



## Mini Review

# Computational approaches for effective CRISPR guide RNA design and evaluation

Guanqing Liu<sup>a,b</sup>, Yong Zhang<sup>a,c,\*</sup>, Tao Zhang<sup>a,b,d,\*</sup><sup>a</sup> Jiangsu Key Laboratory of Crop Genetics and Physiology/ Key Laboratory of Plant Functional Genomics of the Ministry of Education/ Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, Agricultural College of Yangzhou University, Yangzhou 225009, China<sup>b</sup> Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou 225009, China<sup>c</sup> Department of Biotechnology, School of Life Sciences and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China<sup>d</sup> Joint International Research Laboratory of Agriculture and Agri-Product Safety, The Ministry of Education of China, Yangzhou University, Yangzhou 225009, China

## ARTICLE INFO

## Article history:

Received 29 August 2019

Received in revised form 9 November 2019

Accepted 15 November 2019

Available online 29 November 2019

## Keywords:

CRISPR

Guide RNA design

Efficiency

Specificity

Machine-learning

## ABSTRACT

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/ CRISPR-associated (Cas) system has emerged as the main technology for gene editing. Successful editing by CRISPR requires an appropriate Cas protein and guide RNA. However, low cleavage efficiency and off-target effects hamper the development and application of CRISPR/Cas systems. To predict cleavage efficiency and specificity, numerous computational approaches have been developed for scoring guide RNAs. Most scores are empirical or trained by experimental datasets, and scores are implemented using various computational methods. Herein, we discuss these approaches, focusing mainly on the features or computational methods they utilize. Furthermore, we summarise these tools and give some suggestions for their usage. We also recommend three versatile web-based tools with user-friendly interfaces and preferable functions. The review provides a comprehensive and up-to-date overview of computational approaches for guide RNA design that could help users to select the optimal tools for their research.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Contents

1. Introduction	36
2. Evaluation of CRISPR cleavage efficiency	36
2.1. Guide RNA sequence features	36
2.2. Genetic and epigenetic features	37
2.3. Energetics properties	38
2.4. Other considerations for gRNA efficiency	38
3. Prediction of CRISPR cutting specificity	38
3.1. Alignment-based methods	38
3.2. Scoring-based methods	39
3.2.1. Hypothesis-driven methods	39
3.2.2. Learning-based methods	39
3.3. Benchmark of scoring-based methods	39
4. Tools for CRISPR guide RNA design	40
4.1. A comprehensive overview of CRISPR guide RNA design tools	40
4.2. Three comprehensive web platforms for CRISPR guide RNA design	41
4.2.1. CHOPCHOP	41
4.2.2. CRISPR RGEN tools	41

\* Corresponding authors.

E-mail addresses: [zhangyong916@uestc.edu.cn](mailto:zhangyong916@uestc.edu.cn) (Y. Zhang), [zhangtao@yzu.edu.cn](mailto:zhangtao@yzu.edu.cn) (T. Zhang).

4.2.3. CRISPOR ..... 41

5. Summary and perspectives ..... 41

Declaration of Competing Interest ..... 41

Acknowledgements ..... 42

References ..... 42

**1. Introduction**

Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)/CRISPR-associated (Cas) systems, such as Cas9 [46] and Cas12a (formerly Cpf1) [118], are the primary tools used for genome editing due to their various abilities to manipulate, detect, and image certain DNA and RNA sequences in the cell [50]. The CRISPR/Cas system was first adapted for genome editing in 2012 [31,46], and subsequent studies have transformed the CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) into a single guide RNA (sgRNA) that can bind to both the Cas9 protein and the target DNA sequence. Cas9 protein and sgRNA complex first scans the appropriate PAM sequence and binds to the targeted genome loci, then the activated HNH and the RuvC nuclease domain of Cas9 function to make a DNA double-strand break (DSB) in the specific region [22,69].

The most frequently used CRISPR nuclease is type II endonuclease Cas9, which recognises the 5'-NGG-3' PAM (SpCas9) [73]. Another popular nuclease is CRISPR type V endonuclease Cas12a (Cpf1), which recognises the 5'-TTTV-3' PAM, and shows high efficiency in both animal and plant organisms [25,48,49,71,97,118,124]. Recently, several other Cas family proteins have also been discovered and adapted for DNA or RNA editing events, including Cas12b, Cas13a and Cas14 [2,37,94].

CRISPR-based gene editing is implemented with sequence-specific nucleases (SSNs) and a sgRNA to achieve precise gene knock-out (KO) or gene knock-in (KI). Additionally, researchers developed a catalytically inactive Cas9 (dCas9) that loses endonuclease activity, and has been adapted for gene expression regulation (CRISPRa/i) and 3D genome studies [33,66,67,70,84,88]. Furthermore, base editing using modified nCas9 has greatly broadened application of the CRISPR system [32,52,123]. Compared with previous mature gene editing tools such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) [9,10,13,100,121], which use engineered proteins to target and cleave specific genome loci, CRISPR is lower in cost of both time and money. This advancing technology is increasingly being

deployed, and has great potential for clinical detection, gene therapy and agricultural improvement [3,21,50,126].

However, two major challenges hinder the development and application of the CRISPR/Cas system: potential off-target effects, and on-target efficiency (Fig. 1) [112,120]. Successful CRISPR guide RNA (gRNA) design can resolve these issues [23,96], and powerful computational approaches facilitate *in silico* gRNA design [19,34,104], thereby enabling the design of specific gRNAs for particular experiments.

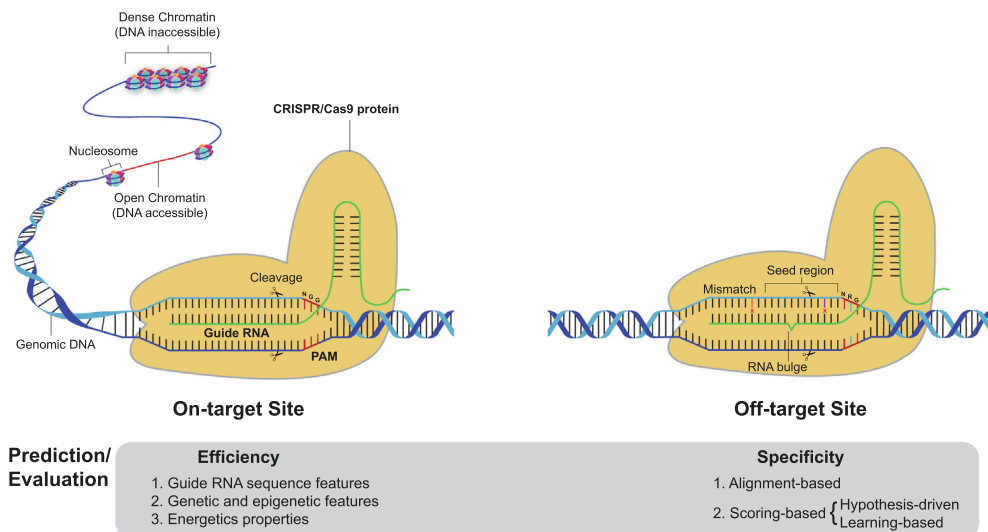
In this review, we summarise existing approaches for CRISPR guide RNA design and evaluation, and assist users in choosing favourable tools for their research. Moreover, it aims to make users aware of the latest computational CRISPR tools and resources.

**2. Evaluation of CRISPR cleavage efficiency**

In theory, the CRISPR/Cas protein scans the PAM sequence, and sgRNA recognises target loci and activates endonuclease activity to cleave specific sites. However, cleavage efficiency varies greatly among different target sites and/or cell lines [14,21,27,28,51,78,87,98,99,103,115,117,122,125], suggesting that several features may influence the binding and cutting efficacy of the sgRNA-Cas complex. Numerous studies have revealed that gRNA sequence features (sequence composition, nucleotide position, GC content), genetic and epigenetic features (chromatin accessibility, gene expression) and energetics properties (RNA secondary structure, melting temperature, free energy) all contribute to gRNA efficacy. Based on these features, many computational tools have been developed for designing highly efficient gRNAs. Herein, we introduce these tools based on their features, and evaluate their efficiency (Table 1).

**2.1. Guide RNA sequence features**

The nucleotide composition of a target sequence is one of the most important determinants of gRNA efficiency



**Fig. 1.** Schematic diagram of CRISPR/Cas9 system at both on-target site and off-target site.

**Table 1**  
Computational methods for evaluation of guide RNA efficiency.

Tool	Enzymes	Data source	Main features	Quantitative metrics
E-CRISP [38]	Cas9	–	SC, GF	–
CRISPRscan [78]	Cas9, Cpf1	Zebrafish	SC	Spearman correlation = 0.309, from [36]
evaluateCrispr [40]	Cas9	<i>Drosophila</i>	SC	Spearman correlation = 0.074, from [36]
sgRNAScorer [14,15]	Cas9, Cpf1	Human	SC, EGF	Spearman correlation = 0.225, from [36]
SSC [112]	Cas9	Human, Mouse	SC	Spearman correlation = 0.274, from [36]
WU-CRISPR [106]	Cas9	Human, Mouse	SC, EP	Spearman correlation = 0.215, from [36]
Azimuth [23,29]	Cas9	Human, Mouse	SC, GF, EP	Spearman correlation = 0.366, from [36]
CRISPRater [55]	Cas9	Human	SC, GF	Pearson correlation = 0.399, from [55]
CRISPRpred [86]	Cas9	Human, Mouse	SC, EP	ROC-AUC = 0.85, from [86]
CASPER [75]	Cas9, Cpf1	–	SC	Pearson correlation = 0.443, from [75]
DeepCpf1 [47]	Cpf1	Human	SC, EGF	Spearman correlation = 0.748, from [47]
TSAM [82]	Cas9	Human, Mouse, Zebrafish	SC, GF, EP	Spearman correlation = 0.4, from [82]
TUSCAN [105]	Cas9	Human, Mouse, Zebrafish	SC, GF	Spearman correlation = 0.12, from [105]
uCRISPR [119]	Cas9	–	SC, EP	Spearman correlation = 0.3, from [119]

SC, sequences composition; GF, genetic features; EGF, epigenetic features; EP, energetics properties.

[24,103,107,112]. Large-scale CRISPR-based screens in mammals have shown that guanines are preferred in positions 1 and 2 before the PAM sequence [103], while thymines are disfavoured within +/-4 nucleotides surrounding the PAM sequence [107]. Additionally, sequences downstream of PAMs can influence gRNA efficiency, while sequences upstream have no significant effect [24]. Cytosine is preferred at the CRISPR/Cas9 cutting site (-3 position proximal to PAM) [21,112], and the GC content of the region 4–13 bases downstream of the PAM sequence contributes to gRNA efficiency. Based on this key information, several efficiency prediction models have been constructed.

Rule Set 1 is a predictive model built using data derived from 1,841 sgRNAs in human and mouse [24], the score is predicted by a support vector machine (SVM) model, and a supervised learning method classifies data in a generalised linear manner. Rule Set 1 is mainly used to investigate sequence features that influence CRISPR cutting efficiency. Results predicted by this model show a high correlation with experimental results. To improve the accuracy, the authors adapted more datasets and built a new model in 2016 called Rule Set 2 [23]. In this model, position-independent nucleotide counts and the location of the sgRNA target site within the gene were considered to improve predictions based on their observations. These optimised models were applied for gRNA design for both CRISPR KO and CRISPRa/i experiments. A package to predict the gRNA efficiency based on the models was also developed and implemented in Broad Institute GPP sgRNA Designer [29].

To unravel the nucleotide preference of gRNA target sites in different CRISPR-based editing events, both CRISPR KO and CRISPRa/i libraries in mammals were screened [112], and significant differences in nucleotide preference between CRISPR KO and CRISPRa/i were detected. Elastic Net is a regularised regression method for fitting and classifying data that performs better than SVM in some cases [128]. The Elastic Net algorithm was used to construct models for both CRISPR KO and CRISPRa/i, and they were applied in Spacer Scoring for CRISPR (SSC) software to predict the efficiency of gRNA. Platforms such as E-CRISP, CHOPCHOP and CRISPR-FOCUS also include this model [12,38,57].

Moreno-Mateos and his colleagues observed that the loading and activity of sgRNA increased with guanine enrichment and adenine depletion [78]. They measured >1,000 sgRNAs targeting 128 genes in zebrafish and used the logistical regression method to construct a predictive model that was integrated into CRISPRscan. WU-CRISPR takes advantage of this data and adds some novel features [24,106], resulting in a model with higher precision than several other predictive models [14,24,112]. Labuhn et al. identified PAM-distal GC content-dependent activity and constructed a model named CRISPRater [55] that was integrated into CRISPR/

Cas9 target online predictor (CCTop), a platform for CRISPR target prediction [93].

The Church laboratory developed software called sgRNA scorer to calculate sgRNA on-target scores based on their SVM model [14]. A second version of the sgRNA scorer software was proposed that improved the on-target prediction power and added prediction for other Cas systems such as SaCas9 and AsCpf1 [15]. Housden et al. then used a drug target method to screen CRISPR KO efficiency in *Drosophila* and developed an efficiency prediction tool [40]. CASPER integrated scores from CRISPRscan and added some new features to maximise correlations between scores and on-target experimental data [75]. This tool can also detect off-target scores and perform multipopulational analysis.

## 2.2. Genetic and epigenetic features

Genetic and epigenetic features like chromatin accessibility, gene position and expression are also important factors that influence sgRNA binding and Cas nucleases cleavage. Many researches have demonstrated that nucleosomes inhibit Cas9 target cleavage, and DNase I hypersensitivity and epigenomic markers alter gRNA efficacy [23,39,44,101,116]. Based on these features, several tools have been developed. By borrowing knowledge from oligonucleotide design and nucleosome occupancy models, an R package called predictSGRNA was proposed for evaluation of sgRNA efficacy [53], and this performed better than other models such as Azimuth and sgRNA scorer [14,23,29].

Non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) are two major pathways which produce heterogeneous repair outcomes when repairing Cas9-mediated DSBs. People use CRISPR/Cas9 system to knock out genes by inducing indels into target genome location. However, CRISPR-based gene KO may induce in-frame variants in which gene functions are retained. Thus, microhomology-based prediction of CRISPR on-target efficiency should be considered. To this end, Bae et al. developed Microhomology-Predictor to improve KO efficiency by reducing in-frame editing [7]. Recent researches have showed that template-free Cas9-editing outcomes are predictable. inDelphi was the first model for precise prediction of CRISPR editing genotype [90]. Soon after, a computational predictor called FORECasT was developed using >40,000 sgRNAs in different cell lines [5]. It was shown that most reproducible mutations are single base insertion, short deletions or longer microhomology-mediated deletions, in addition, Cas9-editing outcomes were cell-line-dependent. The Shendure laboratory also built a predictive model called Lindel for prediction of the insertions and deletions of CRISPR/Cas9-mediated DSB repair based on local sequence context

[16]. All these approaches are sure to assist users in guide RNAs selection for gene disruption.

CRISPR-Cpf1 achieves high efficiency and suffers minor off-targets, besides, Cpf1 prefers AT-enriched regions. Hence, more and more studies have adapted Cpf1 for KO screens. However, models for evaluation of Cpf1 cleavage efficiency are lacking. DeepCpf1 is an algorithm especially for prediction of Cpf1 activity [47]. It is implemented within the deep-learning framework and chromatin accessibility data. This program can significantly improve the accuracy of Cpf1 activity prediction. In addition, models for other Cas experiments, such as Cas9, xCas9, and base edition are also provided on the author's GitHub (<https://github.com/MyungjaeSong/Paired-Library>). CRISPR-DT is a recently developed platform for prediction of Cpf1 target efficiency [127]. This SVM model displayed better performance than the deep learning-based model employed in DeepCpf1.

### 2.3. Energetics properties

The energetics associated with the formation of the DNA, gRNA and Cas protein complex are regular and can be analysed to eliminate bias among different models, since some energetics methods may better illustrate the Cas9 editing efficacy [95,107,113]. CRISPRpred includes the positions of nucleotides as well as secondary structures of sgRNAs to predict the cleavage efficiency, and this performs better than Rule Set 1 [24,86]. Zhang et al. recently demonstrated a free energy scheme called uCRISPR for evaluating the Cas9 editing efficacy, as well as off-target effects [119]. This model is thought to apply to any cleavage-activity Cas9 dataset.

### 2.4. Other considerations for gRNA efficiency

Of note, these predictive models were trained by individual experiments and rules, and each model generates different on-target scores for sgRNAs, hence users must be careful when evaluating or designing guide RNAs using these models in their own experiments. However, several key features are reliable, such as guanine preference, GC content, seed region and the secondary structure of gRNAs [65,91,106,112].

The accuracy of different models is controlled by their learning methods or the approaches of CRISPR activity measurement. Doench et al. tested multiple training methods and selected the best-performing one as the kernel of their model [23]. TUSCAN, a random forest-based model, outperformed models built solely by linear regression [78,105]. A two-step averaging method (TSAM) for the regression of cleavage efficiencies also performed better than many other models [14,23,82,112]. Additionally, measure-

ment using sequencing data rather than phenotypic data may generate fewer false positive results, albeit at a cost [105].

It is critical for users to know which tool best suits their research. A comprehensive evaluation of different efficiency prediction tools was conducted to examine differences among models [36], and it revealed differences in the correlation between different datasets and models. Furthermore, no model performed excellently across all datasets, suggesting that a careful selection of CRISPR gRNA design tools is necessary. Users can also evaluate gRNAs using multiple models and select the best one for their experiments. Among the available tools, Rule Set 2 and DeepCpf1 are the most used and accurate scoring methods for evaluating Cas9 and Cpf1 cutting efficacy, and uCRISPR may be more accurate than some other methods, but it requires further experimental testing.

## 3. Prediction of CRISPR cutting specificity

The main obstacle for the application of CRISPR is off-target effects. CRISPR nucleases may cleave unintended genomic sites and cause unexpected mutations due to sgRNAs recognising DNA sequences with a few mismatches and/or DNA/RNA bulges, referred to as off-target cleavage [41,120]. Off-target effects can be effectively relieved by predicting CRISPR cutting specificity and designing optimal gRNAs [41]. To predict the specificity of CRISPR gRNAs, two main methods have been proposed: (1) alignment-based method. Based on conventional or specialized algorithms, gRNAs are aligned to a given genome and off-target sequences and sites are returned. This method is mainly used for find out all potential off-targets *in silico*, (2) Scoring-based method. sgRNAs should be further scoring and ranking using identified off-targets from alignment process to select the most specific one for experiments. Two scoring approaches are shown: hypothesis-driven, where off-targets are scored based on the contribution of specific genome context factors to gRNA specificity; learning-based, where gRNAs are scored and predicted from a training model that considers the different features affecting specificity. These methods for prediction of gRNA specificity are discussed below and some of them are summarised in Table 2.

### 3.1. Alignment-based methods

In theory, potential off-target sites can be identified by aligning gRNA sequences to the reference genome based on sequence homology. Bowtie [59] and BWA [61] are traditional tools for short sequence alignment that are capable of off-target detection [36,104]. However, there are several potential issues when using these tools. First, tools like Bowtie and BWA cannot identify small PAMs, since these alignment tools were developed for next-

**Table 2**  
Computational methods for prediction of guide RNA specificity.

Tool	Enzymes	Methods	Main features	Quantitative metrics
CasOT [108]	Cas9	alignment	unlimited mismatch number, paired-gRNA mode, annotation	slow
Cas-OFFinder [8]	costom	alignment	unlimited mismatch number, GPU acceleration, web support	middle, fast (use GPU)
sgRNAcas9 [110]	Cas9	alignment	max 5 mismatches, paired-gRNA mode, annotation, risk evaluation	slow
FlashFry [74]	costom	alignment	unlimited mismatch number, multiple on/off-target scores, annotation	fast
Crisflash [43]	Cas9	alignment	unlimited mismatch number, variant data support	fast
MIT [41]	Cas9	scoring	20 bp sgRNA without PAM	ROC-AUC = 0.87, from [36]
CCTop [93]	Cas9, Cpf1	scoring	empirically score based on number of mismatches	ROC-AUC = 0.77, from [36]
CFD [23]	Cas9	scoring	20 bp sgRNA with PAM (enable non-canonical PAM)	ROC-AUC = 0.91, from [36]
CRISPRoff [4]	Cas9	scoring	energetics property and sequences composition	ROC-AUC = 0.98, from [4]
uCRISPR [119]	Cas9	scoring	energetics property and sequences composition	Pearson correlation = 0.75, from [119]
CRISTA [1]	Cas9	scoring	machine learning, sequences composition and epigenetic feature	ROC-AUC = 0.96, from [1]
Elevation [62]	Cas9	scoring	machine learning, integrate both CFD model and epigenetic features	ROC-AUC = 0.98, from [62]
DeepCRISPR [18]	Cas9	scoring	deep learning, sequences composition and epigenetic feature	ROC-AUC = 0.98, from [18]

generation sequencing (NGS) read alignment. Second, these tools allow very limited mismatches in their seed regions, making alignment with default parameters impractical for identifying all potential off-target sites. A survey using bowtie2 [58] to detect off-targets failed to find all possible off-target sites, and could only find off-targets with up to one mismatch [23].

Despite the defects, some gRNA design methods utilise these tools, and parameters have been modified to fit the demand for off-target prediction. CCTop uses Bowtie to find off-target sites by first identifying PAM sites, and matches and mismatches in protospacer sequences are then searched for by bowtie with modified parameters [93]. Up to five mismatches are allowed in the protospacer as more mismatches may prevent DSB induction. CCTop also implements an off-target score for each candidate sgRNA. CHOPCHOP detects off-target sites using Bowtie with parameters -v and -L for searching sgRNA core regions [77]. CRISPOR uses BWA to find all potential off-target sites in iterative mode (“-N”), and can find all validated off-targets as well as Cas-OFFinder [8,36].

In addition to these traditional tools, many other tools and algorithms have been developed for off-target site detection. Cas-OFFinder is one of the most popular tools for searching potential off-target sites, and advantages include no limit to the number of mismatches, PAM types, gRNA length or high running speed with GPUs. Cas-OFFinder can also predict off-target sites with 1 bp deletions or insertions (i.e. DNA/RNA bulges). CasOT is implemented to find Cas9 on-target sites from input sequences, as well as potential off-target sites with up to six mismatches in the seed region (12 nt adjacent to the PAM). This tool can also determine whether off-targets are within a coding exon [108]. Meanwhile, sgRNAs9 utilises the ultrafast short sequence mapping tool SeqMap [45] to find off-targets, and classifies all sites into three categories to generate a final output of the best candidate gRNAs [110]. Recently, two new alignment-based tools have been developed. Crisflash utilises a tree-based algorithm to rapidly design CRISPR guide RNAs and optimise guide accuracy by incorporating user-supplied variant data [43]. FlashFry rapidly searches off-target sites and provides much useful information (annotation of off-target sites, on/off-target scores, GC content, etc.) for candidate gRNAs [74]. Here we still classified Crisflash and FlashFry as alignment-based methods since they both propose novel algorithms for off-target searching, whereas the scoring approaches they use are borrowed from others [23,24,41,78].

Among these tools, Cas-OFFinder may be the best choice for identifying all potential off-target sites with any Cas nucleases, and FlashFry is also worth a try for its high speed and comprehensive outputs.

## 3.2. Scoring-based methods

### 3.2.1. Hypothesis-driven methods

Alignment-based methods are reliable for detection of most potential off-targets, however, not all nucleotide positions containing mismatches have the same decisive effect on off-target cleavage. Additionally, alignment-based prediction always outputs redundant off-target sites, many which are false-positives, although users can reduce the number of outputs by restricting the maximum mismatches when exploring off-target cleavage. One study compared experimentally validated off-targets and off-targets predicted by Cas-OFFinder and CCTop, and the results showed that off-targets detected by the prediction tools only covered some of the validated sites, while some off-target sites cannot be predicted solely based on sequence homology [11]. Thus, features that influence the nonspecific binding of CRISPR gRNAs need to be considered to increase the accuracy of off-target detection.

MIT (Hsu-Zhang) score was proposed for off-target evaluation during the early stages of gene editing by CRISPR. Hsu et al. evalu-

ated >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci [41]. They evaluated the contributions made by different mismatch positions and numbers in the target 20 bp, and calculated a weight matrix to determine off-target efficiency for each sgRNA. The authors claimed that this score accounts for >50% of the variance in cutting frequency. The MIT score has been integrated into many CRISPR gRNA design tools such as CHOPCHOP and CRISPOR [36,56].

Cutting frequency determination (CFD) is another popular score for off-target evaluation. It is noticeable that sgRNA can also bind genome loci with non-canonical PAMs such as NAG, NCG and NGA, which may cause off-target cleavage. Doench et al. added PAM features in their scoring metrics [23]. sgRNAs with mismatches and indels in target sequences were also included for examining correlations between sgRNAs and off-targets. CFD score was validated with GUIDE-seq and proved to perform better than MIT score, hence it was adopted by CRISPRscan [78], GuideScan [83] and CRISPOR [36].

The prediction of sgRNA specificity based on the structural features of the Cas9-sgRNA complex proved to be superior to prediction solely based on sequence features. CRISPROff [4] and uCRISPR [119] both utilise energetics properties for off-target evaluation. Compared with other scoring methods like MIT and CFD, they both yielded better accuracy in off-target prediction. Nevertheless, neither have not been systematically evaluated by large-scale experiments, and caution should be exercised when using them.

### 3.2.2. Learning-based methods

Sometimes, empirical algorithms cannot effectively evaluate off-targets since they only consider minor features, whereas learning-based methods build complex models using combinations of multiple features, and they may better predict off-targets. In recent years, several new approaches for off-target prediction based on machine learning have been developed.

CRISPR Target Assessment (CRISTA) software uses BWA as the off-target search tool and implements multiple features (PAM type, nucleotide composition, GC content, chromatin structure, DNA methylation, RNA secondary structure, etc.) to predict cleavage propensity [1]. CRISTA exhibits better performance than MIT, CCTop and CFD.

The Doench laboratory also developed a linear regression model for prediction of off-target activity called Elevation that takes both sequence and chromatin accessibility features into consideration, as well as observations from others [62]. Elevation predicts individual scores for each off-target site, as well as an aggregate score for gRNAs. This method performs best among MIT, CFD and CCTop. However, this software can only compute off-target scores for the human exome (GRCh38) and cannot support other organisms on their website.

DeepCRISPR is a state-of-the-art computational platform that unifies sgRNA on-target and off-target site prediction into one framework with deep learning [18]. It identifies all possible sequence and epigenetic features that may affect sgRNA KO efficacy by learning from large data. This method was proved to surpass other available off-target prediction tools [23,41,92,93].

## 3.3. Benchmark of scoring-based methods

Different scoring methods are based on different characteristics applied by each individual laboratory. In order to compare the accuracy of different scores, Haeussler et al. assessed four hypothesis-driven methods for off-target prediction using the same datasets [36]. CFD score exhibited the best prediction accuracy, whereas CCTop performed the worst. Data imbalance, where the number of true observed off-target sites (OTS) identified by off-target detection experiments is much less than that of all potential

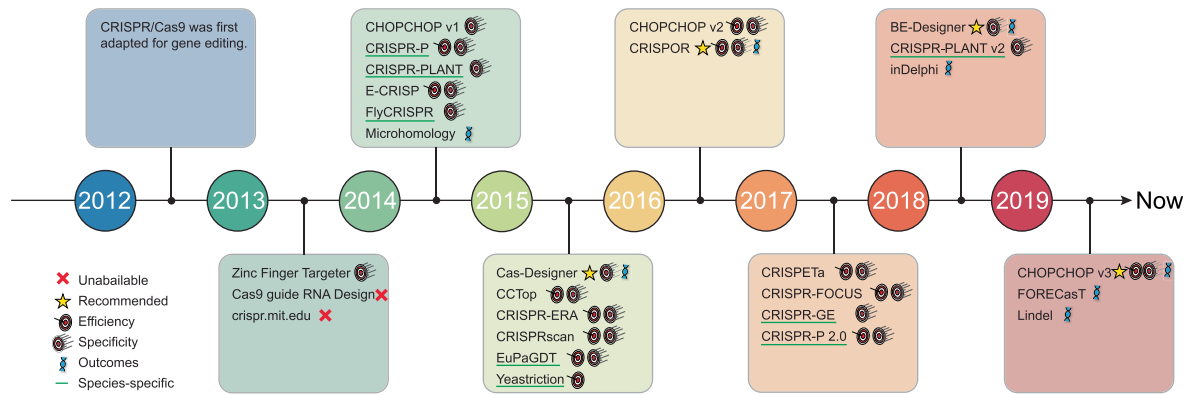


Fig. 2. Timeline of the development of web-based tools for CRISPR guide RNA design.

Table 3  
Web-based CRISPR guide RNA design tools.

Tool	Website
CHOPCHOP [56,57,77]	<a href="https://chopchop.cbu.uib.no">https://chopchop.cbu.uib.no</a>
CRISPR RGEN Tools [42,79]	<a href="http://www.rgenome.net/">http://www.rgenome.net/</a>
CRISPOR [20,36]	<a href="http://crispor.tefor.net">http://crispor.tefor.net</a>
E-CRISP [38]	<a href="http://www.e-crisp.org/E-CRISP">http://www.e-crisp.org/E-CRISP</a>
CCTop [93]	<a href="https://crispr.cos.uni-heidelberg.de">https://crispr.cos.uni-heidelberg.de</a>
CRISPR-ERA [64]	<a href="http://CRISPR-ERA.stanford.edu">http://CRISPR-ERA.stanford.edu</a>
CRISPEta [85]	<a href="http://crispeta.crg.eu">http://crispeta.crg.eu</a>
CRISPRscan [78]	<a href="https://www.crisprscan.org">https://www.crisprscan.org</a>
EuPaGDT [81]	<a href="http://grna.ctegd.uga.edu">http://grna.ctegd.uga.edu</a>
CRISPR-P [60,63]	<a href="http://crispr.hzau.edu.cn/CRISPR2/">http://crispr.hzau.edu.cn/CRISPR2/</a>
CRISPR-PLANT [76,109]	<a href="https://www.genome.arizona.edu/crispr2">https://www.genome.arizona.edu/crispr2</a>
CRISPR-GE [111]	<a href="http://skl.scau.edu.cn">http://skl.scau.edu.cn</a>
inDelphi* [90]	<a href="https://www.crisprindelphi.design/">https://www.crisprindelphi.design/</a>
FORECasT* [5]	<a href="https://partslab.sanger.ac.uk/FORECasT">https://partslab.sanger.ac.uk/FORECasT</a>
Lindel* [16]	<a href="https://shendurelab.github.io/Lindel/">https://shendurelab.github.io/Lindel/</a>

Tools marked by asterisk (\*) are used for prediction of CRISPR editing outcomes.

off-target sites recognized by alignment-based methods, is a common issue in the learning-based methods. To address the problem, a systematic survey was conducted to assess the influence of data imbalance [30]. The authors used bootstrapping sampling and PR-AUC methods to evaluate two well-established models. They emphasized the importance of using symmetric data for model construct and taking unbiased datasets for benchmark. According to current assessment results, we recommend people to use CFD score for off-target prediction. Elevation and DeepCRISPR are suitable for human genome editing, but CRISPROff and uCRISPR may need further evaluation before use.

#### 4. Tools for CRISPR guide RNA design

A CRISPR/Cas complex typically contains a Cas protein and a sgRNA, both of which determine the cutting activity. However, protein engineering is a complicated and uncertain strategy for most researchers, and optimising gRNAs is more feasible and efficient. A robust gRNA requires maximum on-target efficiency as well as minimum off-target activity. Basing on these two criteria, numerous computational tools have been developed for high-efficiency CRISPR gRNA design. However, each tool possesses distinct features, and user-friendliness is also important. To acquaint users with existing gRNA design tools, we provide an overview of most that are available based on the strength of their features, and this help users to make appropriate selections. Furthermore, we recommend several functional and user-friendly websites for gRNA design (Fig. 2, Table 3).

#### 4.1. A comprehensive overview of CRISPR guide RNA design tools

The CRISPR/Cas system was first adapted for gene editing in 2012, and several tools were developed immediately thereafter, such as Zinc Finger Targeter (ZiFiT) [89], Cas9 guide RNA Design [68] and CRISPR (<http://crispr.mit.edu/>) [41]. All these platforms have been implemented in the design of gRNAs for model organisms. ZiFiT is also used to design Zinc Finger and TALEN modules, but Cas9 guide RNA Design and CRISPR are now unavailable.

In the following years, various tools for CRISPR gRNA design emerged, due to both urgent demand for these tools, and because people have learned more and more about how the CRISPR system functions and what influences the efficiency and specificity of Cas cleavage. Many of these tools combine multiple scoring methods and provide alternative options for users. Some others have also proposed their own algorithms to rank designed sgRNAs, and this can assist users in gRNA selection. CHOPCHOP [56,57,77] provides alternative scores for users such as Rule Set ½ [23,24], SSC [112], CRISPRscan [78] and deepCpf1 [47]. E-CRISP utilises its own SAE (Specificity, Annotation, Efficacy) score to determine the quality of each sgRNA, while Rule Set 1 [24] and SSC [112] are also included in E-CRISP. CCTop [93] assigns off-target scores for each sequence empirically, while the CRISPRater score [55] is also included to predict the efficiency of sgRNAs. CRISPOR [36] is a versatile platform that ranks gRNAs according to different scores for evaluating potential off-targets in the genome of interest, and for predicting on-target activity.

A large number of CRISPR/Cas-derived RNA-guided Endonucleases (RGENs) have been identified or modified to improve the cutting efficiency and enlarge the editing range. Some tools enable the design of gRNAs for RGENs. For example, Cas-Designer [79] allows users to choose 20 PAM types from different RGENs, while CRISPOR [20] also offers various PAMs from a defined list. To best serve the specialised purposes of different experiments, several web-based tools have been developed. CRISPR-ERA [64] and CHOPCHOP v3 [57] support sgRNA design for the CRISPRa/i system. CRISPEta [85] is mainly used for genome deletion with paired gRNAs, and BE-Designer [42] can be used for designing gRNA for base editing. Recently, three methods have been employed for successively predicting Cas9-editing outcomes: inDelphi [90], FORECasT [5] and Lindel [16]. These approaches can help to identify gRNAs based on forecasting results.

Researchers should choose suitable tools with caution since different tools are in favour of different genomes or cell types. For instance, Yeastriktion [72] is specialised for yeast, FlyCRISPR [35] is specific for *Drosophila*, EuPaGDT [81] is recommended for eukaryotic pathogens, and CRISPR-P [60,63], CRISPR-PLANT [76,109] and CRISPR-GE [111] are all implemented for genome

editing in plants. Also, user-friendliness is essential for these web-based design tools. Based on these considerations, we propose three comprehensive web-based platforms for CRISPR gRNA design.

#### 4.2. Three comprehensive web platforms for CRISPR guide RNA design

##### 4.2.1. CHOPCHOP

CHOPCHOP has a succinct interface with well-rounded functions, >200 genomes are available on the website, and users can provide gene names, genomic coordinates or target sequences as inputs. If a gene is given, users can set specific regions of the gene as targets, such as coding sequences or promoters. Before designing gRNAs, two off-target detection methods and seven efficiency scores can be chosen, which aids users in selecting optimal gRNAs for their research.

To satisfy different experimental purposes, CRISPR Cas9 nuclease/nickase, Cpf1, CasX, C2C2 and TALEN are all supported by CHOPCHOP, and various modes of DNA targeting are optional such as KO/KI, gene activation/repression, and nanopore enrichment. In activation/repression mode, gRNAs are designed to target the promoter region and its flanking sites in order to bring the activating/repressing domain into close proximity with the transcription start site [17,54]. Meanwhile, nanopore enrichment mode is implemented to design sgRNAs for long fragment excision within 40 kb. Additionally, the inDelphi model has been integrated into CHOPCHOP for repair profile prediction by Cas9 in five cell types [90].

After clicking the 'Find Target Sites' button, a results table is shown, and each candidate gRNA has a rank, genomic information, GC content, self-complementarity score, mismatch number (0–3) and an efficiency score. A graphical view of all gRNAs is also provided in the UCSC Genome Browser, which helps users to select optimal gRNAs.

##### 4.2.2. CRISPR RGEN tools

CRISPR RGEN tools, computational tools and libraries for RNA-guided endonucleases, are maintained by the Bae laboratory, and libraries comprise nine useful tools including Cas-OFFinder [8], Cas-Designer [79] and Digenome-Seq [80]. Herein, we mainly discuss the two gRNA design tools Cas-Designer and BE-Designer [42].

Compared with other designer tools, Cas-Designer allows DNA/RNA bulges when performing off-target detection. Additionally, this detection method is more rapid than others due to the Cas-OFFinder algorithm. Genomic sequences or coordinates and fasta file formats are allowed as inputs. Over 350 genomes and 20 PAM types are specified for users, and outputs include target sequences as well as out-of-frame score (calculate by microhomology), mismatch number (0–2) and off-target sites with up to 1 bp bulge. On/off-target sites are redirected to the Ensembl genome browser [26], which is capable of further inspection.

Unlike Cas-Designer, BE-Designer is primarily implemented for base editing. In this tool, four methods of base editing are specified, and the editing window region is also adjustable. After the design phase, three outcomes are based on different coding types, and gRNAs, editing window sequences and amino acids are highlighted in a user-friendly manner. This tool is a good choice for base editing.

##### 4.2.3. CRISPOR

Since many tools have been developed for highly efficient CRISPR gRNA design, an ensemble of multiple tools can be useful, and CRISPOR does this well [20,36]. At present, CRISPOR contains 417 genomes and 19 PAM types, and can meet the needs of most users. CRISPOR takes genome coordinates and sequences <2,000 bp as inputs and provides comprehensive information as

outputs. By default, results are shown in two parts; visualisation of PAM sites along the given sequence, which allows for downloading using multiple formats including fasta, GenBank and SnapGene; a table including all information for each predicted gRNA is also generated. In the table, two specificity scores (MIT and CFD) and 10 efficiency scores (Rule Set 2, CRISPRscan, etc.) are calculated [23,41,78]. Furthermore, Microhomology and Lindel scores are also provided for outcome prediction [7,16]. Warnings such as extreme GC content and poly-T values are indicated, the table is downloadable, and candidate off-target sites with up to four mismatches can be visualised and downloaded.

In addition to gRNA design, CRISPOR designs primers for each gRNA as well as off-target sites. These primers are useful when conducting on/off-target validation or *in vitro* expression experiments. Furthermore, CRISPOR enables filtering of gRNAs with genomic variants based on well-known variant databases or marked input sequences with N characters. Thus, CRISPOR may be the best tool for designing gRNAs.

## 5. Summary and perspectives

CRISPR/Cas technology has emerged as an advanced strategy for functional genomics, crop breeding and precise medicine. Guide RNAs are indispensable for CRISPR-based editing, and computational approaches provide assistance for efficient gRNA design. Numerous tools have been developed for CRISPR gRNA design and evaluation. However, many issues remain. For example, experimental datasets used to build models for predicting sgRNA specificity or efficiency are disparate; CRISPRscan score performs worse in mammals than in zebrafish, the genome the predictive model was trained in [36]. Researchers should therefore know the strengths and weaknesses of each scoring method, and select the optimal tool for their research.

As we become more aware of the mechanisms by which Cas proteins recognise gRNAs, and bind to and cleave target loci, more and more novel features contributing to cutting efficiency and specificity are being identified, including sequence composition. For instance, low chromatin accessibility may block access of Cas9, while open chromatin regions are more likely to cause undesired mutations [39,92]. In practice, specificity is more important than efficiency for application of CRISPR. However, several off-targets remain indecipherable using current tools, and discovery of novel features may minimise off-target effects.

Too much choice is not always desirable, and an unbiased evaluation can provide guidance. To this end, the Liu laboratory has conducted the benchmarking [30,114], and Haeussler et al. performed a comparison of different predictive scores and integrated most into CRISPOR [36]. None of the tools excel when using heterogeneous data, due to cell-specific or species-specific training models. Therefore, more cell types need to be evaluated. Also, plant and some animal genomes should be analysed thoroughly, since human or mouse have dominated to date. Moreover, it will be beneficial if useful features of multiple tools are combined in future software packages.

The major outcomes of Cas9 cleavage is non-random and predictable [102], and several tools have been created for prediction of CRISPR-Cas9 outcomes [5,16,90]. Such findings facilitate more accurate gene editing. However, existing tools cannot account for large indels, homozygous sites and the activity of nucleases other than SpCas9, and this should be resolved in future [6].

In summary, the development of computational approaches has revolutionised the design of effective gRNAs, and CRISPR-based gene therapies and customised gene editing may come within reach. It is likely that CRISPR technology will continue to become more powerful in the coming years.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the National Transgenic Major Project (2019ZX08010003-001-002, 2018ZX08020-003), Jiangsu Province Government Project (BK2018003)/The open funds of the Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding (PL201801) for T.Z. and Y.Z.; the Fund of Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) for T.Z.

## References

- Abadi S, Yan WX, Amar D, Mayrose I. A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS Comput Biol* 2017;13.
- Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, et al. RNA targeting with CRISPR-Cas13. *Nature* 2017;550:280–4.
- Adli M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 2018;9.
- Alkan F, Wenzel A, Anthon C, Havgard JH, Gorodkin J. CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. *Genome Biol* 2018;19.
- Allen F, Crepaldi L, Alsinet C, Strong AJ, Kleshchevnikov V, De Angeli P, et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol* 2019;37:64–72.
- Bae S, Kim JS. Machine learning finds Cas9-edited genotypes. *Nat Biomed Eng* 2018;2:892–3.
- Bae S, Kweon J, Kim HS, Kim JS. Microhomology-based choice of Cas9 nuclease target sites. *Nat Methods* 2014;11:705–6.
- Bae S, Park J, Kim JS. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 2014;30:1473–5.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009;326:1509–12.
- Bogdanove AJ, Voytas DF. TAL effectors: customizable proteins for DNA targeting. *Science* 2011;333:1843–6.
- Cameron P, Fuller CK, Donohue PD, Jones BN, Thompson MS, Carter MM, et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* 2017;14:600–6.
- Cao Q, Ma J, Chen CH, Xu H, Chen Z, Li W, et al. CRISPR-FOCUS: a web server for designing focused CRISPR screening experiments. *PLoS ONE* 2017;12.
- Carroll D. Genome engineering with zinc-finger nucleases. *Genetics* 2011;188:773–82.
- Chari R, Mali P, Moosburner M, Church GM. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* 2015;12:823–6.
- Chari R, Yeo NC, Chavez A, Church GM. sgRNA Scorer 2.0: a species-independent model to predict CRISPR/Cas9 activity. *ACS Synth Biol* 2017;6:902–4.
- Chen W, McKenna A, Schreiber J, Haeussler M, Yin Y, Agarwal V, et al. Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res* 2019;47:7989–8003.
- Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res* 2013;23:1163–71.
- Chuai GH, Ma HH, Yan JF, Chen M, Hong NF, Xue DY, et al. DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol* 2018;19.
- Chuai GH, Wang QL, Liu Q. In silico meets *in vivo*: towards computational CRISPR-based sgRNA design. *Trends Biotechnol* 2017;35:12–21.
- Concordet JP, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* 2018;46:W242–5.
- Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;339:819–23.
- Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao YJ, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 2011;471:602–7.
- Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 2016;34:184–91.
- Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 2014;32:1262–7.
- Endo A, Masafumi M, Kaya H, Toki S. Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Sci Rep-Uk* 2016;6.
- Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, et al. Ensembl 2011. *Nucleic Acids Res* 2011;39:D800–6.
- Fu YF, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013;31:822–6.
- Fu YF, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 2014;32:279–84.
- Fusi N, Smith I, Doench J, Listgarten J. In silico predictive modeling of CRISPR/Cas9 guide efficiency. *bioRxiv* 2015.
- Gao Y, Chuai G, Yu W, Qu S, Liu Q. Data imbalance in CRISPR off-target prediction. *Brief Bioinform* 2019;bbz069.
- Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *P Natl Acad Sci USA* 2012;109:E2579–86.
- Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A.T to G.C in genomic DNA without DNA cleavage. *Nature* 2017;551:464–71.
- Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 2014;159:647–61.
- Graham DB, Root DE. Resources for the design of CRISPR gene editing experiments. *Genome Biol* 2015;16.
- Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* 2014;196:961–71.
- Haeussler M, Schonig K, Eckert H, Eschstruth A, Mianne J, Renaud JB, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* 2016;17.
- Harrington LB, Burstein D, Chen JS, Paez-Espino D, Ma E, Witte IP, et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* 2018;362:839–42.
- Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. *Nat Meth* 2014;11:122–4.
- Horlbeck MA, Witkowsky LB, Guglielmi B, Replogle JM, Gilbert LA, Villalta JE, et al. Nucleosomes impede Cas9 access to DNA *in vivo* and *in vitro*. *Elife* 2016;5.
- Housden BE, Valvezan AJ, Kelley C, Sopko R, Hu Y, Roessel C, et al. Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. *Sci Signal* 2015;8.
- Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013;31:827–32.
- Hwang GH, Park J, Lim K, Kim S, Yu J, Yu E, et al. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinform* 2018;19:542.
- Jacquain AL, Odum DT, Lukk M. Crisflash: open-source software to generate CRISPR guide RNAs against genomes annotated with individual variation. *Bioinformatics* 2019;35:3146–7.
- Jensen KT, Floe L, Petersen TS, Huang J, Xu F, Bolund L, et al. Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett* 2017;591:1892–901.
- Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* 2008;24:2395–6.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816–21.
- Kim HK, Min S, Song M, Jung S, Choi JW, Kim Y, et al. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat Biotechnol* 2018;36:239–41.
- Kim Y, Cheong SA, Lee JG, Lee SW, Lee MS, Baek JJ, et al. Generation of knockout mice by Cpf1-mediated gene targeting. *Nat Biotechnol* 2016;34:808–10.
- Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* 2016;34:869–74.
- Knott GJ, Doudna JA. CRISPR-Cas guides the future of genetic engineering. *Science* 2018;361:866–9.
- Koike-Yusa H, Li YL, Tan EP, Velasco-Herrera MD, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* 2014;32:267–73.
- Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 2016;533:420–4.
- Kuan PF, Powers S, He S, Li K, Zhao X, Huang B. A systematic evaluation of nucleotide properties for CRISPR sgRNA design. *BMC Bioinform* 2017;18.
- La Russa MF, Qi LS. The New State of the Art: Cas9 for gene activation and repression. *Mol Cell Biol* 2015;35:3800–9.
- Labuhn M, Adams FF, Ng M, Knoess S, Schambach A, Charpentier EM, et al. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res* 2018;46:1375–85.



- [56] Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* 2016;44:W272–6.
- [57] Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* 2019;47:W171–4.
- [58] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
- [59] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10.
- [60] Lei Y, Lu L, Liu HY, Li S, Xing F, Chen LL. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant* 2014;7:1494–6.
- [61] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- [62] Listgarten J, Weinstein M, Kleinstiver BP, Sousa AA, Joung JK, Crawford J, et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat Biomed Eng* 2018;2:38–47.
- [63] 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;10:530–2.
- [64] Liu HL, Wei Z, Dominguez A, Li YD, Wang XW, Qi LS. CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. *Bioinformatics* 2015;31:3676–8.
- [65] 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;6:19675.
- [66] Lowder LG, Zhang D, Baltes NJ, Paul 3rd JW, Tang X, Zheng X, et al. A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol* 2015;169:971–85.
- [67] Lowder LG, Zhou J, Zhang Y, Malzahn A, Zhong Z, Hsieh TF, et al. Robust transcriptional activation in plants using multiplexed CRISPR-Act2.0 and mTALE-Act systems. *Mol Plant* 2018;11:245–56.
- [68] Ma M, Ye AY, Zheng W, Kong L. A guide RNA sequence design platform for the CRISPR/Cas9 system for model organism genomes. *Biomed Res Int* 2013;2013.
- [69] Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* 2006;1.
- [70] Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 2013;31:833–8.
- [71] Malzahn AA, Tang X, Lee K, Ren Q, Sretenovic S, Zhang Y, et al. Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and Arabidopsis. *BMC Biol* 2019;17:9.
- [72] Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NG, van den Broek M, et al. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 2015;15.
- [73] Marraffini LA. CRISPR-Cas immunity in prokaryotes. *Nature* 2015;526:55–61.
- [74] McKenna A, Shendure J. FlashFry: a fast and flexible tool for large-scale CRISPR target design. *BMC Biol* 2018;16.
- [75] Mendoza BJ, Trinh CT. Enhanced guide-RNA design and targeting analysis for precise CRISPR genome editing of single and consortia of industrially relevant and non-model organisms. *Bioinformatics* 2018;34:16–23.
- [76] Minkenberg B, Zhang J, Xie K, Yang Y. CRISPR-PLANT v2: an online resource for highly specific guide RNA spacers based on improved off-target analysis. *Plant Biotechnol J* 2019;17:5–8.
- [77] Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* 2014;42:W401–7.
- [78] Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Meth* 2015;12:982–8.
- [79] Park J, Bae S, Kim JS. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics* 2015;31:4014–6.
- [80] Park J, Childs L, Kim D, Hwang GH, Kim S, Kim ST, et al. Digenome-seq web tool for profiling CRISPR specificity. *Nat Methods* 2017;14:548–9.
- [81] Peng D, Tarleton R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. *Microb Genom* 2015;1.
- [82] Peng H, Zheng Y, Blumenstein M, Tao D, Li J. CRISPR/Cas9 cleavage efficiency regression through boosting algorithms and Markov sequence profiling. *Bioinformatics* 2018;34:3069–77.
- [83] Perez AR, Pritykin Y, Vidigal JA, Chhangawala S, Zampari L, Leslie CS, et al. GuideScan software for improved single and paired CRISPR guide RNA design. *Nat Biotechnol* 2017;35:347–9.
- [84] Pickar-Oliver A, Gersbach CA. The next generation of CRISPR-Cas technologies and applications. *Nat Rev Mol Cell Biol* 2019;20:490–507.
- [85] Pulido-Quetglas C, Aparicio-Prat E, Arnan C, Polidori T, Hermoso T, Palumbo E, et al. Scalable Design of Paired CRISPR Guide RNAs for Genomic Deletion. *PLoS Comput Biol* 2017;13.
- [86] Rahman MK, Rahman MS. CRISPRpred: a flexible and efficient tool for sgRNAs on-target activity prediction in CRISPR/Cas9 systems. *PLoS ONE* 2017;12.
- [87] Ren Q, Zhong Z, Wang Y, You Q, Li Q, Yuan M, et al. Bidirectional promoter-based crisper-cas9 systems for plant genome editing. *Front Plant Sci* 2019;10:1173.
- [88] Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 2014;32:347–55.
- [89] Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res* 2010;38:W462–8.
- [90] Shen MW, Arbab M, Hsu JY, Worstell D, Culbertson SJ, Krabbe O, et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* 2018;563:646–51.
- [91] Shibata M, Nishimasu H, Kodera N, Hirano S, Ando T, Uchihashi T, et al. Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat Commun* 8. 2017.
- [92] Singh R, Kucsu C, Quinlan A, Qi Y, Adli M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res* 2015;43.
- [93] Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: an intuitive, flexible and reliable crisper/cas9 target prediction tool. *PLoS One* 2015;10.
- [94] Strecker J, Jones S, Koopal B, Schmid-Burgk J, Zetsche B, Gao LY, et al. Engineering of CRISPR-Cas12b for human genome editing. *Nat Commun* 2019;10.
- [95] Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* 2014;111:9798–803.
- [96] Tang X, Liu GQ, Zhou JP, Ren QR, You Q, Tian L, et al. A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol* 2018;19.
- [97] Tang X, Lowder LG, Zhang T, Malzahn AA, Zheng X, Voytas DF, et al. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat Plants* 2017;3:17018.
- [98] Tang X, Ren Q, Yang L, Bao Y, Zhong Z, He Y, et al. Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. *Plant Biotechnol J* 2019;17:1431–45.
- [99] Tang X, Zheng X, Qi Y, Zhang D, Cheng Y, Tang A, et al. A single transcript CRISPR-Cas9 system for efficient genome editing in plants. *Mol Plant* 2016;9:1088–91.
- [100] Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005;435:646–51.
- [101] Uusi-Makela MIE, Barker HR, Bauerlein CA, Hakkinen T, Nykter M, Ramet M. Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (*Danio rerio*). *PLoS One* 2018;13.
- [102] van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, et al. DNA repair profiling reveals Nonrandom outcomes at Cas9-mediated breaks. *Mol Cell* 2016;63:633–46.
- [103] Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 2014;343:80–4.
- [104] Wilson LOW, O'Brien AR, Bauer DC. The current state and future of CRISPR-Cas9 gRNA design tools. *Front Pharmacol* 2018;9.
- [105] Wilson LOW, Reti D, O'Brien AR, Dunne RA, Bauer DC. High activity target-site identification using phenotypic independent CRISPR-Cas9 core functionality. *CRISPR J* 2018;1:182–90.
- [106] Wong N, Liu W, Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol* 2015;16:218.
- [107] Wu XB, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* 2014;32:670–6.
- [108] Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, Gao G, et al. CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* 2014;30:1180–2.
- [109] Xie K, Zhang J, Yang Y. Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Mol Plant* 2014;7:923–6.
- [110] Xie S, Shen B, Zhang C, Huang X, Zhang Y. sgRNACas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS One* 2014;9.
- [111] Xie X, Ma X, Zhu Q, Zeng D, Li G, Liu YG. CRISPR-GE: a convenient software toolkit for CRISPR-based genome editing. *Mol Plant* 2017;10:1246–9.
- [112] Xu H, Xiao TF, Chen CH, Li W, Meyer CA, Wu Q, et al. Sequence determinants of improved CRISPR sgRNA design. *Genome Res* 2015;25:1147–57.
- [113] Xu X, Duan D, Chen SJ. CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. *Sci Rep* 7. 2017.
- [114] Yan J, Chua G, Zhou C, Zhu C, Yang J, Zhang C, et al. Benchmarking CRISPR on-target sgRNA design. *Brief Bioinform* 2018;19:721–4.
- [115] Yang H, Wang HY, Shivalila CS, Cheng AW, Shi LY, Jaenisch R. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 2013;154:1370–9.
- [116] Yarrington RM, Verma S, Schwartz S, Trautman JK, Carroll D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc Natl Acad Sci U S A* 2018;115:9351–8.

- [117] You Q, Zhong Z, Ren Q, Hassan F, Zhang Y, Zhang T. CRISPRMatch: an automatic calculation and visualization tool for high-throughput CRISPR genome-editing data analysis. *Int J Biol Sci* 2018;14:858–62.
- [118] Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015;163:759–71.
- [119] Zhang D, Hurst T, Duan D, Chen SJ. Unified energetics analysis unravels SpCas9 cleavage activity for optimal gRNA design. *Proc Natl Acad Sci U S A* 2019;116:8693–8.
- [120] Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther-Nucl Acids* 2015;4.
- [121] Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, et al. Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol* 2013;161:20–7.
- [122] Zheng X, Yang S, Zhang D, Zhong Z, Tang X, Deng K, et al. Effective screen of CRISPR/Cas9-induced mutants in rice by single-strand conformation polymorphism. *Plant Cell Rep* 2016;35:1545–54.
- [123] Zhong Z, Sretenovic S, Ren Q, Yang L, Bao Y, Qi C, et al. Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. *Mol Plant* 2019;12:1027–36.
- [124] Zhong Z, Zhang Y, You Q, Tang X, Ren Q, Liu S, et al. Plant genome editing using FnCpf1 and LbCpf1 nucleases at redefined and altered pam sites. *Mol Plant* 2018;11:999–1002.
- [125] Zhou J, Deng K, Cheng Y, Zhong Z, Tian L, Tang X, et al. CRISPR-Cas9 based genome editing reveals new insights into MicroRNA function and regulation in rice. *Front Plant Sci* 2017;8:1598.
- [126] Zhou J, Xin X, He Y, Chen H, Li Q, Tang X, et al. Multiplex QTL editing of grain-related genes improves yield in elite rice varieties. *Plant Cell Rep* 2019;38:475–85.
- [127] Zhu H, Liang C. CRISPR-DT: designing gRNAs for the CRISPR-Cpf1 system with improved target efficiency and specificity. *Bioinformatics* 2019;35:2783–9.
- [128] Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B (Stat Methodol)* 2005;67:301–20.