Front. Biol. 2015, 10(4): 289–296
DOI 10.1007/s11515-015-1366-y

MINI-REVIEW

Overview of guide RNA design tools for CRISPR-Cas9 genome editing technology

Lihua Julie Zhu (✉)

Department of Molecular, Cell and Cancer Biology, Program in Bioinformatics and Integrated Biology, Program in Molecular Medicine, University of Massachusetts Medical School,Worcester, MA 01605, USA

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

Abstract **CRISPR-Cas (Clustered, Regularly Interspaced, Short Palindromic Repeats – CRISPR-associated (Cas)) RNA guided endonuclease has emerged as the most effective and widely used genome editing technology, which has become the most exciting and rapidly advancing research field. Efficient genome editing by the CRISPR-Cas9 system has been demonstrated in many species, and several laboratories have established CRISPR-Cas9 as a screening tool for systematic genetic analysis, similar to shRNA screening. At least three companies have been founded to leverage this technology for therapeutic uses. To facilitate the implementation of this technology, many software tools have been developed to identify guide RNAs that effectively target a desired genomic region. Here, I provide an overview of the technology, focusing on guide RNA design principles, available software tools and their strengths and weaknesses.**

Keywords CRISPR-Cas9, genome editing, gRNA design, off-target analysis, gRNA efficacy

Introduction

CRISPR-Cas (Clustered, Regularly Interspaced, Short Palindromic Repeats – CRISPR-associated (Cas)) RNA guided endonuclease is the most recent development in genome editing technology (Esvelt et al., 2013, Ran et al., 2013b). CRISPR-Cas editing technology borrows a strategy from the adaptive mechanisms for bacteria and archaea to fight invading viruses and plasmids (Koonin and Makarova, 2009, 2013; Horvath and Barrangou, 2010; Jinek et al., 2012; Sampson et al., 2013; Doudna and Charpentier, 2014). In brief, CRISPR stores DNA sequences from invaded viruses or plasmid in a transcriptional array and when the same type of virus invades again, the system will recognize it using the transcribed RNA sequences and direct the Cas nuclease to make a double stranded break (DSB). One type of nuclease known as Cas9 from the bacterium *Streptococcus pyogenes* (*S. pyogenes*) cuts DNA at the exact location dictated by a single guide RNA (sgRNA) that can be programmed to target a genomic DNA sequence for editing (Hsu et al., 2013). Once Cas9 makes a DSB, random insertion or deletion can be generated via an error-prone non-homologous end-joining (NHEJ) pathway or desired modification can be introduced by homology-directed repair (HDR) pathway templated from exogenous DNA (Wyman and Kanaar, 2006). Recently, efficient genome editing by the CRISPR-Cas9 system has been demonstrated in multiple organisms, including human, mouse, rat, zebrafish, *Drosophila* and *C. elegans* (Cong et al., 2013; Friedland et al., 2013; Gratz et al., 2013; Hou et al., 2013; Hwang et al., 2013; Jinek et al., 2013; Li et al., 2013; Mali et al., 2013b; Yang et al., 2013). In contrast to previous genome-editing techniques, such as zinc-finger nucleases (ZFNs) (Meng et al., 2008; Gupta et al., 2011; Chu et al., 2012; Enuameh et al., 2013) and transcription activator-like effector nucleases (TALENs) (Joung and Sander, 2013), the target specificity of CRISPR-Cas9 is primarily dictated by a Watson-Crick pairing of a 20-base sequence at the 5′-end of the sgRNA with the target site instead of protein-DNA recognition, providing a much easier system to target multiple genes simultaneously. It has been shown that compared with ZFNs and TALENs, CRISPR-Cas-mediated gene targeting has similar or greater efficiency in human cells, zebrafish and metazoan *Nematostella vectensis* (Ding et al., 2013; Ikmi et al., 2014; Smith et al., 2014). Recently, several laboratories have established CRISPR-Cas9 as a screening tool for systematic genetic analysis in mammalian cells, analogous to shRNA screens (Shalem et

Received June 21, 2015; accepted July 8, 2015
Correspondence: Lihua Julie Zhu
E-mail: Julie.zhu@umassmed.edu

al., 2014) (Chen et al., 2015; Koike-Yusa et al., 2014; Wang et al., 2014). At least three companies have been founded to take advantage of this technology for therapeutic uses to correct genetic disorders and battle invading pathogens, named CRISPR Therapeutics, Intellia Therapeutics and Editas Medicine.

Figure 1 depicts the two components of the CRISPR-Cas9 system from *S. pyogenes* and their recognition sites. The first component is the Cas9 nuclease from the bacteria, depicted as purple oval and the second component is a single guide RNA or sgRNA, which is derived from a fusion of the tracrRNA and crRNA found in bacteria (Jinek et al., 2012). In the engineered form, the sgRNA has two parts. One is the constant region, colored in peach, which forms several stem-loop structures serving as scaffolding for Cas9 binding. The second is a 20 base variable region (referred to as gRNA hereafter), colored in green, which can be altered to target different sequences. The target site, recognized by this complex, is composed of two parts. One part of the target site, colored in blue, is complementary to the gRNA. The other part of the target site, colored in red, is called protospacer adjacent motif (PAM) and bound by Cas9. The PAM is a very short region (NGG in Sp) adjacent to the 20 bases that are recognized by the gRNA. In summary, for the most commonly engineered CRISPR-Cas9 system derived from Sp, Cas9 nuclease binds to the NGG PAM sequence, then if the 20 base gRNA base pairs with the target DNA sequence it will make a DSB. Once Cas9 makes a DSB, DNA undergoes repairs using NHEJ or HDR if a donor template is provided, leading to random indels or desired modification of the targeted gene.

Overview of gRNA design tools

Finding target sites is generally quite easy by just scanning for the PAM sequence e.g., NGG for the CRISPR-Cas9 system from *S. pyogenes*. The challenge is to be able to design a predictive algorithm to identify target sites that can be cleaved efficiently (aka efficacy) and for which the cognate gRNAs have little or no cleavage at other genomic locations (aka specificity). Therefore, ideal gRNAs would have high efficacy with great specificity. To help researchers to select the best gRNAs for input sequences, it is essential to identify gRNAs and their potential off-targets, and accurately predict their relative cleavage rates. To facilitate gRNA design, many computational tools have been developed (Hsu et al., 2013; Ma et al., 2013; Doench et al., 2014; Heigwer et al., 2014; Xiao et al., 2014; Zhu et al., 2014; Prykhozhij et al., 2015), and a few representative ones are summarized in Table 1.

The Root laboratory assessed the rules governing the gRNA efficacy by creating a pool of 1841 sgRNAs, tiling across all possible target sites fora panel of six endogenous mouse and three endogenous human genes and quantitatively assessing their ability to produce null alleles by antibody staining and flow cytometry (Doench et al., 2014). The data

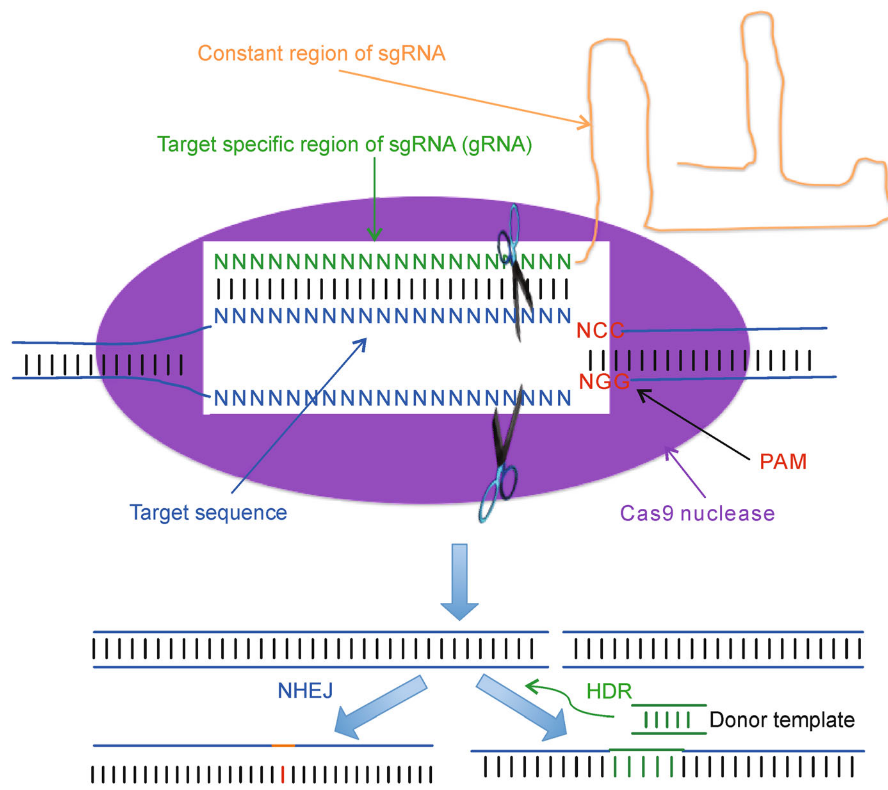

Figure 1 Two components of the CRISPR-Cas9 system from bacteria *S. pyogenes* along with their recognition sites.

from 1841 sgRNAs were used to construct a model to predict the efficacy by fitting a logistic regression using sequence features of the expanded gRNA. The expanded gRNA includes 4 bases upstream of the gRNA and 3 bases downstream of PAM sequence. The predictive model includes 72 features, found statistically significant to contribute to the gRNA efficacy including GC content, some single nucleotide and dinucleotide variants. For example, in position 20, C is highly disfavored and G is strongly favored. The Root laboratory provides an online tool (sgRNA Designer, the 6th tool in Table 1) for predicting gRNA efficacy to facilitate design of highly active sgRNAs for any gene of interest based on this model. Recently, the Liu laboratory further refined the model by incorporating additional features glean from genome-wide sgRNA screens such as a preference for cytosine at the cleavage site (Xu et al., 2015). However, both tools only output efficacy score, the cleavage likelihood for a given gRNA on an intended target, without consideration of potential off-target cleavage.

To effectively apply CRISPR-Cas9 genome editing system, we not only need to select gRNAs with high efficacy, but also need to find gRNAs with low off-target cleavage, i.e., high specificity. Most tools have adopted a simple counting approach to predict off-target effects by listing all genomic sequences containing 0-3 or user-defined maximum number of mismatches to the gRNA (Cradick et al., 2014; Heigwer et al., 2014), and some provide a relative cleavage score for each potential off-target by classifying the target region as seed and non-seed region, and equally penalizing seed region mismatches (Bae et al., 2014; Xiao et al., 2014).

The Zhang laboratory studied the effect of number of mismatches and the mismatch positions of gRNAs on the predicted cleavage rate, by tested >700 gRNA variants for 15 target sequences in a human cell line (Hsu et al., 2013). Briefly, the cells were transfected with gRNA variants containing all possible single nucleotide mismatches and a subset of multiple mismatches and the lesion rates were compared to the cognate gRNAs by deep sequencing PCR products spanning the region of each target site. It turns out that not only the number of mismatch but also their position impacts the activity of the gRNA. Table 2 contains the penalty weights (0-1) to capture the position-dependent mismatch effect on cleavage, where 0 means no mismatch effect and 1 indicates the biggest effect on cleavage. For example, mismatch at position 1 (most distal to PAM) to 5 has almost no effect on the cleavage activity while mismatches at positions 13 to 20 has a large influence on activity. The Zhang laboratory developed a position specific penalty matrix from this experimental data and used it to develop a web application to evaluate gRNAs based on an aggregated off-target score calculated from the top 100 off-target cleavage scores within the genome (the second tool in Table 1). However, this web application does not provide gRNA efficacy prediction. To date, CRISPRseekis the only tool that performs both efficacy and specificity prediction (the first tool in Table 1) (Zhu et al., 2014).

Other considerations of gRNA design

CRISPR-Cas9 technology evolves rapidly with the characterization of new CRISPR-Cas from different species,

Table 1 Overview of several representative gRNA design tools, that are available at [1]http://www.bioconductor.org/packages/release/bioc/html/CRISPRseek.html, [2]http://crispr.mit.edu, [3]http://www.e-crisp.org/E-CRISP/, [4]http://cas9.cbi.pku.edu.cn, [5]http://www.multicrispr.net, and [6]http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design

Functions/Limits	sgRNA designer[1]	Zhang laboratory[2]	eCRISP[3]	Kong laboratory[4]	CRISPR MultiTargeter[5]	CRISPRseek[6]
Off-target analysis	No	**Yes with PDMM penalty scoring**	Yes without scoring	Yes without scoring	Yes without scoring	**Yes with PDMM penalty scoring**
gRNA efficacy prediction	**Yes**	No	No	No	No	**Yes**
Batch search	No	**Yes**	**Yes**	No	No	**Yes**
Paired configuration	No	**Nickase**	No	No	No	**Nickase, dCas9-Fok1 and Cas9-DBD fusion (dev)**
Cas9 type	SpCas9 only	SpCas9 only	SpCas9 only	SpCas9 only	SpCas9 only	**Flexible**
Alternative scoring matrix for efficacy and off-targets	No	No	No	No	No	**Yes**
Target species	Human and mouse	A specified set	A specified set	A specified set		**Any**
Restriction enzyme sites	No	No	No	No	No	**Yes**
Compare set of sequences	No	No	No	No	**Yes**	**Yes**
Predict secondary structure	No	No	No	**Yes**	No	**Yes**

Table 2 Penalty weights to capture the position-dependent mismatch effect of gRNA on target cleavage, where 0 means no mismatch effect and 1 indicates the biggest effect on cleavage, and position 1 is the most distal from PAM sequence (Hsu et al., 2013)

Mismatch position	Penalty weight
1	0
2	0
3	0.014
4	0
5	0
6	0.395
7	0.317
8	0
9	0.389
10	0.079
11	0.445
12	0.508
13	0.613
14	0.851
15	0.732
16	0.828
17	0.615
18	0.804
19	0.685
20	0.583

which will likely have different preference for PAM sequence and different gRNA length. For example, Cas9 from *Neisseria minigenitis* (*N. minigenitis*) has a different PAM preference of NNNNGATT instead of NGG for *S. pyogeneous* (Hou et al., 2013). As new off-target analysis data becomes available, more informative and accurate penalty matrix and scoring system will be generated (Tsai et al., 2015). Strategies have been developed to reduce off-target cleavage, such as using paired Cas9 nickases (Mali et al., 2013a; Ran et al., 2013a; Cho et al., 2014). RNA-guided Cas9 nickases function as a pair to generate a DSB by binding to genomic neighboring genomic sequences with a flexible spacing but defined orientation to generates two single-stranded breaks. The requirement for two nickases to create a DSB increase specificity since the likelihood of a pair of nickases binding neighboring sites is low. Another paired configuration uses dimeric RNA-guided dCas9FokI nucleases (RFNs) (Tsai et al., 2014), which function similarly to RNA-guided nickases but have more restricted spacing requirements. There will be likely more novel configurations emerging to increase the cleavage specificity as paired nickases and FokI dimerization. Furthermore, there are different methods for synthesis and delivery of nucleases to cells. Each method might impose different constraints on the gRNAs. For example, synthesis of gRNAs *in vivo* from host U6 promoters is more efficient if the first base is guanine and gRNA synthesis *in vitro* using T7 promoters is most efficient when the first two bases are GG.

Once mutations are introduced, methods are needed to screen the resulting cells or animals for sequence alterations at the target sites. One of the simplest and least expensive methods is by restriction enzyme digestion as shown in Fig. 2. In this example, the target site contains the recognition site of a restriction enzyme *Pst*1, colored in red, which overlaps with Cas9 cleavage site shown as green arrow. After PCR amplification of the target locus, if there is no mutation in the target sequence, then the *Pst*1 site will stay intact. Addition of *Pst*1 enzyme will produce two bands while the untreated DNA will produce one band. However, if there is a modification such as one A insertion, then the *Pst*1 site will be disrupted and *Pst*1 will not be able to recognize this site and make a cut. Now the sample treated with *Pst*1 will produce only one band just like that of the untreated sample. It is useful to be able to identify restriction enzymes whose recognition sites overlap with the Cas9 cleavage site so that users can choose to filter out gRNAs without overlapping restriction enzyme sites (RES).

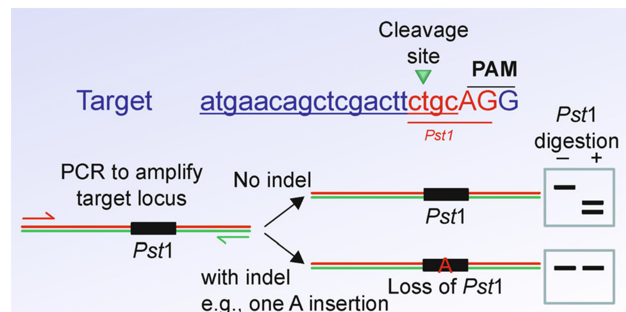

Figure 2 Identify INDELs by restriction enzyme digestion (Courtesy of Dr. Huang Yang at Dr. Michael Green's laboratory, University of Massachusetts).

In addition, there is a need to be able to design gRNAs to analyze closely related sequences, such as targeting one allele but not the other or target both well (Zhu et al., 2014). Recently, a specialized web tool CRISPR MultiTargeter was developed to find common and unique CRISPR single guide RNA targets in a set of similar sequences (Prykhozhij et al., 2015), which is also implemented as *compare2Sequences* function in CRISPRseek package (Zhu et al., 2014).

As described in Fig. 1, the variable region of the sgRNA (gRNA) base-pairs with the target sequence, and the constant region of sgRNA forms several stem-loop structure serving as scaffolding. Therefore, it is important to avoid gRNAs that disrupt the secondary structure of the constant region of the sgRNA, e.g., GUUUUAGAGCUAGAAAUAGCAA-GUUAAAAUAAGGCUAGUCCGUUAUCAACUU-GAAAAAGUGGCACCGAGUCGGUGCUUUUUU. Thus it is important to predict the secondary structure of sgRNA. To date, there are two tools that output the secondary structure of the concatenated sequence of sgRNA (Ma et al., 2013; Zhu et al., 2014).

Overview of *CRISPRseek* functionalities

CRISPRseek was developed with the above considerations in mind to be versatile, flexible and adaptive to rapidly changing needs. Besides efficacy and off-target prediction, *CRISPR-seek* provides flexibility to incorporate alternative paired configuration, other Cas9 types and to plug in alternative penalty matrix and scoring system for efficiency and off-target score prediction from newly published/unpublished source, and to require or exclude specific features within the target site. Additional features include RES annotation, secondary structure prediction and comparison of two sets of sequences. To make it easy to use, all the above functions have been wrapped into two main workflow functions in *CRISPRseek*. One is *offTargetAnalysis* workflow for gRNA (paired or not paired) searching and off-target analysis for one or a set of input sequences (Fig. 3). Several report files are generated including gRNAs in different format, i.e., fasta format, GenBank format, bed format to be visualized in UCSC genome browser (Fig. 4), a tab delimited file containing gRNAs overlap with restriction sites, a tab delimited file containing gRNAs in paired configuration, a tab delimited file containing detailed off-target information such as genomic locations, inside exon or not, mismatch positions, sequence and cleavage score

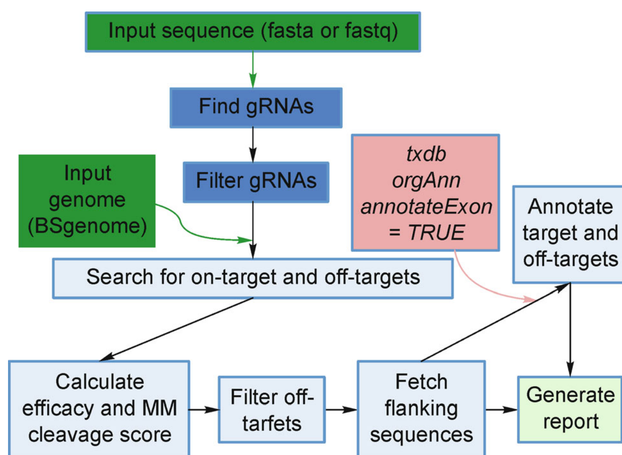

Figure 3 The *offTargetAnalysis* workflow. Major steps in gRNA design are gRNA finding and filtering, on-target and off-target searching, scoring, filtering and annotation, flanking sequence fetching and report generation. The gRNA analysis is colored in blue and off-target analysis is colored in light blue. Required inputs for gRNA and off-target analysis are colored in dark green, optional input is colored in dark salmon, and output is colored in mint green. Several report files are generated including gRNAs in different format, i.e., fasta format, GenBank format, bed format to be visualized in UCSC genome browser, a tab delimited file containing gRNAs overlap with restriction sites, a tab delimited file containing gRNAs in paired configuration, a tab delimited file containing detailed off-target analysis results, a tab delimited file containing a summary of the gRNAs.

(OfftargetAnalysis.xls), and a tab delimited file containing a summary of the gRNAs such as efficacy, RES annotation and top 5 (or a user-specified number) off-target cleavage score (Summary.xls). If RNA secondary structure prediction software *ViennaRNA* (Lorenz et al., 2011) and *GeneRfold* are installed, then the minimum free energy and bracket notation of secondary structure of sgRNA will be generated and included in the summary file. *ViennaRNA* and *GeneRfold* are available at http://www.tbi.univie.ac.at/RNA/index.html#download and http://www.bioconductor.org/packages/2.9/bioc/html/GeneRfold.html.

There are 44 parameters in *offTargetAnalysis* for creating customized search. To make it easy to use, all parameters are set for the widely used CRISPR-Cas9 system from *S. pyogenes*, composed of a 20 base gRNA sequence and a 3 base preferred PAM sequence (NGG). In default setting, you only need to enter the input sequence file path and the genome you are interested in search for off-targets. The gRNA efficacy and off-target cleavage score calculations are based on the models from the Root laboratory (Doench et al., 2014) and the Zhang laboratory respectively (Hsu et al., 2013). Alternative efficacy scoring matrix and off-target mismatch weight matrix can be plugged in as more data and accurate prediction algorithms become available. To identify guide sequence for CRISPR-Cas9 systems from other species that utilize different PAM/gRNA lengths (Hou et al., 2013) or from truncated gRNAs (Fu et al., 2014), which may provide greater specificity, simply adjust the parameters *gRNA.size, PAM, PAM.size, weights, PAM.pattern* and *allowed.mismatch. PAM* accordingly. There is evidence that even though the preferred site is NGG for SpCas9, there is some reduced activity at sites with NAG (Hsu et al., 2013). Therefore, it is recommended to scan for NGG to identify target sites and include both NGG and NAG for off-target search. Parameter *PAM* specifies PAM preference for gRNA search while *PAM. pattern* specifies degenerative PAM for off-target search.

The other workflow function is *compare2Sequences* for identifying gRNAs that specifically target one of the two sets of input sequences or both (Fig. 5). The parameters are almost the same as *offTargetAnalysis* workflow function. In the default setting, all it needs is two sequence/sequence sets file paths. The *compare2sequences* first identifies gRNAs that target one of the input sequences with the same parameters available for the *offTargetSequence* function. Next, for each gRNA, off-target search and scoring were performed against the other sequence(s). Please note that once you identified gRNAs that fit your need, you will still need to run the other workflow function *offTargetAnalysis* to perform genome wide off-target analysis on the chosen gRNAs to ensure that the one you selected not only target one/all input sequences but also cut rarely elsewhere in the genome. For detailed information on parameter setting and example use cases, please refer to the reference manual and user guide at http://www.bioconductor.org/packages/release/bioc/manuals/CRIS

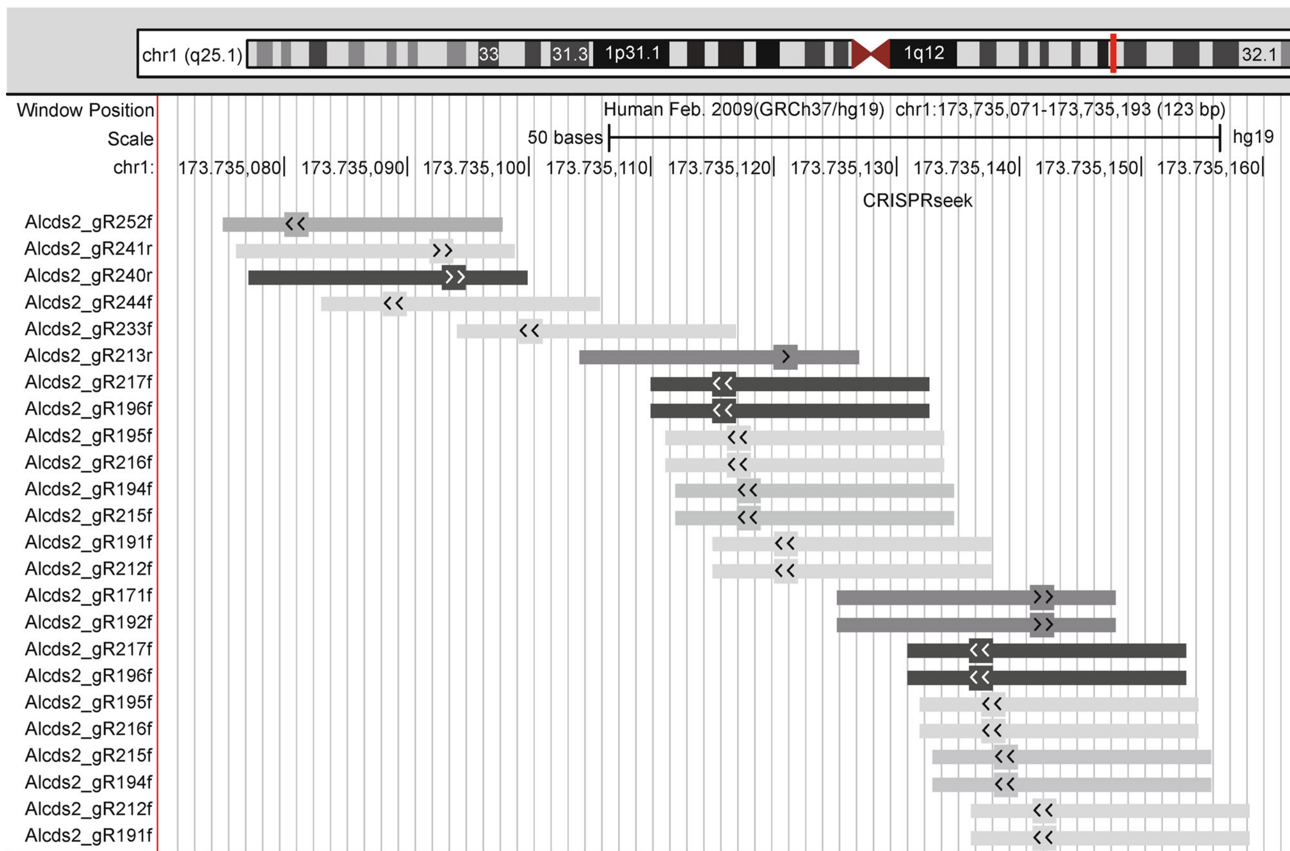

Figure 4 An example of gRNA bed output, visualized in UCSC genome browser, one row per gRNA plotted in gray scale to reflect the gRNA efficacy. The darker the color, the higher efficacy the gRNA has. The thicker portion of the rectangle is the cas9 cleavage site with arrow indicating the gRNA orientation either in positive or negative strand.

PRseek/man/CRISPRseek.pdf and http://www.bioconductor.org/packages/release/bioc/vignettes/CRISPRseek/inst/doc/CRISPRseek.pdf. The ability to easily alter all parameters in both workflow functions is the key in adapting to a rapidly advancing field.

Future directions

There are hurdles to overcome before CRISPR-Cas9 genome editing technology can be successfully applied for therapeutic uses. Computationally, there is a need to develop a more precise gRNA efficacy and off-target cleavage rate prediction models. Although additional features have been discovered to improve the gRNA efficacy prediction (Xu et al., 2015), ~40% of inefficient sgRNAs are not predictable with the improved sequence model, probably due to the small size of the training and testing data set, or/and other sequence determinants not included in the model such as chromatin structure and sgRNA secondary structure. Recently, Cradick and colleagues developed a web application for searching off-targets allowing indels for the human, mouse, *Caenorhabditis elegans*, and rhesus macaque genomes but without off-target cleavage score prediction (Cradick et al., 2014). It is unclear how bulge formed in gRNA or protospacer due to indels affects off-target cleavage. In addition, experiments suggest that not only mismatch positions, but also mismatch types, e.g., A- > T, A- > G and A- > C affect off-target cleavage (Hsu et al., 2013).With the development of GUIDE-seq and the expanding of CRISPR experimental data sets (Tsai et al., 2015), more comprehensive and accurate predictive models expect to be developed, which can be easily plugged into CRISPRseek to improve gRNA design.

Compliance with ethics guidelines

Lihua Julie Zhu declares that she have no conflict of interest. This article does not contain any studies with human or animal subjects performed by the author.

Acknowledgements

I would like to thank Dr. Scot Wolfe at Department of Molecular, Cell and Cancer Biology in University of Massachusetts Medical School for his critical review of the manuscript and his excellent suggestions.

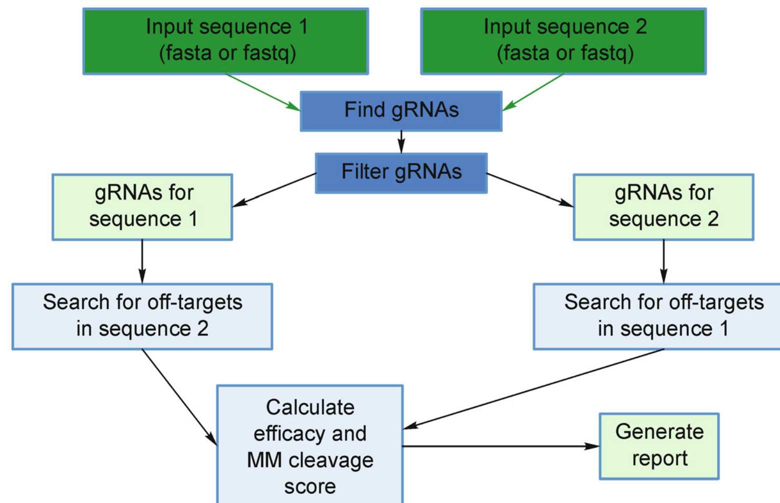

Figure 5 The *compare2Sequences* workflow for design gRNAs that target one but not the other sequences, or that target all sequences equally well. The gRNA analysis is colored in blue and off-target analysis is colored in gray. Required inputs for gRNA and off-target analysis are colored in dark green and output is colored in mint green. Several report files are generated including gRNAs in different format for each input sequence, i.e., fasta format, GenBank format, bed format to be visualized in UCSC genome browser, a tab delimited file containing gRNAs overlap with restriction sites, a tab delimited file containing gRNAs in paired configuration. In addition, it outputs a tab-delimited file containing gRNAs with their predicted efficacy, cleavage score for both sequences and score difference. For applications to target both sequences equally well, you would need to select the gRNAs with minimum absolute score difference. For applications to target one of the sequences but not the other, you would want to select gRNAs with the largest score difference.

Reference

Bae S, Park J, Kim J S (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics, 30(10): 1473–1475

Chen S, Sanjana N E, Zheng K, Shalem O, Lee K, Shi X, Scott D A, Song J, Pan J Q, Weissleder R, Lee H, Zhang F, Sharp P A (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell, 160(6): 1246–1260

Cho S W, Kim S, Kim Y, Kweon J, Kim H S, Bae S, Kim J S (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res, 24(1): 132–141

Chu S W, Noyes M B, Christensen R G, Pierce B G, Zhu L J, Weng Z, Stormo G D, Wolfe S A (2012). Exploring the DNA-recognition potential of homeodomains. Genome Res, 22(10): 1889–1898

Cong L, Ran F A, Cox D, Lin S, Barretto R, Habib N, Hsu P D, Wu X, Jiang W, Marraffini L A, Zhang F (2013). Multiplex genome engineering using CRISPR/Cas systems. Science, 339(6121): 819–823

Cradick T J, Qiu P, Lee C M, Fine E J, Bao G (2014). COSMID: A Web-based tool for identifying and validating CRISPR/Cas off-target sites. Mol Ther Nucleic Acids, 3(12): e214

Ding Q, Regan S N, Xia Y, Oostrom L A, Cowan C A, Musunuru K (2013). Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell, 12(4): 393–394

Doench J G, Hartenian E, Graham D B, Tothova Z, Hegde M, Smith I, Sullender M, Ebert B L, Xavier R J, Root D E (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol, 32(12): 1262–1267

Doudna J A, Charpentier E (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science, 346(6213): 1258096

Enuameh M S, Asriyan Y, Richards A, Christensen R G, Hall V L, Kazemian M, Zhu C, Pham H, Cheng Q, Blatti C, Brasefield J A, Basciotta M D, Ou J, McNulty J C, Zhu L J, Celniker S E, Sinha S, Stormo G D, Brodsky M H, Wolfe S A (2013). Global analysis of *Drosophila* Cys_2-His_2 zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. Genome Res, 23 (6): 928–940

Esvelt K M, Mali P, Braff J L, Moosburner M, Yaung S J, Church G M (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods, 10(11): 1116–1121

Friedland A E, Tzur Y B, Esvelt K M, Colaiácovo M P, Church G M, Calarco J A (2013). Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. Nat Methods, 10(8): 741–743

Fu Y, Sander J D, Reyon D, Cascio V M, Joung J K (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol, 32(3): 279–284

Gratz S J, Cummings A M, Nguyen J N, Hamm D C, Donohue L K, Harrison M M, Wildonger J, O'Connor-Giles K M (2013). Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. Genetics, 194(4): 1029–1035

Gupta A, Meng X, Zhu L J, Lawson N D, Wolfe S A (2011). Zinc finger protein-dependent and-independent contributions to the *in vivo* off-target activity of zinc finger nucleases. Nucleic Acids Res, 39(1): 381–392

Heigwer F, Kerr G, Boutros M (2014). E-CRISP: fast CRISPR target site identification. Nat Methods, 11(2): 122–123

Horvath P, Barrangou R (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science, 327(5962): 167–170

Hou Z, Zhang Y, Propson N E, Howden S E, Chu L F, Sontheimer E J, Thomson J A (2013). Efficient genome engineering in human

pluripotent stem cells using Cas9 from *Neisseria meningitidis*. Proc Natl Acad Sci USA, 110(39): 15644–15649

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 (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol, 31(9): 827–832

Hwang W Y, Fu Y, Reyon D, Maeder M L, Tsai S Q, Sander J D, Peterson R T, Yeh J R, Joung J K (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol, 31(3): 227–229

Ikmi A, McKinney S A, Delventhal K M, Gibson M C (2014). TALEN and CRISPR/Cas9-mediated genome editing in the early-branching metazoan *Nematostella vectensis*. Nat Commun, 5: 5486

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J A, Charpentier E (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337(6096): 816–821

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J (2013). RNA-programmed genome editing in human cells. eLife, 2: e00471

Joung J K, Sander J D (2013). TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol, 14(1): 49–55

Koike-Yusa H, Li Y, Tan E P, Velasco-Herrera M C, Yusa K (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol, 32(3): 267–273

Koonin E V, Makarova K S (2009). CRISPR-Cas: an adaptive immunity system in prokaryotes. F1000 Biol Rep, 1: 95

Koonin E V, Makarova K S (2013). CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. RNA Biol, 10(5): 679–686

Li D, Qiu Z, Shao Y, Chen Y, Guan Y, Liu M, Li Y, Gao N, Wang L, Lu X, Zhao Y, Liu M (2013). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat Biotechnol, 31(8): 681–683

Lorenz R, Bernhart S H, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler P F, Hofacker I L (2011). ViennaRNA Package 2.0. Algorithms Mol Biol, 6(1): 26

Ma M, Ye A Y, Zheng W, Kong L (2013). A guide RNA sequence design platform for the CRISPR/Cas9 system for model organism genomes. BioMed Res Int, 2013: 270805

Mali P, Aach J, Stranges P B, Esvelt K M, Moosburner M, Kosuri S, Yang L, Church G M (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol, 31(9): 833–838

Mali P, Yang L, Esvelt K M, Aach J, Guell M, DiCarlo J E, Norville J E, Church G M (2013b). RNA-guided human genome engineering via Cas9. Science, 339(6121): 823–826

Meng X, Noyes M B, Zhu L J, Lawson N D, Wolfe S A (2008). Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. Nat Biotechnol, 26(6): 695–701

Prykhozhij S V, Rajan V, Gaston D, Berman J N (2015). CRISPR multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. PLoS ONE, 10(3): e0119372

Ran F A, Hsu P D, Lin C Y, Gootenberg J S, Konermann S, Trevino A E, Scott D A, Inoue A, Matoba S, Zhang Y, Zhang F (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell, 154(6): 1380–1389

Ran F A, Hsu P D, Wright J, Agarwala V, Scott D A, Zhang F (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc, 8 (11): 2281–2308

Sampson T R, Saroj S D, Llewellyn A C, Tzeng Y L, Weiss D S (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature, 497(7448): 254–257

Shalem O, Sanjana N E, Hartenian E, Shi X, Scott D A, Mikkelsen T S, Heckl D, Ebert B L, Root D E, Doench J G, Zhang F (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science, 343(6166): 84–87

Smith C, Gore A, Yan W, Abalde-Atristain L, Li Z, He C, Wang Y, Brodsky R A, Zhang K, Cheng L, Ye Z (2014). Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. Cell Stem Cell, 15 (1): 12–13

Tsai S Q, Wyvekens N, Khayter C, Foden J A, Thapar V, Reyon D, Goodwin M J, Aryee M J, Joung J K (2014). Dimeric CRISPR RNA-guided *Fok*I nucleases for highly specific genome editing. Nat Biotechnol, 32(6): 569–576

Tsai S Q, Zheng Z, Nguyen N T, Liebers M, Topkar V V, Thapar V, Wyvekens N, Khayter C, Iafrate A J, Le L P, Aryee M J, Joung J K (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol, 33(2): 187–197

Wang T, Wei J J, Sabatini D M, Lander E S (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science, 343(6166): 80–84

Wyman C, Kanaar R (2006). DNA double-strand break repair: all's well that ends well. Annu Rev Genet, 40(1): 363–383

Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, Gao G, Zhang B (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. Bioinformatics

Xu H, Xiao T, Chen C H, Li W, Meyer C, Wu Q, Wu D, Cong L, Zhang F, Liu J S, Brown M, Liu S X (2015). Sequence determinants of improved CRISPR sgRNA design. Genome Res: gr.191452.115

Yang H, Wang H, Shivalila C S, Cheng A W, Shi L, Jaenisch R (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell, 154(6): 1370–1379

Zhu L J, Holmes B R, Aronin N, Brodsky M H (2014). CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PLoS ONE, 9(9): e108424