретный ген, причем можно даже задать размеры фланкирующих его последовательностей (5 тысяч и одна тысяча для предшествующего и последующего участков соответственно). После чего следует запустить поиск, результаты которого программа выдаст в табличной форме с 9-ю столбцами. В первом столбце можно отметить соответствующие квадратики и для этих гидРНК будет в дальнейшем подобран комплект праймеров для конструирования нужной последовательности ДНК для ее клонирования в соответствующем векторе. Второй столбец сообщит экспериментатору о неудачных гидРНК, которые будут отмечены разным количеством красных восклицательных знаков (!, !!, !!!). В третьем столбце будут приведены подобранные последовательности спейсерной области гидРНК с отдельным указанием РАМ участка. Следующий ряд столбцов сообщит информацию о позиции найденного участка в конкретном гене, его GC-составе и укажет цепь ДНК и кодирующая ли это область. Последние столбцы позволят оценить вероятность нецелевого редактирования, которое проводится с помощью программы **offTarget**. Далее можно перейти к дизайну праймеров с помощью программы **primerDesign** и ее

двух подпрограмм - **primerDesign-V** и **primerDesign-A**. Первая, как уже говорилось, отвечает за дизайн олигонуклеотидов для клонирования, а вторая – за амплификацию участков редактирования с целью подтверждения имевшего места данного процесса.

Программа **offTarget** может работать независимо от пакета **targetDesign** и анализировать введенную в нее последовательность гидРНК на предмет возможности редактирования с ее участием геномов в нецелевых местах. Порядок работы сходен с таковым при поиске гидРНК. Также из тех же предлагаемых перечней необходимо выбрать тип нуклеазы и конкретный геном (за небольшим исключением представлены те же геномы), после чего следует ввести спейсерную последовательность гидРНК и запустить процесс поиска. Для удобства экспериментатора имеется демо-версия такого анализа.

Программа **seqDownload** облегчает экспериментатору поиск нужного участка генома из представленных во встроенной базе данных, который можно найти в том числе введя известную последовательность длиной не менее 35 нуклеотидов в соответствующее окошко.

Программа **ZiFiT Targeter** (**Z**inc **F**inger **T**argeter) (<http://zifit.partners.org/ZiFiT/>) была написана еще до того, как появился CRISPR/Cas метод редактирования геномов [Sander et al., 2007], да и ее обновление было сделано несколько раньше [Sander et al., 2010] и поэтому она наряду с появившимся в ней модулем для дизайна гидРНК, пригодна и для осуществления экспериментов по редактированию геномов с помощью ZFN и TALEN подходов. Тем не менее, в современной версии 4.2 (http://zifit.partners.org/ZiFiT/Program_use.aspx) предлагается также использовать CRISPR подход. Причем имеются три подпрограммы - CRISPR RFNs (RNA-guided FokI Nucleases); CRISPR/Cas Nucleases; CRISPR/Cas Nickase, запускающиеся независимо. Дизайн начинается с введения в FASTA формате нуклеотидной последовательности, в которой необходимо произвести редактирование, причем в квадратных скобках можно указать нуклеотид, на который хотелось, чтобы приходился двуцепочечный разрыв. Программа рекомендует проверить введенную последовательность на отсутствие в ней повторов с помощью программы RepeatMasker (<http://www.repeatmasker.org>). Можно выбрать тип промотора и длину спейсерного участка гидРНК после чего дается команда поиска и удаленный сервер выдаст приемлемые варианты гидРНК, которые можно затем проверить на наличие off-target сайтов в одном из нескольких геномов на выбор (растений нет). Также изначально для дизайна экспериментов по редактированию геномов с помощью TALEN технологии была написана

программа **Mojo Hand** [Neff et al., 2013], однако в современной ее версии (<http://www.talendesign.org>) исследователям предоставляется возможность осуществлять дизайн экспериментов, в том числе для систем CRISPR/Cas9 и Cas9 nickase.

Несмотря на то, что в нем допускается работать всего с несколькими геномами животных, включая человека, определенный интерес представляет web-ресурс **grID (gRNA Identification)** (<http://crispr.technology>) [Jaskula-Ranga, Zack, 2017], в котором сосредоточен целый ряд программ и утилит, а также собственная база данных. Так, с помощью программы **Advanced gRNA Search** можно для человека осуществить дизайн гидПНК для нокаутных или нокин экспериментов, выбрав в геноме или целевые экзоны, или целевые интроны, или 5'-, 3'-нетранслируемые области либо промоторные последовательности, введя идентификационный номер гена или его название из базы данных RefSeq. Выданные программой последовательности гидПНК ранжируются трижды по трем заложенным принципам, один из которых выработан для этого ресурса непосредственно. Далее, с помощью программы **Oligo Generator** для конкретной гидПНК будут подобраны последовательности олигонуклеотидов, необходимые для клонирования в подходящем векторе. Помимо этого, на данном сайте содержатся подробные протоколы для амплификации и проведения клонирования и трансфекции созданных конструкций, сопровождаемые вычислением эффективности нокин и нокаутных экспериментов для выбранной мишени и рекомендациями для проведения детекции разными методами произведенного редактирования.

Фирма General Electric Company, имеющая подразделение GE Healthcare, предоставляет on-line ресурс **CRISPR RNA Configurator** (<https://dharmacon.gelifsciences.com/gene-editing/crispr-rna-configurator/>) с помощью которого (требуется регистрация) можно осуществлять дизайн гидПНК, в том числе для нокин-экспериментов. При этом среди 33 предлагаемых геномов для поиска в них нецелевых мишеней нет ни одного растительного. Следует отметить, что руководство пользователя содержит раздел, где представлены типичные неудачи и пути их преодоления.

Программа **CRISPR-RT (RNA Targeting)** (<http://bioinfolab.miamioh.edu/CRISPR-RT/>) [Zhu et al., 2017] является первой программой дизайна гидПНК, где в качестве нуклеазы используется новый фермент C2c2, отнесенный к VI типу и второму классу Cas нуклеаз. В режиме on-line предлагается сначала ввести анализируемую последовательность в FASTA формате, затем выбрать один из 10 предзагруженных транскриптомов, из которых три – растительные –

Arabidopsis thaliana, *Oryza sativa* и *Zea mays*. При необходимости анализировать другие организмы рекомендуется связаться с разработчиками, указав свою электронную почту. На следующем шаге задаются параметры гидПНК и места локализации с учетом прилегающего нуклеотида к протоспейсеру, последовательность PFS, а также указать общую длину протоспейсера и длину якорного участка (по умолчанию – 28 и 10 соответственно). На заключительном шаге выбираются опции, где указывается допустимое число неспариваний в обеих областях варибельной части гидПНК и процесс поиска запускается. Имеется еще ряд опций, позволяющих сузить выдаваемые результаты.

Помимо использования CRISPR/Cas систем для геномного редактирования они находят применение еще для иных целей, рассмотренных нами довольно подробно в другой статье этого выпуска [Кулуев и др., 2017a]. Среди них - исследования с помощью CRISPR/dCas системы с инактивированной нуклеазой процессов функционирования отдельных генов и генных сетей путем их репрессии CRISPRi (interference или inhibition) или активации CRISPRa (activation). И кратко рассмотренные ниже две компьютерные программы нацелены на дизайн гидПНК не только для геномного редактирования, но и для использования CRISPR/dCas системы для CRISPRi/a экспериментов.

В названии программы **CRISPR-ERA** (<http://crispr-era.stanford.edu>) [Liu et al., 2015a] можно усмотреть игру слов (хотя сами авторы об этом не упоминают), так как наступившая сейчас в молекулярной биологии «эра» действительно может быть названа «CRISPRовской», но в аббревиатуре **ERA** на самом деле закодированы различные процессы использования CRISPR/Cas систем, поскольку с помощью этой программы предполагается осуществлять дизайн гидПНК для геномного редактирования (**E**ditting), ингибирования или репрессии (**R**epression) отдельных генов и их активации (**A**ctivation). При работе с CRISPR-ERA первоначально предлагается выбрать тип манипуляции с геномом (редактирование с помощью нуклеазы, редактирование с помощью никаз, репрессия гена или активация гена, после чего появляется возможность выбрать организм из небольшого списка (растений не содержит), указать конкретный ген или ввести его последовательность (не длиннее 5 т.п.н.) и запустить процесс поиска, выбрав еще ряд опций.

Web-ресурс **SSC (Sequence Scan for CRISPR)** [Xu et al., 2015], находящийся по адресу <http://cistrome.org/SSC/>, также предназначен для дизайна не только гидПНК для нокаутирования отдельных генов, но и при проведении работ по

CRISPR репрессии и CRISPR активации генных систем. Для нокаутных экспериментов при подборе гидРНК с помощью программы SSC первым делом необходимо ввести через буфер обмена интересующую исследователя последовательность ДНК (не более 10 тысяч нуклеотидов), выбрать длину спейсера (19 или 20 нуклеотидов) и запустить процесс поиска подходящих участков. Выдаваемые результаты могут быть сохранены в разных форматах, где наиболее эффективные гидРНК будут вверху списка. Причем оказывается предпочтение тем гидРНК, использование которых запрограммировано так, чтобы в месте разрыва ДНК в положении -3 от PAM находится цитозин.

В Интернете имеется еще целый ряд программ для дизайна гидРНК - **CasBLASTR** (<http://www.casblastr.org>), **Jack Lin's CRISPR/Cas9 gRNA finder** (<http://spot.colorado.edu/~slin/cas9.html>), **Azimuth** (<https://www.microsoft.com/en-us/research/project/azimuth>), **CRISPy-web** (<http://crispy.secondarymetabolites.org>) [Blin et al., 2016], **CRISPy CHO** (<http://staff.biosustain.dtu.dk/laeb/crispy>) [Ronda et al., 2014], **CRISPRer** (<http://jstacs.de/index.php/CRISPRer>), **CRISPR Specificity Analysis Tool** (<http://dharmacon.gelifsciences.com/resources/tools-and-calculators/crispr-specificity-tool/>), **CRISPR Mapper** (<http://crdd.osdd.net/servers/crisprge/mapper.php>), **CRISPR Domesticator** и **CRISPR Assembler** (<https://gbcloining.upv.es/tools/crisprs>) [Vazquez-Vilar et al., 2016], **Stupar Lab's CRISPR Design** (<http://stuparcrispr.cfans.umn.edu/CRISPR/>) [Michno et al., 2015], **CRISPR Efficiency Predictor** (<http://www.flyrnai.org/evaluateCrispr/>), **CRISPR Helper** (<http://123.57.239.141:8081>), **CRISPR sgRNA Design Tool** (<http://www.genscript.com/gRNA-design-tool.html>), **CREATE Designer (CRISPR Enabled Trackable genome Engineering)** (<http://www.thebioverse.org>) [Garst et al., 2017].

Основные принципы работы программ дизайна гидРНК

В работе программ по дизайну гидРНК для геномного CRISPR/Cas редактирования основной многократно выполняемой процедурой является подсчет количества вхождений подстрок некоторой длины k в длинной строке, классификация этих подстрок. Практически все программы работают по следующей схеме:

1. считывание всей последовательности генома (обычно задается в формате FASTA);
2. разделение исходной последовательности на части (в зависимости от выделяемого программой объема памяти);
3. поиск всех участков, которые соответствуют определенному образцу в геноме;

4. реорганизация последовательности для нового массива;
5. сравнение целевого участка с запрашиваемой последовательностью;
6. подсчет количества не совпавших оснований;
7. выбор последовательности в пределах допустимых критериев поиска, задаваемых при проведении анализа;
8. вывод отчета.

Можно выделить несколько основных методов, которые применяются в решении таких задач: метод «разделяй и властвуй» (разбиение большой задачи на несколько подзадач), рекурсивный метод, бинарный поиск и т.д. Для решения задачи классификации часто применяются инструменты регрессионного анализа, логистическая регрессия — это статистическая модель, используемая для предсказания вероятности возникновения некоторого события путем подгонки данных к логистической кривой. Во многих программных продуктах упоминается метод Bowtie (Bowtie 2 – более поздняя версия) – сверхбыстрый механизм короткого считывания, который выравнивает короткие последовательности ДНК. Bowtie проводит качественный поиск во множестве возможных выравниваний.

В результате проведения анализа последовательности возникает задача множественного сравнения данных. Для этого применяются статистические методы решения множественных сравнений. В одной из работ [Туско et al., 2016] упоминаются методы Холмма, Бенджамини-Хохберга, метод поправки Бонферрони, причем метод Холмма считается более мощным, жестким методом, чем поправка Бонферрони. Данные методы получили широкое применение в области биоинформатики в целом, так как при работе с геномами довольно часто возникают рутинные задачи многократного сравнения, оценки случайных величин. В ряде программ разработаны собственные эвристические алгоритмы, а также применяются методы построения суффиксного дерева или суффиксного массива (более экономная структура с точки зрения занимаемого объема памяти). Эти структуры данных позволяют выяснить, входит ли определенная строка S в некоторую строку M .

Для разработки программного обеспечения используются высокоуровневые языки программирования такие как C, C++, Python, C# с применением фреймворка OpenCL (например, модуль Cas-OFFinder), Perl с применением специализированной библиотеки BioPerl. Кроме того, существующие на сегодняшний день разработки, представленные в виде web-приложений, имеют

удобный, интуитивно понятный пользователю интерфейс. Данные web-приложения разработаны с помощью языков Java и PHP. Поддержка работы с базами данных осуществляется посредством системы управления базами данных MySQL.

Базы данных по подобранным гидПНК

В некоторых из упоминавшихся выше программах предусмотрено формирование неких библиотек подобранных гидПНК, однако в Интернете имеются и специализированные базы данных по гидПНК для CRISPR/Cas редактирования геномов различных организмов. Так, корейскими авторами создана база данных гидПНК **Cas-Database** [Park et al., 2016], входящая в комплект программных продуктов, представленных на специализированном портале **CRISPR RGEN Tools** (<http://www.rgenome.net>). На сайте The Wellcome Trust Sanger Institute в Англии поддерживается база данных по гидПНК человека (<http://www.sanger.ac.uk/science/tools/wge>) Wellcome Trust Sanger Institute Genome Editing database - **WGE** [Hodgkins et al., 2015]. Аналогичная база данных **CRISPRz** (<https://research.nhgri.nih.gov/CRISPRz/>) создана для такого модельного объекта как аквариумная рыбка *Danio rerio* [Varshney et al., 2015]. Универсальная база данных по гидПНК различных организмов **GenomeCRISPR** (<http://genomecrispr.org>) недавно создана, и авторы отмечают, что она быстро пополняется [Rauscher et al., 2017]. Ранее была создана база данных **CrisprGE** (<http://crdd.osdd.net/servers/crisprge/index.php>), претендующая на единый хаб для гидПНК для геномного редактирования разных видов организмов [Kaur et al., 2015]. Эта база данных имеет также встроенный модуль **Crispr Mapper**, позволяющий создавать гидПНК, анализируя введенную в FASTA формате нуклеотидную последовательность, однако при этом никаких опций исследователю, которые он хотел бы менять, не предусмотрено. Уже упоминавшийся web-ресурс **CRISPRScan** (<http://www.crisprscan.org>) [Moreno-Mateos et al., 2015] содержит довольно удобную в работе базу данных по предварительно подобранным для нескольких видов животных организмов гидПНК, скомпонованным по-хромосомно с выделением тремя цветами, символизирующими хорошую эффективность редактирования, слабую эффективность и вероятность нецелевого редактирования. Созданная для редактирования с помощью ZFN и TALEN технологий база данных **EENdb** [Xiao et al., 2012] в настоящее время начинает собирать сведения и для CRISPR/Cas редактирования.

Весьма интересна база данных **Grape-Crispr** (<http://biodb.sdau.edu.cn/gc/>) по гидПНК винограда *Vitis vinifera*, позволяющая проводить в ней поиск

спейсеров, которые (как специфичные, так и не очень) можно искать тройко – выбирая конкретную хромосому и указывая координаты в нуклеотидах; вводя идентификационный номер интересующего гена; проводя поиск по семействам белков. В статье, посвященной этой базе данных Grape-Crispr [Wang et al., 2016], проведен подробный анализ найденных в геноме винограда 35767960 потенциальных сайтов для CRISPR/Cas9 редактирования. Поиск протоспейсеров с PAM (NGG) участками проводился с помощью написанного авторами специального скрипта, разработанного на языке Perl, в результате чего удалось разделить их на несколько групп, исходя из мест локализации в геноме.

Недавно создана база данных **grID** (gRNA Identification) (<http://crispr.technology/database.html>) [Jaskula-Ranga, Zack, 2017], в которой хранятся заранее подобранные оптимальные гидПНК для редактирования генома человека с помощью нуклеаз Cas9, выделенных из разных микроорганизмов, включая их модифицированные генно-инженерные варианты, а также для нуклеазы Cpfl. Кроме генома человека эта база данных содержит подобранные гидПНК для еще нескольких геномов (мышь, крыса, нематода, аквариумная рыбка, дрожжи) для их редактирования с помощью нуклеазы Cas9.

Для пользователя довольно удобно, что базы данных работают в интерактивном режиме. На некоторых ресурсах допускается создавать свои небольшие базы, в которых может храниться необходимая информация для конкретного пользователя. В «личных» таблицах удобно проводить сортировку данных.

Программы для анализа отредактированных геномов

Учитывая, что геномное редактирование с помощью CRISPR/Cas технологии, может приводить к нежелательным (незапрограммированным) изменениям генома, то помимо того, что рекомендуется очень тщательно подбирать гидПНК, в важных случаях необходимо убеждаться, что редакция генома имела место только там, где предполагалась. Недавно была опубликована статья, в которой сообщалось о большом числе ошибок, внесенных Cas9 нуклеазой в геномы мышей, превышающих уровень обычных мутаций приблизительно в 20 раз [Shaefer et al., 2017]. Обнаружили они эти мутации путем полногеномного секвенирования отредактированных геномов двух мышей. Чтобы облегчить поиски измененных участков генома написано уже немало специальных программ, анализирующих результаты секвенирования ДНК, включая полногеномное – **CRISPR-GA** [Guell et al., 2014], **MAGeCK** и

MAGeCK-VISPR [Li et al., 2014; 2015], **GUIDE-seq** (**G**enome-wide, **U**nbiased **I**dentification of **D**SBs **E**nabled by **s**equencing) [Tsai et al., 2014; Zhu et al., 2017], **TIDE** (**T**racking of **I**ndels by **D**Ecomposition) [Brinkman et al., 2014], **AGEseq** [Xue, Tsai, 2015], **Protospacer Workbench** [MacPherson, Scherf, 2015], **CRISPResso** [Pinello et al., 2016], **BATCH-GE** [Boel et al., 2016], **CrisprVariants** [Lindsay et al., 2016], **Digenome-seq** [Park et al., 2017], **Cas-analyzer** [Park et al., 2017a], **Circle-seq** [Tsai et al., 2017], **DSDecodeM** [Xie X. et al., 2017], **caRpoools** [Winter et al., 2016], **CRISPRAnalyzeR** [Winter et al., 2017], **CRISPRcloud** [Jeong et al., 2017], **SAVE** (**S**creening **A**nalysis **V**isual **E**xplorer) [Jeong et al., 2017].

Несмотря на то, что эффективность и точность геномного редактирования с помощью CRISPR/Cas систем растет и будет дальше продолжаться улучшаться, необходимость в подобных анализах (имеется в виду в полногеномных данных) отредактированных геномов будет сохраняться и подобные программы будут появляться и совершенствоваться вместе с CRISPR/Cas технологией.

Заключение

Для дизайна гидРНК для CRISPR/Cas редактирования геномов написано уже довольно много компьютерных программ. Большинство из них представляют собой web-ресурсы, находящиеся в свободном доступе. Часть таких программных продуктов являются специализированными для определенных групп организмов. Некоторые программы рассчитаны на дизайн гидРНК и поиск нецелевых сайтов редактирования для ограниченного числа геномов, чаще всего человека и мыши. Не так много программ анализируют геномы растений. При этом есть ряд программ, которые позволяют искать нецелевые сайты редактирования практически во всех известных геномах.

Мы сознательно не стали здесь давать оценок различным программам дизайна гидРНК и указывать какие из них больше подходят для начинающих пользователей, а какие для продвинутых, поскольку это могло быть до некоторой степени субъективно, а во-вторых, стоящие перед разными экспериментаторами задачи могут сами диктовать тот или иной выбор конкретных инструментов. Поэтому главной своей целью в этой статье посчитали составить фактически перечень таких программ и очень кратко описать их возможности, с чем смеем надеяться справились.

В Приложении к данной статье в табличной форме дана чуть более подробная информация о некоторых on-line программах, позволяющая экспериментатору составить общее представление об

их возможностях. Все же основываясь на своем опыте, можем порекомендовать для дизайна гидРНК в рамках одного эксперимента использовать сразу несколько программ, сопоставляя выдаваемые результаты. При этом необходимо заметить, что написание программ дизайна гидРНК не совсем успевает за бурным развитием технологий геномного редактирования с помощью CRISPR/Cas систем. Так, например, при все более частом использовании химически синтезированных молекул гидРНК снимаются проблемы присутствия в их последовательностях неких участков, могущих служить преждевременными терминаторами транскрипции при ферментативном (*in vivo* или *in vitro*) синтезе гидРНК. Однако ряд программ *a priori* не рассматривают варианты гидРНК с такими последовательностями, сужая возможности редактирования. Представляется актуальным при дизайне гидРНК в самом начале работы с той или программой предлагать исследователю выбор вариантов редактирования геномов с использованием ферментативно или химически синтезированной РНК и соответственно подстраивать под них требования к последовательностям молекул гидРНК. Помимо программ дизайна гидРНК, уже создано несколько баз данных, хранящих подобранные гидРНК для некоторых видов организмов, включая растения. Сведения о таковых приведены в данной статье. Также кратко дана информация о ряде программ, используемых для оценки точности редактирования путем анализа данных секвенирования ДНК, в том числе полногеномного для выявления мест нецелевого редактирования.

Благодарности

Надеемся, что в данной статье нам удалось собрать сведения о практически всех компьютерных программах, рассчитанных на дизайн гидРНК для CRISPR/Cas редактирования геномов, однако не исключаем, что какие-то могли оказаться за пределами нашего внимания. При этом мы будем крайне признательны за информацию о таких пропущенных нами программах, а их разработчикам сразу приносим свои извинения.

Литература

1. Баймиев Ан.Х., Чемерис Д.А., Кирьянова О.Ю., Матниязов Р.Т., Валеев А.Ш., Баймиев Ал.Х., Губайдуллин И.М., Чемерис А.В. Биоинформатические ресурсы для *in silico* поиска CRISPR локусов в геномах прокариот // Биомика. 2017 Т.9. С.229-244.
2. Кулуев Б.Р., Герашенков Г.А., Рожнова Н.А., Баймиев Ан.Х., Вершинина З.Р., Князев А.В., Матниязов Р.Т., Гумерова Г.Р., Михайлова Е.В.,

- Никонов Ю.М., Чемерис Д.А., Баймиев Ал.Х., Чемерис А.В. CRISPR/Cas редактирование геномов растений // Биомика. 2017. Т.9. С.155-182.
3. Кулуев Б.Р., Баймиев Ан.Х., Чемерис Д.А., Матниязов Р.Т., Никонов Ю.М., Баймиев Ал.Х., Чемерис А.В. Применение CRISPR-локусов не для редактирования геномов // Биомика. 2017а. Т.9. С.271-283.
 4. Чемерис Д.А., Кирьянова О.Ю., Губайдуллин И.М., Чемерис А.В. Дизайн праймеров для полимеразной цепной реакции. (Краткий обзор компьютерных программ и баз данных) // Биомика. 2016. Т.8. С.215-238.
 5. Aach J., Mali P., Church G.M. CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes // bioRxiv preprint first posted online. 2014. doi: <https://doi.org/10.1101/005074>.
 6. Bae S., Kweon J., Kim H.S., Kim J.S. Microhomology-based choice of Cas9 nuclease target sites // Nat Methods. 2014. V. 11. P. 705–706.
 7. Bae S., Park J., Kim J.S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases // Bioinformatics. 2014. V. 30. P. 1473–1475.
 8. Blin K., Pedersen E., Weber T., Lee S.Y. CRISPy-web: An online resource to design sgRNAs for CRISPR applications // Synthetic and Systems Biotechnology. 2016. V.1. P.118-121.
 9. Brinkman E.K., Chen T., Amendola M., van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition // Nucleic Acids Res. 2014. V.42(22):e168.
 10. Brazelton V.A. Jr., Zarecor S., Wright D.A., Wang Y., Liu J., Chen K., Yang B., Lawrence-Dill C.J., A quick guide to CRISPR sgRNA design tools // GM Crops Food. 2015. 6. P. 266–276.
 11. Cao Q., Ma J., Chen C.H., Xu H., Chen Z., Li W., Liu X.S. CRISPR-FOCUS: A web server for designing focused CRISPR screening experiments // PLoS One. 2017. V.12(9):e0184281.
 12. Chari R., Mali P., Moosburner M., Church G.M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach // Nat Methods. 2015. V. 12. P. 823–826.
 13. Chari R., Yeo N.C., Chavez A., Church G.M. sgRNA Scorer 2.0: A species-independent model to predict CRISPR/Cas9 activity // ACS Synth Biol. 2017. V. 19. P. 902–904.
 14. Chu V.T., Graf R., Wirtz T., Weber T., Favret J., Li X., Petsch K., Tran N.T., Sieweke M.H., Berek C., Kühn R., Rajewsky K. Efficient CRISPR-mediated mutagenesis in primary immune cells using CrispRGold and a C57BL/6 Cas9 transgenic mouse line // Proc. Natl. Acad. Sci. USA. 2016. V.113. P.12514-12519.
 15. Chuai G.H., Wang Q.L., Liu Q. *In silico* meets *in vivo*: Towards computational CRISPR-based sgRNA design // Trends Biotechnol. 2017. V.35. P.12-21.
 16. Cradick T.J., Qiu P., Lee C.M., Fine E.J., Bao G. COSMID: A Web-based tool for identifying and validating CRISPR/Cas off-target sites // Molecular Therapy—Nucleic Acids. 2014. V.3. e214.
 17. 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. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation // Nat. Biotechnol. 2014. V.32. P.1262-1267.
 18. Garst A.D., Bassalo M.C., Pines G., Lynch S.A., Halweg-Edwards A.L., Liu R., Liang L., Wang Z., Zeitoun R., Alexander W.G., Gill R.T. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering // Nat. Biotechnol. 2017. V.35. P.48-55.
 19. Graham D.B., Root D.E. Resources for the design of CRISPR gene editing experiments // Genome Biol. 2015. V. 16:260.
 20. Guell M., Yang L., Church G.M. Genome editing assessment using CRISPR genome analyzer (CRISPR-GA) // Bioinformatics. 2014. V. 30. P. 2968–2970.
 21. Guo D., Li X., Zhu P., Feng Y., Yang J., Zheng Z., Yang W., Zhang E., Yu Y., Zhou S., Wang H. Online High-throughput Mutagenesis Designer Using Scoring Matrix of Sequence-specific Endonucleases // J. Integrative Bioinformatics. 2015. V. 12. doi: 10.2390/biecoll-jib-2015-283.
 22. Haeussler M., Schönig K., Eckert H., Eschstruth A., Mianné J., Renaud J.B., Schneider-Maunoury S., Shkumatava A., Teboul L., Kent J., Joly J.S., Concordet J.P. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR // Genome Biol. 2016. V. 17:148.
 23. Heigwer F., Kerr G., Boutros M. E-CRISP: fast CRISPR target site identification // Nat Methods. 2014. V. 11. P. 122–123.
 24. Heigwer F., Zhan T., Breinig M., Winter J., Brügemann D., Leible S., Boutros M. CRISPR library designer (CLD): software for multispecies design of single guide RNA libraries // Genome Biol. 2016. V.24. 17:55.
 25. Hodgkins A., Farne A., Perera S., Grego T., Parry-Smith D.J., Skarnes W.C., Iyer V. WGE: a CRISPR database for genome engineering // Bioinformatics. 2015. V. 31:3078–3080.
 26. Hough S.H., Kancleris K., Brody L., Humphries-Kirilov N., Wolanski J., Dunaway K., Ajetunmobi A., Dillard V. Guide Picker is a comprehensive design tool for visualizing and

- selecting guides for CRISPR experiments // BMC Bioinformatics. 2017. V. 18:167.
27. Hsu P.D., Scott D.A., Weinstein J.A., Ran F.A., Konermann S., Agarwala V., Li Y., Fine E.J., Wu X., Shalem O., Cradick T.J., Marraffini L.A., Bao G., Zhang F. DNA targeting specificity of RNA-guided Cas9 nucleases // Nat. Biotechnol. 2013. V.31. P.827-832.
 28. Jaskula-Ranga V., Zack D.J. grID: A CRISPR-Cas9 guide RNA database and resource for genome-editing // bioRxiv 2017. 097352; doi.org/10.1101/097352
 29. Jeong H.H., Kim S.Y., Rousseaux M.W., Zoghbi H.Y., Liu Z. CRISPRcloud: A secure cloud-based pipeline for CRISPR pooled screen deconvolution // Bioinformatics. 2017.
 30. Jeong H.H., Kim S.Y., Rousseaux M.W., Zoghbi H.Y., Liu Z. SAVE: A secure cloud-based pipeline for CRISPR pooled screen deconvolution // bioRxiv. 2017. doi:10.1101/110262.
 31. Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity // Science. 2012. V. 337. P. 816–821.
 32. Kaur K., Tandon H., Gupta A.K., Kumar M. CrisprGE: a central hub of CRISPR/Cas-based genome editing // Database (Oxford). 2015. V. 27.
 33. Kelley M.L., Strezoska Z., He K., Vermeulen A., Smith A. Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing // J Biotechnol. 2016. V. 233. P. 74–83.
 34. Kuan P.F., Powers S., He S., Li K., Zhao X., Huang B. A systematic evaluation of nucleotide properties for CRISPR sgRNA design // BMC Bioinformatics. 2017. V.18:297.
 35. Labun K., Montague T.G., Gagnon J.A., Thyme S.B., Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering // Nucleic Acids Res. 2016. V. 44. P. W272-W276.
 36. Lei Y., Lu L., Liu H.Y., Li S., Xing F., Chen L.L. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants // Mol Plant. 2014. V. 7. P. 1494–1496.
 37. Lindsay H., Burger A., Biyong B., Felker A., Hess C., Zaugg J., Chiavacci E., Anders C., Jinek M., Mosimann C., Robinson M.D. CrispRVariants charts the mutation spectrum of genome engineering experiments // Nat Biotechnol. 2016. V. 34. P. 701–702.
 38. Liu X., Homma A., Sayadi J., Yang S., Ohashi J., Takumi T. Sequence features associated with the cleavage efficiency of CRISPR/Cas9 system // Sci. Rep. 2016. V.6:19675.
 39. Liu H., Wei Z., Dominguez A., Li Y., Wang X., Qi L.S. CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation // Bioinformatics. 2015a. V.31. P. 3676–3678.
 40. Liu H., Ding Y., Zhou Y., Jin W., Xie K., Chen LL. CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants // Mol Plant. 2017. V. 10. P. 530–532.
 41. Ma J., Köster J., Qin Q., Hu S., Li W., Chen C., Cao Q., Wang J., Mei S., Liu Q., Xu H., Liu X.S. CRISPR-DO for genome-wide CRISPR design and optimization // Bioinformatics. 2016. V.32. P.3336-3338.
 42. Ma M., Ye A.Y., Zheng W., Kong L. A guide RNA sequence design platform for the CRISPR/Cas9 system for model organism genomes // Biomed. Res. Int. 2013;2013:270805.
 43. Michno J.M., Wang X., Liu J., Curtin S.J., Kono T.J., Stupar R.M. CRISPR/Cas mutagenesis of soybean and *Medicago truncatula* using a new web-tool and a modified Cas9 enzyme // GM Crops Food. 2015. V.6. P.243-252.
 44. Montague T.G., Cruz J.M., Gagnon J.A., Church G.M., Valen E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing // Nucleic Acids Res. 2014. V. 42. P. W401–407.
 45. Moreno-Mateos M.A., Vejnar C.E., Beaudoin J.D., Fernandez J.P., Mis E.K., Khokha M.K., Giraldez A.J. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo // Nat. Methods. 2015. V.12. P.982-988.
 46. Naito Y., Hino K., Bono H., Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites // Bioinformatics. 2015. V. 31. P. 1120–1123.
 47. Nakajima I., Ban Y., Azuma A., Onoue N., Moriguchi T., Yamamoto T., Toki S., Endo M. CRISPR/Cas9-mediated targeted mutagenesis in grape // PLoS One. 2017. V. 18:e0177966.
 48. Neff K.L., Argue D.P., Ma A.C., Lee H.B., Clark K.J., Ekker S.C. Mojo Hand, a TALEN design tool for genome editing applications // BMC Bioinformatics. 2013. V.14: 1.
 49. O'Brien A., Bailey T.L. GT-Scan: Identifying unique genomic targets // Bioinformatics. 2014. V. 30. P. 2673–2675.
 50. Oliveros J.C., Franch M., Tabas-Madrid D., San-León D., Montoliu L., Cubas P., Pazos F. Breaking-Cas-interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes // Nucleic Acids Res. 2016. V.44. P.W267-271.
 51. Park J., Bae S., Kim J.S. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites // Bioinformatics. 2015. V. 31. P. 4014–4016.
 52. Park J., Kim J.S., Bae S. Cas-Database: web-based genome-wide guide RNA library design for gene knockout screens using CRISPR-Cas9 // Bioinformatics. 2016. V. 32. V. 2017–2023.

53. Park J., Lim K., Kim J.S., Bae S. Cas-analyzer: an online tool for assessing genome editing results using NGS data // *Bioinformatics*. 2017. V. 33. P. 286–288.
54. Peng D., Tarleton R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens // *Microb. Genom.* 2015. V.30. e000033.
55. Peng R., Lin G., Li J. Potential pitfalls of CRISPR/Cas9-mediated genome editing // *FEBS J.* 2016. V.283. P.1218-1231.
56. Perez A.R., Pritykin Y., Vidigal J.A., Chhangawala S., Zamparo L., Leslie C.S., Ventura A. GuideScan software for improved single and paired CRISPR guide RNA design // *Nat Biotechnol.* 2017. V. 35. P. 347–349.
57. Periwal V. A comprehensive overview of computational resources to aid in precision genome editing with engineered nucleases // *Brief Bioinform.* 2017. V.18. P.698-711.
58. Pinello L., Canver M.C., Hoban M.D., Orkin S.H., Kohn D.B., Bauer D.E., Yuan G.C. Analyzing CRISPR genome-editing experiments with CRISPResso // *Nat Biotechnol.* 2016. V. 34. P. 695–697.
59. Pliatsika V., Rigoutsos I. "Off-Spotter": very fast and exhaustive enumeration of genomic lookalikes for designing CRISPR/Cas guide RNAs // *Biol. Direct.* 2015. V.10:4.
60. Prykhozhij S.V., Rajan V., Gaston D., Berman J.N. CRISPR Multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences // *PLoS One.* 2015. V. 10:e0119372.
61. Prykhozhij S.V., Rajan V., Gaston D., Berman J.N. Correction: CRISPR MultiTargeter: A Web Tool to Find Common and Unique CRISPR Single Guide RNA Targets in a Set of Similar Sequences // *PLoS One.* 2015. V.10:e0138634.
62. Pulido-Quetglas C., Aparicio-Prat E., Arnan C., Polidori T., Hermoso T., Palumbo E., Ponomarenko J., Guigo R., Johnson R. Scalable Design of Paired CRISPR Guide RNAs for Genomic Deletion .. *PLoS Comput Biol.* 2017. V.13:e1005341.
63. Ran F.A., Hsu P.D., Wright J., Agarwala V., Scott D.A., Zhang F. Genome engineering using the CRISPR-Cas9 system // *Nat. Protoc.* 2013. V.8. P.2281-308.
64. Rauscher B., Heigwer F., Breinig M., Winter J., Boutros M. GenomeCRISPR – a database for high-throughput CRISPR/Cas9 screens // *Nucleic Acids Res.* 2017. V. 45. P. D679–D686.
65. Ronda C., Pedersen L.E., Hansen H.G., Kallehauge T.B., Betenbaugh M.J., Nielsen A.T., Kildegaard H.F. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool // *Biotechnol. Bioeng.,* 2014. V.111. P.1604–1616.
66. Sander, J.D., Zaback, P.Z., Joung, J.K., Voytas, D.F., Dobbs, D. (2007) Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Research*, 35, W599-605; Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K., Dobbs, D. (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Research*, 38:W462-468
67. Rodríguez-López M., Cotobal C., Fernández-Sánchez O., Bravo N.B., Oktriani R., Abendroth H., Uka D., Hoti M., Wang J., Zaratiegui M., Bähler J. A CRISPR/Cas9-based method and primer design tool for seamless genome editing in fission yeast // *Wellcome Open Research.* 2016. V.1:19. [version 2; referees: 2 approved]
68. Stemmer M., Thumberger T., Del Sol Keyer M., Wittbrodt J., Mateo J.L. CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool // *PLoS One.* 2015. V. 24:e0124633. Correction: *PLoS One.* 2017. V.12:e0176619.
69. Tsai S.Q., Zheng Z., Nguyen N.T., Liebers M., Topkar V.V., Thapar V., Wyvekens N., Khayter C., Iafrate A.J., Le L.P., Aryee M.J. Joung J.K. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases // *Nat Biotechnol.* 2015. V. 33. P. 187–197.
70. Tsai S.Q., Nguyen N.T., Malagon-Lopez J., Topkar V.V., Aryee M.J., Joung J.K. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets // *Nat Methods.* 2017. V. 14. P. 607–614.
71. Tycko J., Myer V.E., Hsu P.D. Methods for optimizing CRISPR-Cas9 genome editing specificity // *Mol. Cell.* 2016. V.63. P.355-370.
72. Varshney G.K., Zhang S., Pei W., Adomako-Ankomah A., Fohntung J., Schaffer K., Carrington B., Maskeri A., Slevin C., Wolfsberg T., Ledin J., Sood R., Burgess S.M. CRISPRz: a database of zebrafish validated sgRNAs // *Nucleic Acids Res.* 2016. V. 44. P. D822-D826.
73. Vazquez-Vilar M., Bernabé-Orts J.M., Fernandez-Del-Carmen A., Ziarsolo P., Blanca J., Granell A., Orzaez D. A modular toolbox for gRNA-Cas9 genome engineering in plants based on the GoldenBraid standard // *Plant Methods.* 2016. V.12:10.
74. Wang Y, Liu X, Ren C, Zhong GY, Yang L, Li S, Liang Z. Identification of genomic sites for CRISPR/Cas9-based genome editing in the *Vitis vinifera* genome // *BMC Plant Biol.* 2016 Apr 21;16:96. doi: 10.1186/s12870-016-0787-3.
75. Winter J., Breinig M., Heigwer F., Brügemann D., Leible S., Pelz O., Zhan T., Boutros M. caRools: an R package for exploratory data

- analysis and documentation of pooled CRISPR/Cas9 screens // *Bioinformatics*. 2016. V.32. P.632-634.
76. Winter J., Schwering M., Pelz O., Rauscher B., Zhan T., Heigwer F., Boutros M. CRISPRAnalyzer: Interactive analysis, annotation and documentation of pooled CRISPR screens // *bioRxiv. The Preprint Server for Biology*. 2017.
 77. Wong N., Liu W., Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system // *Genome Biol*. 2015. V. 16:218.
 78. Xiao A., Wu Y., Yang Z., Hu Y., Wang W., Zhang Y., Kong L., Gao G., Zhu Z., Lin S., Zhang B. EENdb: a database and knowledge base of ZFNs and TALENs for endonuclease engineering // *Nucleic Acids Res*. 2013. V41. (Database issue). D415-422.
 79. Xiao A., Cheng Z., Kong L., Zhu Z., Lin S., Gao G., Zhang B. CasOT: a genome-wide Cas9/gRNA off-target searching tool // *Bioinformatics*. 2014. V. 30. P. 1180–1182.
 80. Xie S., Shen B., Zhang C., Huang X., Zhang Y. sgRNAs9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites // *PLoS One*. 2014 V. 23:e100448.
 81. Xie X., Ma X., Zhu Q., Zeng D., Li G., Liu Y.-G. CRISPR-GE: A convenient software toolkit for CRISPR-based genome editing // *Mol. Plant*. 2017.
 82. Xu H., Xiao T., Chen C.H., Li W., Meyer C.A., Wu Q., Wu D., Cong L., Zhang F., Liu J.S., Brown M., Liu X.S. Sequence determinants of improved CRISPR sgRNA design // *Genome Res*. 2015. V.25. P.1147-1157.
 83. Xue L.J., Tsai C.J. AGEseq: analysis of genome editing by sequencing // *Mol Plant*. 2015. V.8. P.1428–1430.
 84. Yan M., Zhou S.R., Xue H.W. CRISPR Primer Designer: Design primers for knockout and chromosome imaging CRISPR-Cas system // *J. Integr. Plant Biol*. 2015. V.57. P.613-617.
 85. Yan J., Chuai G., Zhou C., Zhu C., Yang J., Zhang C., Gu F., Xu H., Wei J., Liu Q. Benchmarking CRISPR on-target sgRNA design // *Brief Bioinform*. 2017.
 86. Yennmalli R.M., Kalra S., Srivastava P.A., Garlapati V.K. Computational tools and resources for CRISPR/Cas 9 genome editing method // *MOJ Proteomics Bioinform*. 2017. V.5:00164.
 87. Zetsche B., Gootenberg J.S., Abudayyeh O.O., Slaymaker I.M., Makarova K.S., Essletzbichler P., Volz S.E., Joung J., van der Oost J., Regev A., Koonin E.V., Zhang F. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system // *Cell*. 2015. V. 163. P. 759–771.
 88. Zhu L.J., Holmes B.R., Aronin N., Brodsky M.H. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems // *PLoS One*. 2014. V. 9:e108424.
 89. Zhu L.J., Lawrence M., Gupta A., Pages H., Kucukural A., Garber M., Wolfe S.A. GUIDEseq: a bioconductor package to analyze GUIDE-Seq datasets for CRISPR-Cas nucleases // *BMC Genomics*. 2017. V. 18:379.
 90. Zhu H., Misel L., Graham M., Robinson M.L., Liang C. CT-Finder: A web service for CRISPR optimal target prediction and Visualization // *Sci Rep*. 2016. V. 23:25516.
 91. Zhu H., Richmond E., Liang C. CRISPR-RT: A web service for designing CRISPR-C2c2 crRNA with improved target specificity // *bioRxiv preprint first posted online*. 2017. doi.org/10.1101/099895.

BIOINFORMATIC RESOURCES FOR CRISPR/Cas GENOME EDITING

Chemersis D.A.¹, Kiryanova O.Yu.^{2,3}, Gerashchenkov G.A.¹, Kuluev B.R.¹, Rozhnova N.A.¹, Matniyazov R.T.¹, Baymiev An.Kh.¹, Baymiev Al.Kh.¹, Gubaydullin I.M.^{2,3}, Chemersis A.V.¹

¹Institute of Biochemistry and Genetics, Ufa Scientific Center of RAS, Ufa, Russia, chemeris@anrb.ru

²Institute of Petrochemistry and Catalysis of Russian Academy of Sciences, Ufa, Russia

³Ufa State Petroleum Technological University, Ufa, Russia

Resume

The features of numerous programs for gRNA designing with their actual web pages and the stages of work with them are described. For a number of programs are described more detailed, that is reflected in a special Appendix. Programs for mutations search after CRISPR/Cas editing of genomes of various organisms are listed. Databases containing previously selected gRNA for a number of genomes of plants and animals, including humans are considered. The analysis of almost all the existing literature on computer designing of gRNA and relevant Internet resources are shown.

Keywords: CRISPR, Cas, PAM, gRNA, spacer, protospacer, On-target site, Off-target site, database, software, web-resource

Основные характеристики некоторых on-line программ для дизайна гидРНК для CRISPR/Cas редактирования геномов

	<i>i</i>	<i>ii</i>	<i>iii</i>	<i>iv</i>	<i>v</i>	<i>vi</i>	<i>vii</i>	<i>viii</i>	<i>ix</i>	<i>x</i>	<i>xi</i>	<i>xii</i>	<i>xiii</i>	<i>xiv</i>	<i>xv</i>	<i>xvi</i>	<i>xvii</i>	<i>xviii</i>	<i>xix</i>	<i>xx</i>	<i>xxi</i>	
1	✓	✓				✓	✓		67/>100	✓			15-25				✓	✓			✓	✓
2	✓						✓		13/42	✓	✓		ф ¹	✓							✓	
3	✓		✓		✓	✓			45/0	✓	✓	✓	15-22									
4	✓	✓	✓	✓	✓	✓			9/90			✓	19-20	✓	✓							
5	✓		✓		✓				>50/>80	✓	✓	✓	ф								✓	
6	✓	✓	✓				✓		15/41				15-23	✓	✓						✓	✓
7	✓		✓	✓	✓		✓		3/9	✓	✓	✓	в ²		✓							
8	✓				✓	✓			6/0	✓	✓		21,23									
9	✓	✓	✓	✓	✓	✓			0/14	✓			14-24									
10	✓	✓	✓						44/много ³	✓			ф, 18-25	✓							✓	
11	✓	✓	✓		✓	✓			40/180	✓			ф				✓		✓/✓	✓	✓	✓
12	✓								0/5		✓	✓	19, 20									
13	✓			✓					3/11	✓			в								✓	✓
14	✓				✓				8/0		✓	✓	20								✓	✓
15	✓			✓	✓				0/5	✓			20	✓								
16	✓	✓	✓		✓		✓ ⁴		8/0	✓	✓		ф, в			✓			✓/✓	✓		
17	✓						✓		13/0	✓			20		✓				✓/-	✓	✓	✓
18	✓				✓	✓			1/16	✓			20								✓	
19	✓		✓	✓	✓	✓			44/85	✓			18-30		✓						✓	✓
20	✓				✓	✓	✓	✓	0/2	✓	✓		20									
	<i>i</i>	<i>ii</i>	<i>iii</i>	<i>iv</i>	<i>v</i>	<i>vi</i>	<i>vii</i>	<i>viii</i>	<i>ix</i>	<i>x</i>	<i>xi</i>	<i>xii</i>	<i>xiii</i>	<i>xiv</i>	<i>xv</i>	<i>xvi</i>	<i>xvii</i>	<i>xviii</i>	<i>xix</i>	<i>xx</i>	<i>xxi</i>	

i – Cas9; *ii* – SpF1; *iii* – вариации PAM последовательностей, в том числе вводимые с клавиатуры; *iv* – никазы/парные гидРНК; *v* – ранжирование оптимальных гидРНК; *vi* – нахождение нецелевых сайтов; *vii* – анализ ранее подобранных гидРНК (валидация); *viii* – выравнивание множественных последовательностей; *ix* – доступные геномы для анализа (царство растений/прочие); *x* – ввод последовательностей ДНК через буфер обмена или в виде файла; *xi* – ввод конкретных генов с помощью их идентификатора (имя, Accession number); *xii* – ввод последовательностей ДНК с помощью геномных координат; *xiii* – изменяемая длина спейсерной последовательности; *xiv* – допускаемые неспаривания; *xv* – выбор 5'-конца гидРНК в виде G или GG; *xvi* – анализ фланкирующих последовательностей; *xvii* – микромология; *xviii* – ДНК/РНК-выпегливание (инделы) в спейсерах; *xix* – подбор праймеров/олигонуклеотидов для клонирования / ПЦР детекции; *xx* – обучающие демоверсии/видео; *xxi* – возможность установки на ПК

¹ ф – фиксированные длины спейсеров для конкретных нуклеаз разных типов
² в – варибельная длина спейсера (нет ограничений), по умолчанию обычно 20
³ все доступные геномы в последней версии Ensembl (<http://ensemblgenomes.org>), среди которых около 700 эукариотических
⁴ off-target участки позволяется искать в 27 растительных геномах (23 вида) и трех прочих (человек, мышь, рыба)

**Перечень проанализированных в Приложении компьютерных программ дизайна гидПНК
(с комментариями)**

- 1) **CRISPR RGEN Tools (Cas-OFFinder, Microhomology-Predictor, Cas-Designer** (<http://www.rgenome.net>) - допускается сделать запрос для введения генома дополнительного организма; допускается до трех неспариваний в спейсерной области.
- 2) **E-CRISP** (<http://www.e-crisp.org/E-CRISP/>)
- 3) **CRISPR-P 2.0** (<http://cbi.hzau.edu.cn/crispr/>) – можно выбрать U6 или U3 промоторы для транскрипции гидПНК; приведена последовательность константной части гидПНК.
- 4) **CHOPCHOP v2** (<http://chopchop.cbu.uib.no>) – ведется анализ фланкирующих последовательностей; допускается максимально 3 неспаривания при обнаружении off target мишеней, в том числе в затравочной части; возможно ранжирование гидПНК на основе отличающихся способов выявления оценок эффективности редактирования.
- 5) **CRISPRdirect** (<http://crispr.dbcls.jp>)
- 6) **CCTop** (<http://crispr.cos.uni-heidelberg.de>) - можно выбрать T7 или U3 промоторы для транскрипции гидПНК; длина затравочной последовательности может варьировать от 5 до 20 нуклеотидов с максимальным количеством в виде двух неспариваний.
- 7) **CRISPR MultiTargeter** (<http://www.multicrispr.net>) – допускается 5'- или 3'-ориентация PAM; можно указать допустимы или нет неспаривания в затравочной части (8 нуклеотидов).
- 8) **CGAT** (<http://cbc.gdcb.iastate.edu/cgat/>) - можно выбрать конкретную хромосому или ее плечо; можно варьировать GC-состав гидПНК (по умолчанию от 40 до 60%).
- 9) **sgRNA Scorer 2.0** (<https://crispr.med.harvard.edu/sgRNAScorer/>) – допускается 5'- или 3'-ориентация PAM; допускается введение нуклеотидной последовательности длиной от 30 до 1 млн. нуклеотидов.
- 10) **Breaking Cas** (<http://bioinfogp.cnb.csic.es/tools/breakingcas/>) – в спейсерной области гидПНК допускается до 4 неспариваний с указанием их «веса».
- 11) **CRISPOR v4.3** (<http://crispor.tefor.net>) - допускается сделать запрос для введения генома дополнительного организма; ранжирование гидПНК с помощью разных подходов.
- 12) **CRISPR-DO** (<http://cistrome.org/crispr/>) - можно задавать параметры ранжирования гидПНК.
- 13) **CT-Finder** (<http://bioinfolab.miamioh.edu/ct-finder/>) – отдельно выделяется затравочный участок из 8 нуклеотидов, который можно варьировать, в нем допускается одно неспаривание, в целом до 5 неспариваний. Для обнаружения off-target мишеней допускается два неспаривания в затравочном участке.
- 14) **CRISPR-Plant** (<http://genome.arizona.edu/crispr/>)
- 15) **CRISPETA** (<http://crispeta.crg.eu>) – допускается до 5 неспариваний в спейсерной области гидПНК.
- 16) **CRISPR-GE (seqDownload, targetDesign, offTarget, primerDesign-V, primerDesign-A)** (<http://skl.scau.edu.cn>) - можно вводить праймеры, ограничивающие интересующий экспериментатора участок генома, который будет вовлечен в анализ; допускается анализ фланкирующих участков до 100 тыс. п.н. (по умолчанию – 2000 и 500 для 5'- и 3'-сторон; допускается 5' или 3'-ориентация PAM.
- 17) **PhytoCRISP-Ex** (<http://www.phytocrispex.biologie.ens.fr/CRISP-Ex/>)
- 18) **Optimized CRISPR Design** (<http://crispr.mit.edu:8079>) - допускается до 4 неспариваний в спейсерной области гидПНК.
- 19) **CasOT** (<http://casot.cbi.pku.edu.cn>) - допускается до 6 неспариваний в спейсерной области гидПНК.
- 20) **sgRNA Designer (CRISPRko)** (<http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>)

Ограничившись в данном Приложении анализом этих 20 программ дизайна гидПНК, хотим заметить, что эта цифра была продиктована отчасти желанием уместить всю информацию на одной странице формата А4 в альбомном варианте. Произведенный выбор конкретных 20 программ дизайна гидПНК отнюдь не означает, что все упомянутые в этой статье такие программные продукты нами как-то специально ранжированы и выбрана лучшая двадцатка. В данном Приложении сознательно представлены программы, решающие

несколько отличающиеся задачи и имеющие разные возможности, чтобы показать их многообразие. Хотя некоторый приоритет был отдан программам, в которых в анализ вовлекаются геномы различных растений. Подобной таблицы из 21 информационной колонки нет ни в одной обзорной статье по этой теме, где такие сравнения ведутся по четырем/шести/семи/десяти колонкам, включая название программы, ее URL и ссылки на публикацию [Brazelton, Jr. et al., 2015; Graham, Root, 2015; Zhu, 2015; Ceasar et al., 2016; Kanchiswamy,

2016; Chuai et al., 2016; Peng et al., 2016; Туско et al., 2016; Yennmalli et al., 2017]. Лишь в одной очень краткой (двустраничной) статье [MacPherson, Scherf, 2015] приведены сведения для шести программ по 34 пунктам, но заметная часть из них носит неосновной характер, отмечая, например, скорость работы программы (самого поиска и дизайна гидРНК), операционные системы, включая их битность и легкость инсталляции (для программ, которые требуют установки на компьютер пользователя), наличие web-сайта и др. При этом наши колонки **ix** и **xix** каждая несут к тому же фактически двойную информацию о количествах вовлеченных в анализ геномов видов растений и прочих организмов и подборе праймеров для клонирования гидРНК и детекции с помощью ПЦР результатов редактирования геномов соответственно. Здесь надо сказать, что некоторые программы дизайна гидРНК, у которых подобная встроенная опция отсутствует, отсылают пользователей на другие компьютерные программы, предназначенные для дизайна праймеров для проведения ПЦР, большинство которых находятся в свободном доступе и таковых имеется около полутора сотен, при этом их анализ был проведен нами ранее [Чемерис и др., 2016]. Что касается геномов и видов организмов (колонка **ix**), то во избежание недоразумений необходимо отметить, что имеющиеся в ряде случаев расхождения в указанных нами количествах с таковыми, фигурирующими в программах дизайна гидРНК, вызваны тем, что мы ведем речь о числе видов организмов, тогда как в программных продуктах часто приводятся количества разных секвенированных геномов, в том числе принадлежащих одному виду. Безусловно, число геномов, которые могут быть подвергнуты анализу – важная информация, но мы сочли, что число видов организмов представляет не меньший, если не больший интерес, а делать еще две дополнительные колонки с количествами доступных для анализа геномов растений и прочих видов посчитали излишним.

В целом многие программы дизайна гидРНК вместе с сопутствующими атрибутами и различными опциями, настолько разнообразны, что число информационных колонок, требующихся для описания их всех, может превысить и полусотню. Так, в наших комментариях к перечисленным программам содержатся не менее полутора десятков пунктов, которые могли бы стать самостоятельными колонками, но мы решили поступить иначе, намеренно уменьшив число столбцов в таблице. Предлагаем тем, кому это интересно самостоятельно изучить приведенные во всей данной статье (а не только в Приложении) компьютерные программы дизайна гидРНК, обращая внимание на именно те моменты, которые будут в первую очередь важны для конкретных экспериментов, поскольку конечным пользователям должно быть виднее, что им нужнее. При этом надеемся, что сделанная нами подборка и первичный краткий анализ таких программ в этом могут им помочь и окажутся полезными.

Здесь еще надо сказать, что при установке некоторых программ дизайна гидРНК на компьютер пользователя у него часто появляется возможность менять те или иные параметры поиска под свои задачи, поскольку в целом ряде программных продуктов, работающих on-line, они установлены по умолчанию и даже не видны экспериментатору. И это прямо отмечают некоторые разработчики. Другой важной причиной поиска гидРНК в режиме off-line может стать желание экспериментаторов скрыть от хозяев web-сайта, где размещена та или иная программа, над чем они работают, поскольку по нуклеотидной последовательности из 20 и даже меньшего числа нуклеотидов можно легко понять объекты исследований конкретной лаборатории и догадаться о стоящих перед ними задачах, что может оказаться нежелательным. К тому же перед началом работы исследователю на некоторых web-ресурсах предлагается указать свой действующий электронный адрес, на который высылаются результаты анализа.