# SNP-CRISPR: a web tool for SNP-specific genome editing

Chiao-Lin Chen[1], Jonathan Rodiger[1,2], Verena Chung[1,2], Raghuvir Viswanatha[1], Stephanie E. Mohr[1,2], Yanhui Hu[1,2], Norbert Perrimon[1,2,3]

[1]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115
[2]*Drosophila* RNAi Screening Center, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115
[3]Howard Hughes Medical Institute, 77 Avenue Louis Pasteur, Boston, MA 02115

ORCID IDs:
Norbert Perrimon, https://orcid.org/0000-0001-7542-472X
Stephanie E. Mohr, https://orcid.org/0000-0001-9639-7708
Yanhui Hu, https://orcid.org/0000-0003-1494-1402
Raghuvir Viswanatha, https://orcid.org/0000-0002-9457-6953

**Corresponding author:** Norbert Perrimon

**Key words:** genome editing, CRISPR, genome variant

## ABSTRACT

CRISPR-Cas9 is a powerful genome editing technology in which a single guide RNA (sgRNA) confers target site specificity to achieve Cas9-mediated genome editing. Numerous sgRNA design tools have been developed based on reference genomes for humans and model organisms. However, existing resources are not optimal as genetic mutations or single nucleotide polymorphisms (SNPs) within the targeting region affect the efficiency of CRISPR-based approaches by interfering with guide-target complementarity. To facilitate identification of sgRNAs (1) in non-reference genomes, (2) across varying genetic backgrounds, or (3) for specific targeting of SNP-containing alleles, for example, disease relevant mutations, we developed a web tool, SNP-CRISPR (https://www.flyrnai.org/tools/snp_crispr/). SNP-CRISPR can be used to design sgRNAs based on public variant data sets or user-identified variants. In addition, the tool computes efficiency and specificity scores for sgRNA designs targeting both the variant and the reference. Moreover, SNP-CRISPR provides the option to upload multiple SNPs and target single or multiple nearby base changes simultaneously with a single sgRNA design. Given these capabilities, SNP-CRISPR has a wide range of potential research applications in model systems and for design of sgRNAs for disease-associated variant correction.


## INTRODUCTION

The CRISPR-Cas9 system, a repurposed bacterial adaptive immune system, is a powerful programmable genome editing tool for research, including in eukaryotic systems, that also has potential for gene therapy (PICKAR-OLIVER and GERSBACH 2019). With this system, *Streptococcus pyogenes* Cas9 nuclease is directed to a target site or sites in the genome that have a unique 20 nt sequence followed by a 3 bp sequence conforming to *NGG* known as the protospacer adjacent motif (PAM). A double-strand break (DSB) induced by Cas9 nuclease recruits the cellular machinery, which can repair the break either through the error-prone non-homologous end-joining (NHEJ) pathway or through homology directed repair (HDR). NHEJ often results in insertions and/or deletions (indels), which can result in frameshift mutations. HDR allows researchers to introduce or 'knock in' specific DNA sequences, such as precise nucleotide changes or reporter cassettes.

In addition, catalytically dead forms of Cas9 have been fused with different effector proteins to manipulate DNA or gene expression (PICKAR-OLIVER and GERSBACH 2019). For example, to correct disease-causative point mutations, CRISPR-Cas9 mediated DNA base editing has been developed as a promising method to convert undesired spontaneous point mutations to the wild-type nucleotide (GAUDELLI *et al.* 2017; KOMOR *et al.* 2016; PICKAR-OLIVER and GERSBACH 2019). DNA base editing can be achieved by fusing a Cas9 nickase with a cytidine deaminase enzyme

and uracil glycosylate inhibitor to achieve a C->T (or G->A) substitution. Similarly, a transfer RNA adenosine deaminase is fused to a catalytically dead Cas9 to generate A->G (or T->C) conversion. Notably, unlike for knock-in, DNA editing-induced changes occur without a DSB and without the need for introduction of a donor template. Disease-relevant mutations in mammalian cells can be corrected with base editing strategies (DANDAGE *et al.* 2019). Prime Editing based on the fusion of Cas9 and reverse transcriptase, is another recently published technique that could add more precision and flexibility to CRISPR editing (ANZALONE *et al.* 2019). Thus, programmable editing of a target base in genomic DNA provides a potential therapy for genetic diseases that arise from point mutations.

Single-nucleotide polymorphisms (SNPs) can be defined as single-nucleotide differences from reference genomes. The targeting efficiency of Cas9 has been examined using data from genome-wide studies combined with machine learning (CHUAI *et al.* 2018; DOENCH *et al.* 2014; LISTGARTEN *et al.* 2018; NAJM *et al.* 2018; TYCKO *et al.* 2019). The position of specific nucleotides in the target sequences has been shown to affect targeting efficiency, which is the major determinant of CRISPR-Cas9 dependent genetic modification (DOENCH *et al.* 2014; HOUSDEN *et al.* 2015). Therefore, the presence of a SNP (or of an indel) can cause inefficient binding of the Cas9-sgRNA ribonucleoprotein (RNP) complex, resulting in inefficient genome editing.

Many rules for sgRNA design are generalizable and many web tools have been developed to predict sgRNA sequences for the human genome and genomes of numerous model organisms. There are two types of input that sgRNA design tools typically accept: (1) gene symbols or genome coordinates and (2) sequences. Resources that support the former typically precompute sgRNAs based on annotated reference genome information. Moreover, these sgRNA sequences are designed based on a single wild-type reference sequence without considering variants (e.g. CHOPCHOP, GuideScan; Table 1). With the second type of input, some tools BLAST user input sequence against a reference genome and correct any differences introduced by the user, thus making it impossible to design sgRNAs against a variant allele (this is the case for example for CRISPR-ERA and CRISPR-DT; Table 1). With others, it is possible to design a sgRNA to target variant allele (e.g. E-CRISPR, CRISPOR and CRISPRscan; Table 1). However, these tools require that the user retrieve the genomic sequences surrounding the variant and select designs that specifically target the variant region after the program sends back all the results. For a bench scientist, this is a time-consuming and error-prone process. For example, when the coding variant is near an exon-intron boundary, the user needs to retrieve the exon sequence as well as the intron sequence and enter these into the program. In addition, the user cannot do batch entries with most of the online tools that take sequence as an input (e.g. E-CRISPR, CRISPOR and CRISPRscan; Table 1). A few command line programs that take sequence as the input were developed for batch design; however, based on our experience,

these tools or specific features either do not work or are not easily configured by bench scientists without programing experience (Table 1). In addition, researchers might need features that are missing from current tools, such as an option to target SNPs either together or independently of one another when the SNPs are nearby one another. Moreover, the ability to compare sgRNA designs targeting the same locus in the wild-type and the variant allele in terms of efficiency and specificity would also be very useful when targeting a heterozygous variant.

To broaden the application of sgRNA design tools to better accommodate SNPs and small indels, we developed SNP-CRISPR. SNP-CRISPR is a web-based tool that accepts variant annotations as the input and uses rigorous off-target search algorithms to predict the specificity of each target site in the genome for wild-type and variant sequences. SNP-CRISPR offers customized options and allows users to easily and rapidly select optimal variant-specific CRISPR-Cas9 target sequences in genes from a variety of organisms.

## METHODS

### Pipeline development

The SNP-CRISPR pipeline environment is managed using the Conda package and environment management system (ANACONDA 2016). This allows for convenient reproduction of the necessary software dependencies and versions on different machines. The majority of the pipeline logic at SNP-CRISPR is written in Python using Biopython, with some Perl used for the BLAST and efficiency score analysis (COCK *et al.* 2009). Potential off-target loci are evaluated by performing a BLAST search of each design against the species reference genome. An off-target score is assigned based on both the number of hits found in the BLAST results and the number of mismatched nucleotides per off-target hit. Designs are also assigned an efficiency score that was computed using a position matrix; detailed information about the input dataset and algorithm can be found in (HOUSDEN *et al.* 2015). GNU Parallel is used to allow for parallelized computation of designs on different chromosomes and with different parameters for improved performance on multi-core systems (TANGE 2018). The full source code of the pipeline, including instructions for installation and use, is available at https://github.com/jrodiger/snp_crispr.

### Implementation of the web-based tool

The SNP-CRISPR web tool (https://www.flyrnai.org/tools/snp_crispr/) is located at the web site of the Drosophila RNAi Screening Center (DRSC). The back-end is written in PHP using the Symfony framework and the front end HTML pages take advantage of the Twig template

engine. The JQuery JavaScript library with the DataTables plugin is used for handling Ajax calls and displaying table views. The Bootstrap framework and some custom CSS is also used on the user interface. Hosting by Harvard Medical School Research Computing makes it possible to provide a web-facing user interface to run the SNP-CRISPR core pipeline on Harvard Medical School's "O2" high-performance computing cluster. When jobs are submitted from the website, the form parameters and uploaded input file path are passed to a bash script controlling the pipeline, which is then run as a cluster job. When the job is complete, an email is sent to the user with a URL that contains a unique ID used to retrieve the corresponding results.

**Availability**

SNP-CRISPR is available for online use without any restrictions at https://www.flyrnai.org/tools/snp_crispr.
The source code for the pipeline, including instructions for installation and use, is available at https://github.com/jrodiger/snp_crispr

**RESULTS**

**SNP-CRISPR web tool**

The web-based version of SNP-CRISPR provides the functionality of the design pipeline with an easy-to-use interface and interactive results view. Users can select up to 2,000 variants of interest in Variant Call Format (VCF) or a csv file in the provided format, and then upload this file on the SNP-CRISPR homepage. The acceptable variants include single nucleotide changes, small insertions and small deletions. The user then chooses the species and whether to create designs that target each input variant individually or to target all SNPs within each potential sgRNA sequence. When a user submits input, the web logic starts a job on the Harvard Medical School "O2" high-performance computing cluster, using the uploaded file and parameters as input for the pipeline. After the pipeline finishes running, an automated email is sent to the user with a link to a webpage at which the user can view and export results. For a couple of variants, the design pipeline usually takes up to a few minutes and with an input of 2,000 human SNP variants, it takes about half an hour for users to receive the results by email. The result page shows the wild-type and variant designs with corresponding scores in a tabular view that can be sorted by one or more columns. The output table also lists the genome targeting position of each sgRNA and the position of the variant within the sgRNA sequence relevant to the PAM sequence. The variant is shown in lowercase, which can be easily spotted by users. Using the checkboxes in the left-most column, users can opt to export all or only selected rows to an Excel or csv file. Currently, SNP-CRISPR supports reference genomes from human, mouse, rat, fly and zebrafish (Figure 1).

**Computation of potential variant-targeting sgRNAs**

Users are required to upload variant information in one of the supported formats including the genome coordinates, the sequence of the reference alleles and the sequence of the variant alleles. First, SNP-CRISPR validates the input reference sequences and will warn users if the submitted reference sequences does not match, which might reflect a different version of the genome assembly being used in the user input vs. SNP-CRISPR. After validation, SNP-CRISPR then re-constructs the template sequence, swapping the reference nucleotide with the variant nucleotide for SNPs, while inserting or deleting the corresponding fragment for indel type variants. Second, SNP-CRISPR computes potential variant-targeting sgRNAs based on availability of PAM sequences in the neighboring region since the presence of a PAM sequence (NGG or NAG) is one of the few requirements for binding. Third, sgRNA designs that contain four or more consecutive thymine residues, which can result in termination of RNA transcription by RNA polymerase III, are filtered out (Gao *et al.* 2018). Cas9 can have off-target activity across the genome and tolerance to mismatches shows significant variance depending on the position within the sgRNA (Fu *et al.* 2013; Hsu *et al.* 2013). Therefore, for each sgRNA design, SNP-CRISPR computes an efficiency score (Housden *et al.* 2015) and a specificity score calculated based on BLAST results against the reference genome. All possible sgRNAs are provided to the user along with specificity and efficiency scores, without further filtering; filtering options are available for custom applications based on user needs (Figure 2).

To facilitate identification of the best variant-specific sgRNAs, we provide information about both sgRNAs targeting specific variants and sgRNAs targeting the reference sequence in the same region. The efficiency score and an off-target score are provided, and the positions of relevant SNPs or indels in the sgRNA are included so that users can select the most suitable sgRNA or filter out less optimal ones.

The web tool supports up to 2,000 variants per batch while the command line version has no limit with the number of variants and can be used for any annotated genome. The command line version also provides better performance on large inputs when run multi-threaded. For example, a multi-threaded test run was able to process over 1,000 human SNPs per minute on Harvard Medical School's “O2” high-performance computing cluster. We pre-computed sgRNA designs (NGG-PAM) for all clinically associated SNPs annotated at the Ensembl genome browser (ftp://ftp.ensembl.org/pub/release-97/variation/gvf/homo_sapiens/homo_sapiens_clinically_associated.gvf.gz) using the command line version of the pipeline and the designs can be found at https://github.com/jrodiger/snp_crispr/tree/master/results.

## CONCLUSION

SNP-CRISPR is a unique web tool that designs sgRNAs targeting specific SNPs or indels. SNP-CRISPR is user-friendly and provides all possible CRISPR-Cas9 target sites in a given genomic region with required parameters, allowing users to select an optimal sgRNA. SNP-CRISPR provides not only efficiency scores but also off-target information for sgRNAs targeting sequences with and without SNPs and/or indels of interest in the same genomic region. SNP-CRISPR supports the human reference genome and genomes from major model organisms; namely, mouse, rat, fly and zebrafish. Conveniently, SNP-CRISPR displays the positions of variant nucleotides in each sgRNA region as part of the design output. Moreover, SNP-CRIPSR accepts up to 2,000 inputs per batch for design of large-scale experiments at the website. The command line version has no limit as to the number of variants and can be used for any genome that has been properly annotated. Altogether, SNP-CRISPR improves the ability of researchers to edit SNP or indel-containing loci by facilitating the design of sgRNAs that target specific variants. As such, SNP-CRISPR provides a valuable new resource to the genome editing technology field.

More and more variant data has become available in recent years, and much current research focuses on the biological impact of variants (Amberger and Hamosh 2017; Bragin *et al.* 2014; Landrum *et al.* 2014; Song *et al.* 2016), motivating us to develop a variant-centered tool. For instance, a CRISPR/Cas9-based targeting approach has been used to specifically correct heterozygous missense mutations associated with dominantly inherited conditions by including the mutated base in the sgRNA sequence (Courtney *et al.* 2016). CRISPR/Cas9-based therapeutic approaches show great promise for permanent correction of genetic disorders in somatic cells. In addition, to facilitate direct research in gene therapy of human diseases, SNP-CRISPR will be valuable for modeling human disease using model organisms. With a vast and growing amount of sequences from different strains of model organisms such as *Drosophila melanogaster*, millions of novel sequence variants have been identified (Huang *et al.* 2014; Wang *et al.* 2015). However, the biological significance of most of these sequence variants is still unclear. By facilitating design of sgRNAs targeting variant-specific alleles, including at a large scale, SNP-CRISPR makes it more feasible to study these variants systematically.

## ACKNOWLEDGEMENTS

Relevant grant support includes NIH NIGMS R01 GM067761 and P41 GM132087. In addition, C.C. is supported by R21 ES025615, and S.E.M. is supported in part by the Dana

Farber/Harvard Cancer Center, which is supported in part by NIH NCI Cancer Center Support Grant P30 CA006516. N.P. is an investigator of Howard Hughes Medical Institute.

## FIGURE LEGENDS

**Figure 1. Features of the SNP-CRISPR user interface (UI)**. Users select the species of interest, enter an email address, upload variant information including the genome coordinates and sequence changes, choose to target nearby variants individually or together, and then submit the job. Usually within half an hour, an email is sent automatically to the user with a link to a results page that displays the designs for wild type as well as mutant alleles, side by side with calculated scores. The mutant base(s) are shown in lower case and the wild type sequence in upper case.

**Figure 2. SNP-CRISPR sgRNA design pipeline**. Graphic display of the major steps of sgRNA design (blue), and input files and output files for the command line version of the pipeline (red).

# REFERENCES


AMBERGER, J. S., and A. HAMOSH, 2017 Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. Curr Protoc Bioinformatics **58:** 1 2 1-1 2 12.

ANACONDA, 2016 Anaconda Software Distribution. Computer software. Vers. 2-2.4.0. Web. <https://anaconda.com>.

ANZALONE, A. V., P. B. RANDOLPH, J. R. DAVIS, A. A. SOUSA, L. W. KOBLAN *et al.*, 2019 Search-and-replace genome editing without double-strand breaks or donor DNA. Nature.

BRAGIN, E., E. A. CHATZIMICHALI, C. F. WRIGHT, M. E. HURLES, H. V. FIRTH *et al.*, 2014 DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res **42:** D993-D1000.

CHUAI, G., H. MA, J. YAN, M. CHEN, N. HONG *et al.*, 2018 DeepCRISPR: optimized CRISPR guide RNA design by deep learning. Genome Biol **19:** 80.

COCK, P., T. ANTAO, J. CHANG, B. CHAPMAN, C. COX *et al.*, 2009 Biopython: freely available Python tools

COURTNEY, D. G., J. E. MOORE, S. D. ATKINSON, E. MAURIZI, E. H. ALLEN *et al.*, 2016 CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting. Gene Ther **23:** 108-112.

DANDAGE, R., P. C. DESPRES, N. YACHIE and C. R. LANDRY, 2019 beditor: A Computational Workflow for Designing Libraries of Guide RNAs for CRISPR-Mediated Base Editing. Genetics **212:** 377-385.

DOENCH, J. G., E. HARTENIAN, D. B. GRAHAM, Z. TOTHOVA, M. HEGDE *et al.*, 2014 Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol **32:** 1262-1267.

FU, Y., J. A. FODEN, C. KHAYTER, M. L. MAEDER, D. REYON *et al.*, 2013 High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol **31:** 822-826.

GAO, Z., E. HERRERA-CARRILLO and B. BERKHOUT, 2018 Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. Mol Ther Nucleic Acids **10:** 36-44.

GAUDELLI, N. M., A. C. KOMOR, H. A. REES, M. S. PACKER, A. H. BADRAN *et al.*, 2017 Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature **551:** 464-471.

HAEUSSLER, M., K. SCHONIG, H. ECKERT, A. ESCHSTRUTH, J. MIANNE *et al.*, 2016 Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol **17:** 148.

HEIGWER, F., G. KERR and M. BOUTROS, 2014 E-CRISP: fast CRISPR target site identification. Nat Methods **11:** 122-123.

HOUSDEN, B. E., A. J. VALVEZAN, C. KELLEY, R. SOPKO, Y. HU *et al.*, 2015 Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. Sci Signal **8:** rs9.

HSU, P. D., D. A. SCOTT, J. A. WEINSTEIN, F. A. RAN, S. KONERMANN *et al.*, 2013 DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol **31:** 827-832.

HUANG, W., A. MASSOURAS, Y. INOUE, J. PEIFFER, M. RAMIA *et al.*, 2014 Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res **24:** 1193-1208.

KEOUGH, K. C., S. LYALINA, M. P. OLVERA, S. WHALEN, B. R. CONKLIN *et al.*, 2019 AlleleAnalyzer: a tool for personalized and allele-specific sgRNA design. Genome Biol **20:** 167.

KOMOR, A. C., Y. B. KIM, M. S. PACKER, J. A. ZURIS and D. R. LIU, 2016 Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature **533:** 420-424.

LABUN, K., T. G. MONTAGUE, M. KRAUSE, Y. N. TORRES CLEUREN, H. TJELDNES *et al.*, 2019 CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res **47:** W171-W174.

LANDRUM, M. J., J. M. LEE, G. R. RILEY, W. JANG, W. S. RUBINSTEIN *et al.*, 2014 ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res **42:** D980-985.

LISTGARTEN, J., M. WEINSTEIN, B. P. KLEINSTIVER, A. A. SOUSA, J. K. JOUNG *et al.*, 2018 Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. Nat Biomed Eng **2:** 38-47.

LIU, H., Z. WEI, A. DOMINGUEZ, Y. LI, X. WANG *et al.*, 2015 CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. Bioinformatics **31:** 3676-3678.

MORENO-MATEOS, M. A., C. E. VEJNAR, J. D. BEAUDOIN, J. P. FERNANDEZ, E. K. MIS *et al.*, 2015 CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods **12:** 982-988.

NAITO, Y., K. HINO, H. BONO and K. UI-TEI, 2015 CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics **31:** 1120-1123.

NAJM, F. J., C. STRAND, K. F. DONOVAN, M. HEGDE, K. R. SANSON *et al.*, 2018 Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol **36:** 179-189.

OLIVEROS, J. C., M. FRANCH, D. TABAS-MADRID, D. SAN-LEON, L. MONTOLIU *et al.*, 2016 Breaking-Cas-interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes. Nucleic Acids Res **44:** W267-271.

PARK, J., S. BAE and J. S. KIM, 2015 Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics **31:** 4014-4016.

PEREZ, A. R., Y. PRITYKIN, J. A. VIDIGAL, S. CHHANGAWALA, L. ZAMPARO *et al.*, 2017 GuideScan software for improved single and paired CRISPR guide RNA design. Nat Biotechnol **35:** 347-349.

PICKAR-OLIVER, A., and C. A. GERSBACH, 2019 The next generation of CRISPR-Cas technologies and applications. Nat Rev Mol Cell Biol **20:** 490-507.

PLIATSIKA, V., and I. RIGOUTSOS, 2015 "Off-Spotter": very fast and exhaustive enumeration of genomic lookalikes for designing CRISPR/Cas guide RNAs. Biol Direct **10:** 4.

RABINOWITZ, R., R. DARNELL and D. OFFEN, 2019 CrisPam – a tool for designing gRNA sequences to specifically target a variant allele using CRISPR. Cytotherapy **21:** e6.

SANSON, K. R., R. E. HANNA, M. HEGDE, K. F. DONOVAN, C. STRAND *et al.*, 2018 Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat Commun **9:** 5416.

SONG, W., S. A. GARDNER, H. HOVHANNISYAN, A. NATALIZIO, K. S. WEYMOUTH *et al.*, 2016 Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. Genet Med **18:** 850-854.

STEMMER, M., T. THUMBERGER, M. DEL SOL KEYER, J. WITTBRODT and J. L. MATEO, 2015 CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS One **10:** e0124633.

TANGE, O., 2018 GNU Parallel 2018, ISBN 9781387509881, DOI https://doi.org/10.5281/zenodo.1146014.

TYCKO, J., M. WAINBERG, G. K. MARINOV, O. URSU, G. T. HESS *et al.*, 2019 Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. Nat Commun **10:** 4063.

WANG, F., L. JIANG, Y. CHEN, N. A. HAELTERMAN, H. J. BELLEN *et al.*, 2015 FlyVar: a database for genetic variation in Drosophila melanogaster. Database (Oxford) **2015**.

ZHU, H., and C. LIANG, 2019 CRISPR-DT: designing gRNAs for the CRISPR-Cpf1 system with improved target efficiency and specificity. Bioinformatics **35:** 2783-2789.

ZHU, L. J., B. R. HOLMES, N. ARONIN and M. H. BRODSKY, 2014 CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PLoS One **9:** e108424.

**Table 1.** A Survey of CRISPR design tools

| Tool | Type | Web Input | Web batch entry? | Consider variant? | Variant specific designs compared to SNP-CRISPR | URL/Reference |
|---|---|---|---|---|---|---|
| CHOPCHOP | web-based | Gene<br>Genome coordinates | No | No | None | chopchop.rc.fas.harvard.edu<br>(LABUN *et al.* 2019) |
| GuideScan | web-based | Gene<br>Genome coordinates | Yes | No | None | guidescan.com<br>(PEREZ *et al.* 2017) |
| DRSC find CRISPR Tool | web-based | Gene<br>Genome coordinates | No | No | None | flyrnai.org/crispr<br>(HOUSDEN *et al.* 2015) |
| E-CRISPR | web-based | Gene<br>Sequence | No | Possible | Fewer[2] | e-crisp.org<br>(HEIGWER *et al.* 2014) |
| CRISPOR | web-based | Gene<br>Sequence | No (seq)<br>Yes (gene) | Possible | Same[2] | crispor.tefor.net<br>(HAEUSSLER *et al.* 2016) |
| CRISPRscan | web-based[1] | Gene<br>Sequence | No | Possible | Fewer[2] | crisprscan.org<br>(MORENO-MATEOS *et al.* 2015) |
| CRISPRdirect | web-based | Gene<br>Sequence | No | Possible | Same[2] | crispr.dbcls.jp<br>(NAITO *et al.* 2015) |
| CRISPR-ERA | web-based | Gene, Sequence<br>Genome coordinates | No | No | None | crispr-era.stanford.edu<br>(LIU *et al.* 2015) |
| CRISPR-DT | web-based | sequence | No | No | None | bioinfolab.miamioh.edu/CRISPR-DT<br>(ZHU and LIANG 2019) |
| DeepCRISPR | web-based | sequence | No | Undetermined[3] | | deepcrispr.net<br>(CHUAI *et al.* 2018) |
| GT-Scan | web-based | sequence | No | Possible | Same[2] | gt-scan.csiro.au/<br>(OLIVEROS *et al.* 2016) |
| GPP sgRNA Designer | web-based | Gene<br>Sequence | Yes | Possible | Same[2] | portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design<br>(SANSON *et al.* 2018) |
| CCTop | web-based | sequence | Yes | Possible | Same[2] | crispr.cos.uni-heidelberg.de<br>(STEMMER *et al.* 2015) |
| Cas-Designer | web-based | sequence | Yes | Possible | Same[2] | rgenome.net/cas-designer<br>(PARK *et al.* 2015) |
| Breaking-Cas | web-based[1] | sequence | Yes | Possible | Same[2] | bioinfogp.cnb.csic.es/tools/breakingcas<br>(OLIVEROS *et al.* 2016) |
| Off-Spotter | web-based[1] | sequence | No | Possible | Same[2] | cm.jefferson.edu/Off-Spotter<br>(PLIATSIKA and RIGOUTSOS 2015) |
| Protospacer | GUI (OSX only) | NA | NA | No | None | protospacer.com |
| CrisPam | command line | NA | NA | Undetermined[3] | | github.com/ristllin/CrisPam<br>(RABINOWITZ *et al.* 2019) |
| CRISPRseek | command line | NA | NA | Possible | Same[2] | bioconductor.org/packages/release/bioc/html/CRISPRseek.html<br>(ZHU *et al.* 2014) |
| AlleleAnalyzer | command line | NA | NA | Yes | Same | github.com/keoughkath/AlleleAnalyzer<br>(KEOUGH *et al.* 2019) |
| SNP-CRISPR | web-based[1] | Variants (eg. VCF file) | Yes | Yes | NA | flyrnai.org/tools/snp_crispr |

Note: We limited our survey to CRISPR design tools that do not require registration or user login.

1. A command line version is also available
2. Users need to provide flanking sequence and filter out irrelevant designs
3. Test was attempted but results were not obtained

HOME ABOUT CONTACT

# SNP-targeted CRISPR Designs

**Are you interested in targeting SNP or INDEL variants in your experiment? Look no further!** We can help you find CRISPR guide de
(both -NGG and -NAG) in Fly, Human, Mouse, Zebrafish, and Rat genomes. Each guide will be given an efficiency and off-target score bas
our algorithms so that you can plan accordingly. Targeting multiple variants within the same guide is also available (select the checkbox)!

Select species →

**Species**

✓ Fly
Human
Mouse
Zebrafish
Rat

Enter email →

**Email**

**List of variants**

Upload a .txt or .csv file of **2000 or less variants** (sample file here) -- listed alleles can either be uppercase or lowercase.

Browse... No file selected. ← Upload variant info

Option to group nearby variants → ☐ Create guides where all SNPs are targeted within the 23-mer

Create guides!

| gene_symbol | chromosome | position | strand | reference | variant | group(optional) |
|---|---|---|---|---|---|---|
| NA | 2L | 8471348 | - | C | G | snp1 |
| NA | 3L | 14275690 | + | T | G | snp2 |
| NA | 3L | 14275690 | + | T | C | snp3 |
| NA | 2L | 7310396 | + | CAACTGA | C | indel1 |
| NA | 2L | 7310843 | + | C | CCCGTT | indel2 |

HOME ABOUT CONTACT

# RESULTS

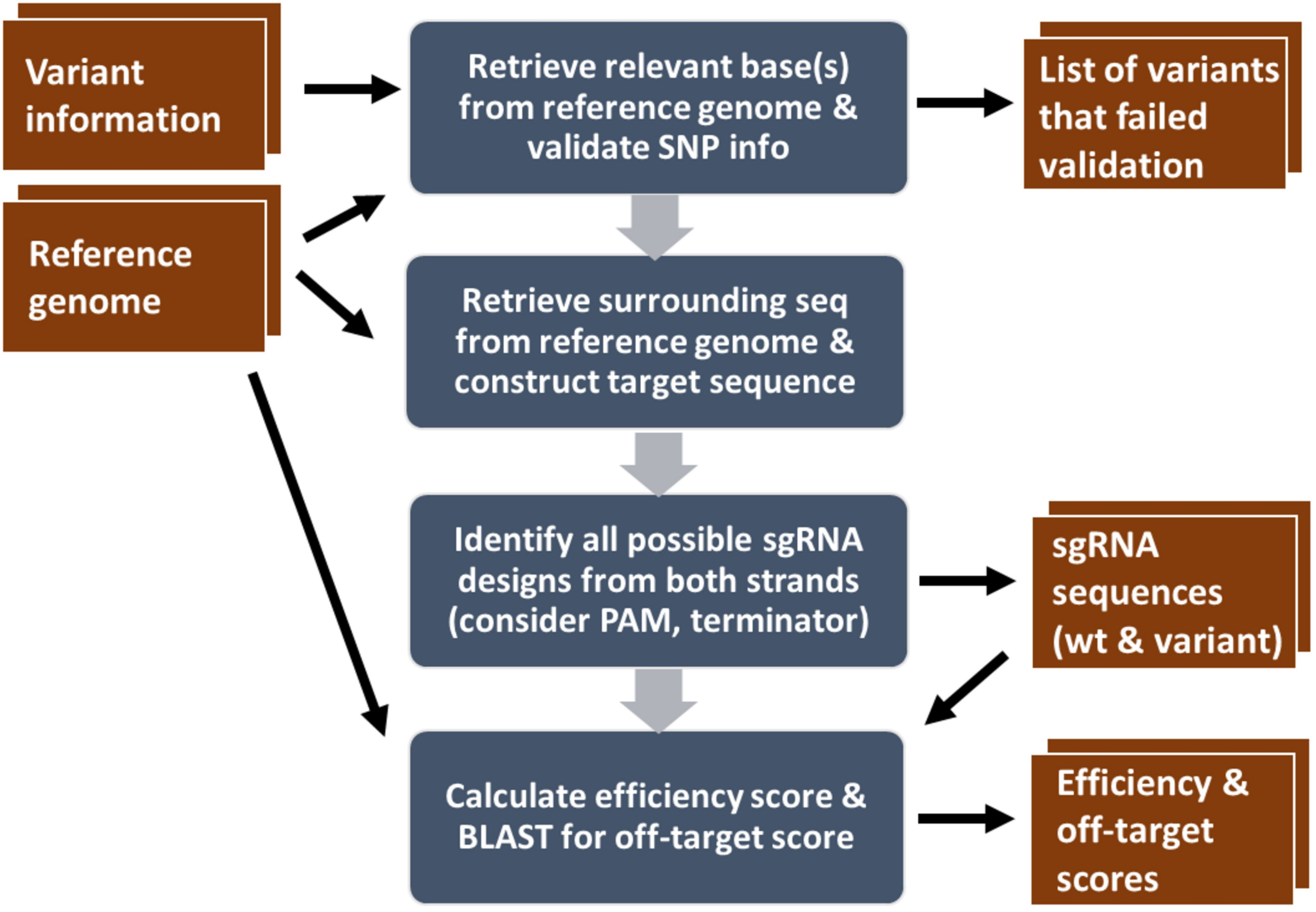
**Housden Efficiency Score**
These scores were computed using a position matrix. Detailed information about the input dataset and the algorithm can be found in Housden et al. Sci Signal. 2015 https://www.ncbi.nlm.nih.gov/pubmed/26350902
Scores range from 1.47-12.32 (higher is better, > 5 recommended)

**Off Target Score**
Calculated based on sgRNA sequence blast results.
Scores range from 0-5441.73 (lower is better, < 1 recommended)

Copy CSV Excel Search/filter:

| Gene | Chr | Variant | Wild-type Design | Wild-type Off Target Score | Wild-type Efficiency Score | Variant Design | Variant Off Target Score | Variant Efficiency Score | Start | End | Strand | Variant Position | Distance to PAM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NA (indel1) | 2L | TCAGTTG>G | gttgCAGGATCTCAGGAAACAGG | 1.04 | 5.27 | GTTgCAGGATCTCAGGAAACAGG | 1.04 | 5.27 | 7310399 | 7310377 | - | 7310396 | 16 |
| NA (indel1) | 2L | TCAGTTG>G | AGTTtcagttgCAGGATCTCAGG | 0.11 | 5.82 | AGTTTCAGTTgCAGGATCTCAGG | 0.11 | 5.82 | 7310406 | 7310384 | - | 7310396 | 9 |
| NA (indel1) | 2L | TCAGTTG>G | GTAGTTTCAGTTtcagttgCAGG | 9.85 | 3.60 | CATCAGGTAGTTTCAGTTgCAGG | 0.23 | 4.50 | 7310414 | 7310392 | - | 7310396 | 1 |
| NA (indel2) | 2L | G>AACGGG | AGAgAACCCAGTCAAGAGACAGG | 0.11 | 4.20 | cgggAACCCAGTCAAGAGACAGG | 1.7 | 4.39 | 7310846 | 7310824 | - | 7310843 | 16 |
| NA (indel2) | 2L | G>AACGGG | AAAACGACAAAAAAAAAAAAGAg | 0 | 3.59 | GACAAAAAAAAAAAAGAaacggg | 0 | 3.55 | 7310865 | 7310843 | - | 7310843 | -3 |
| NA (indel2) | 2L | G>AACGGG | AAAAACGACAAAAAAAAAAAAGA | 0 | 4.14 | CGACAAAAAAAAAAAAGAaacgg | 0 | 3.90 | 7310866 | 7310844 | - | 7310843 | -3 |
| NA (snp1) | 2L | G>C | AACAGCTAGgATCCAAGATCAGG | 0 | 5.56 | AACAGCTAGcATCCAAGATCAGG | 1.02 | 7.56 | 8471339 | 8471361 | + | 8471348 | 10 |
| NA (snp1) | 2L | C>G | cCTAGCTGTTTAAGAGCCACAGG | 0.01 | 4.80 | gCTAGCTGTTTAAGAGCCACAGG | 1.02 | 4.60 | 8471348 | 8471326 | - | 8471348 | 19 |
| NA (snp2) | 3L | T>G | CGTGCAACTGAAAACGCCtCTGG | 0.24 | 5.70 | CGTGCAACTGAAAACGCCgCTGG | 1.14 | 5.54 | 14275672 | 14275694 | + | 14275690 | 1 |
| NA (snp2) | 3L | T>G | GTGCAACTGAAAACGCCtCTGGG | 0.15 | 5.82 | GTGCAACTGAAAACGCCgCTGGG | 1.21 | 5.72 | 14275673 | 14275695 | + | 14275690 | 2 |
| NA (snp2) | 3L | T>G | ACTGAAAACGCCtCTGGGAATGG | 1.14 | 3.90 | ACTGAAAACGCCgCTGGGAATGG | 2.24 | 4.80 | 14275678 | 14275700 | + | 14275690 | 7 |
| NA (snp2) | 3L | T>G | CTGAAAACGCCtCTGGGAATGGG | 0.41 | 9.34 | CTGAAAACGCCgCTGGGAATGGG | 2.15 | 8.95 | 14275679 | 14275701 | + | 14275690 | 8 |
| NA (snp3) | 3L | T>C | CGTGCAACTGAAAACGCCtCTGG | 0.24 | 5.70 | CGTGCAACTGAAAACGCCcCTGG | 2.22 | 5.63 | 14275672 | 14275694 | + | 14275690 | 1 |
| NA (snp3) | 3L | T>C | GTGCAACTGAAAACGCCtCTGGG | 0.15 | 5.82 | GTGCAACTGAAAACGCCcCTGGG | 2.05 | 5.61 | 14275673 | 14275695 | + | 14275690 | 2 |
| NA (snp3) | 3L | T>C | ACTGAAAACGCCtCTGGGAATGG | 1.14 | 3.90 | ACTGAAAACGCCcCTGGGAATGG | 1.42 | 4.03 | 14275678 | 14275700 | + | 14275690 | 7 |
| NA (snp3) | 3L | T>C | CTGAAAACGCCtCTGGGAATGGG | 0.41 | 9.34 | CTGAAAACGCCcCTGGGAATGGG | 2.14 | 9.14 | 14275679 | 14275701 | + | 14275690 | 8 |

Showing 1 to 16 of 16 entries Previous 1 Next


Variant information
Reference genome
Retrieve relevant base(s) from reference genome & validate SNP info
List of variants that failed validation
Retrieve surrounding seq from reference genome & construct target sequence
Identify all possible sgRNA designs from both strands (consider PAM, terminator)
sgRNA sequences (wt & variant)
Calculate efficiency score & BLAST for off-target score
Efficiency & off-target scores