

1    **SNP-CRISPR: a web tool for SNP-specific genome editing**

2

3    Chiao-Lin Chen<sup>1</sup>, Jonathan Rodiger<sup>1,2</sup>, Verena Chung<sup>1,2</sup>, Raghuvir Viswanatha<sup>1</sup>, Stephanie E.  
4    Mohr<sup>1,2</sup>, Yanhui Hu<sup>1,2</sup>, Norbert Perrimon<sup>1,2,3</sup>

5

6    <sup>1</sup>Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115

7    <sup>2</sup>*Drosophila* RNAi Screening Center, Harvard Medical School, 77 Avenue Louis Pasteur, Boston,  
8    MA 02115

9    <sup>3</sup>Howard Hughes Medical Institute, 77 Avenue Louis Pasteur, Boston, MA 02115

10

11

12    ORCID IDs:

13    Norbert Perrimon, <https://orcid.org/0000-0001-7542-472X>

14    Stephanie E. Mohr, <https://orcid.org/0000-0001-9639-7708>

15    Yanhui Hu, <https://orcid.org/0000-0003-1494-1402>

16    Raghuvir Viswanatha, <https://orcid.org/0000-0002-9457-6953>

17

18    **Corresponding author:** Norbert Perrimon

19

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

21

22

23

24 **ABSTRACT**

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

40  
41 **INTRODUCTION**

42  
43 The CRISPR-Cas9 system, a repurposed bacterial adaptive immune system, is a powerful  
44 programmable genome editing tool for research, including in eukaryotic systems, that also has  
45 potential for gene therapy (PICKAR-OLIVER and GERSBACH 2019). With this system, *Streptococcus*  
46 *pyogenes* Cas9 nuclease is directed to a target site or sites in the genome that have a unique 20  
47 nt sequence followed by a 3 bp sequence conforming to NGG known as the protospacer  
48 adjacent motif (PAM). A double-strand break (DSB) induced by Cas9 nuclease recruits the  
49 cellular machinery, which can repair the break either through the error-prone non-homologous  
50 end-joining (NHEJ) pathway or through homology directed repair (HDR). NHEJ often results in  
51 insertions and/or deletions (indels), which can result in frameshift mutations. HDR allows  
52 researchers to introduce or ‘knock in’ specific DNA sequences, such as precise nucleotide  
53 changes or reporter cassettes.

54  
55 In addition, catalytically dead forms of Cas9 have been fused with different effector  
56 proteins to manipulate DNA or gene expression (PICKAR-OLIVER and GERSBACH 2019). For example,  
57 to correct disease-causative point mutations, CRISPR-Cas9 mediated DNA base editing has been  
58 developed as a promising method to convert undesired spontaneous point mutations to the  
59 wild-type nucleotide (GAUDELLI *et al.* 2017; KOMOR *et al.* 2016; PICKAR-OLIVER and GERSBACH 2019).  
60 DNA base editing can be achieved by fusing a Cas9 nickase with a cytidine deaminase enzyme

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

70

71 Single-nucleotide polymorphisms (SNPs) can be defined as single-nucleotide differences  
72 from reference genomes. The targeting efficiency of Cas9 has been examined using data from  
73 genome-wide studies combined with machine learning (CHUAI *et al.* 2018; DOENCH *et al.* 2014;  
74 LISTGARTEN *et al.* 2018; NAJM *et al.* 2018; TYCKO *et al.* 2019). The position of specific nucleotides in  
75 the target sequences has been shown to affect targeting efficiency, which is the major  
76 determinant of CRISPR-Cas9 dependent genetic modification (DOENCH *et al.* 2014; HOUSDEN *et al.*  
77 2015). Therefore, the presence of a SNP (or of an indel) can cause inefficient binding of the  
78 Cas9-sgRNA ribonucleoprotein (RNP) complex, resulting in inefficient genome editing.

79

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

99 these tools or specific features either do not work or are not easily configured by bench  
100 scientists without programing experience (Table 1). In addition, researchers might need  
101 features that are missing from current tools, such as an option to target SNPs either together or  
102 independently of one another when the SNPs are nearby one another. Moreover, the ability to  
103 compare sgRNA designs targeting the same locus in the wild-type and the variant allele in terms  
104 of efficiency and specificity would also be very useful when targeting a heterozygous variant.  
105

106 To broaden the application of sgRNA design tools to better accommodate SNPs and  
107 small indels, we developed SNP-CRISPR. SNP-CRISPR is a web-based tool that accepts variant  
108 annotations as the input and uses rigorous off-target search algorithms to predict the specificity  
109 of each target site in the genome for wild-type and variant sequences. SNP-CRISPR offers  
110 customized options and allows users to easily and rapidly select optimal variant-specific CRISPR-  
111 Cas9 target sequences in genes from a variety of organisms.  
112  
113

## 114 METHODS

### 115 Pipeline development

116 The SNP-CRISPR pipeline environment is managed using the Conda package and  
117 environment management system (ANACONDA 2016). This allows for convenient reproduction of  
118 the necessary software dependencies and versions on different machines. The majority of the  
119 pipeline logic at SNP-CRISPR is written in Python using Biopython, with some Perl used for the  
120 BLAST and efficiency score analysis (Cock *et al.* 2009). Potential off-target loci are evaluated by  
121 performing a BLAST search of each design against the species reference genome. An off-target  
122 score is assigned based on both the number of hits found in the BLAST results and the number  
123 of mismatched nucleotides per off-target hit. Designs are also assigned an efficiency score that  
124 was computed using a position matrix; detailed information about the input dataset and  
125 algorithm can be found in (HOUSDEN *et al.* 2015). GNU Parallel is used to allow for parallelized  
126 computation of designs on different chromosomes and with different parameters for improved  
127 performance on multi-core systems (TANGE 2018). The full source code of the pipeline, including  
128 instructions for installation and use, is available at [https://github.com/jrodiger/snp\\_crispr](https://github.com/jrodiger/snp_crispr).  
129  
130

### 131 Implementation of the web-based tool

132 The SNP-CRISPR web tool ([https://www.flyrnai.org/tools/snp\\_crispr/](https://www.flyrnai.org/tools/snp_crispr/)) is located at the  
133 web site of the Drosophila RNAi Screening Center (DRSC). The back-end is written in PHP using  
134 the Symfony framework and the front end HTML pages take advantage of the Twig template  
135  
136

137 engine. The JQuery JavaScript library with the DataTables plugin is used for handling Ajax calls  
138 and displaying table views. The Bootstrap framework and some custom CSS is also used on the  
139 user interface. Hosting by Harvard Medical School Research Computing makes it possible to  
140 provide a web-facing user interface to run the SNP-CRISPR core pipeline on Harvard Medical  
141 School's "O2" high-performance computing cluster. When jobs are submitted from the website,  
142 the form parameters and uploaded input file path are passed to a bash script controlling the  
143 pipeline, which is then run as a cluster job. When the job is complete, an email is sent to the  
144 user with a URL that contains a unique ID used to retrieve the corresponding results.  
145

#### 146 **Availability**

147 SNP-CRISPR is available for online use without any restrictions at

148 [https://www.flyrnai.org/tools/snp\\_crispr](https://www.flyrnai.org/tools/snp_crispr).

149 The source code for the pipeline, including instructions for installation and use, is available at

150 [https://github.com/jrodiger/snp\\_crispr](https://github.com/jrodiger/snp_crispr)

151

## 152 **RESULTS**

153

### 154 **SNP-CRISPR web tool**

155

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

175

176 **Computation of potential variant-targeting sgRNAs**

177

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

196

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

202

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

213

214

215 **CONCLUSION**

216

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

231

232 More and more variant data has become available in recent years, and much current  
233 research focuses on the biological impact of variants (AMBERGER and HAMOSH 2017; BRAGIN *et al.*  
234 2014; LANDRUM *et al.* 2014; SONG *et al.* 2016), motivating us to develop a variant-centered tool.  
235 For instance, a CRISPR/Cas9-based targeting approach has been used to specifically correct  
236 heterozygous missense mutations associated with dominantly inherited conditions by including  
237 the mutated base in the sgRNA sequence (COURTNEY *et al.* 2016). CRISPR/Cas9-based  
238 therapeutic approaches show great promise for permanent correction of genetic disorders in  
239 somatic cells. In addition, to facilitate direct research in gene therapy of human diseases, SNP-  
240 CRISPR will be valuable for modeling human disease using model organisms. With a vast and  
241 growing amount of sequences from different strains of model organisms such as *Drosophila*  
242 *melanogaster*, millions of novel sequence variants have been identified (HUANG *et al.* 2014;  
243 WANG *et al.* 2015). However, the biological significance of most of these sequence variants is  
244 still unclear. By facilitating design of sgRNAs targeting variant-specific alleles, including at a  
245 large scale, SNP-CRISPR makes it more feasible to study these variants systematically.

246

247 **ACKNOWLEDGEMENTS**

248

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

251 Farber/Harvard Cancer Center, which is supported in part by NIH NCI Cancer Center Support  
252 Grant P30 CA006516. N.P. is an investigator of Howard Hughes Medical Institute.  
253  
254

255 **FIGURE LEGENDS**

256

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

264

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

267

268

269

270 REFERENCES

- 271
- 272 AMBERGER, J. S., and A. HAMOSH, 2017 Searching Online Mendelian Inheritance in Man (OMIM): A  
273 Knowledgebase of Human Genes and Genetic Phenotypes. *Curr Protoc Bioinformatics*  
274 **58:** 1 2 1-1 2 12.
- 275 ANACONDA, 2016 Anaconda Software Distribution. Computer software. Vers. 2-2.4.0. Web.  
276 <<https://anaconda.com>>.
- 277 ANZALONE, A. V., P. B. RANDOLPH, J. R. DAVIS, A. A. SOUSA, L. W. KOBLAN *et al.*, 2019 Search-and-  
278 replace genome editing without double-strand breaks or donor DNA. *Nature*.
- 279 BRAGIN, E., E. A. CHATZIMICHALI, C. F. WRIGHT, M. E. HURLES, H. V. FIRTH *et al.*, 2014 DECIPHER:  
280 database for the interpretation of phenotype-linked plausibly pathogenic sequence and  
281 copy-number variation. *Nucleic Acids Res* **42:** D993-D1000.
- 282 CHUAI, G., H. MA, J. YAN, M. CHEN, N. HONG *et al.*, 2018 DeepCRISPR: optimized CRISPR guide RNA  
283 design by deep learning. *Genome Biol* **19:** 80.
- 284 COCK, P., T. ANTAKO, J. CHANG, B. CHAPMAN, C. COX *et al.*, 2009 Biopython: freely available Python  
285 tools
- 286 COURTNEY, D. G., J. E. MOORE, S. D. ATKINSON, E. MAURIZI, E. H. ALLEN *et al.*, 2016 CRISPR/Cas9 DNA  
287 cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific  
288 targeting. *Gene Ther* **23:** 108-112.
- 289 DANDAGE, R., P. C. DESPRES, N. YACHIE and C. R. LANDRY, 2019 beditor: A Computational Workflow  
290 for Designing Libraries of Guide RNAs for CRISPR-Mediated Base Editing. *Genetics* **212:**  
291 377-385.
- 292 DOENCH, J. G., E. HARTENIAN, D. B. GRAHAM, Z. TOTHOVA, M. HEGDE *et al.*, 2014 Rational design of  
293 highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32:**  
294 1262-1267.
- 295 FU, Y., J. A. FODEN, C. KHAYTER, M. L. MAEDER, D. REYON *et al.*, 2013 High-frequency off-target  
296 mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31:** 822-  
297 826.
- 298 GAO, Z., E. HERRERA-CARRILLO and B. BERKHOUT, 2018 Delineation of the Exact Transcription  
299 Termination Signal for Type 3 Polymerase III. *Mol Ther Nucleic Acids* **10:** 36-44.
- 300 GAUDELLI, N. M., A. C. KOMOR, H. A. REES, M. S. PACKER, A. H. BADRAN *et al.*, 2017 Programmable  
301 base editing of A\*T to G\*C in genomic DNA without DNA cleavage. *Nature* **551:** 464-471.
- 302 HAEUSSLER, M., K. SCHONIG, H. ECKERT, A. ESCHSTRUTH, J. MIANNE *et al.*, 2016 Evaluation of off-target  
303 and on-target scoring algorithms and integration into the guide RNA selection tool  
304 CRISPOR. *Genome Biol* **17:** 148.
- 305 HEIGWER, F., G. KERR and M. BOUTROS, 2014 E-CRISP: fast CRISPR target site identification. *Nat  
306 Methods* **11:** 122-123.
- 307 HOUSDEN, B. E., A. J. VALVEZAN, C. KELLEY, R. SOPKO, Y. HU *et al.*, 2015 Identification of potential drug  
308 targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based  
309 knockouts with RNAi. *Sci Signal* **8:** rs9.
- 310 HSU, P. D., D. A. SCOTT, J. A. WEINSTEIN, F. A. RAN, S. KONERMANN *et al.*, 2013 DNA targeting  
311 specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31:** 827-832.

- 312 HUANG, W., A. MASSOURAS, Y. INOUE, J. PEIFFER, M. RAMIA *et al.*, 2014 Natural variation in genome  
313 architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines.  
314 *Genome Res* **24**: 1193-1208.
- 315 KEOUGH, K. C., S. LYALINA, M. P. OLVERA, S. WHALEN, B. R. CONKLIN *et al.*, 2019 AlleleAnalyzer: a tool  
316 for personalized and allele-specific sgRNA design. *Genome Biol* **20**: 167.
- 317 KOMOR, A. C., Y. B. KIM, M. S. PACKER, J. A. ZURIS and D. R. LIU, 2016 Programmable editing of a  
318 target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**: 420-  
319 424.
- 320 LABUN, K., T. G. MONTAGUE, M. KRAUSE, Y. N. TORRES CLEUREN, H. TJELDNES *et al.*, 2019 CHOPCHOP v3:  
321 expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**: W171-  
322 W174.
- 323 LANDRUM, M. J., J. M. LEE, G. R. RILEY, W. JANG, W. S. RUBINSTEIN *et al.*, 2014 ClinVar: public archive  
324 of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**:  
325 D980-985.
- 326 LISTGARTEN, J., M. WEINSTEIN, B. P. KLEINSTIVER, A. A. SOUSA, J. K. JOUNG *et al.*, 2018 Prediction of off-  
327 target activities for the end-to-end design of CRISPR guide RNAs. *Nat Biomed Eng* **2**: 38-  
328 47.
- 329 LIU, H., Z. WEI, A. DOMINGUEZ, Y. LI, X. WANG *et al.*, 2015 CRISPR-ERA: a comprehensive design tool  
330 for CRISPR-mediated gene editing, repression and activation. *Bioinformatics* **31**: 3676-  
331 3678.
- 332 MORENO-MATEOS, M. A., C. E. VEJNAR, J. D. BEAUDOIN, J. P. FERNANDEZ, E. K. MIS *et al.*, 2015  
333 CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat  
334 Methods* **12**: 982-988.
- 335 NAITO, Y., K. HINO, H. BONO and K. UI-TEI, 2015 CRISPRdirect: software for designing CRISPR/Cas  
336 guide RNA with reduced off-target sites. *Bioinformatics* **31**: 1120-1123.
- 337 NAJM, F. J., C. STRAND, K. F. DONOVAN, M. HEGDE, K. R. SANSON *et al.*, 2018 Orthologous CRISPR-Cas9  
338 enzymes for combinatorial genetic screens. *Nat Biotechnol* **36**: 179-189.
- 339 OLIVEROS, J. C., M. FRANCH, D. TABAS-MADRID, D. SAN-LEON, L. MONTOLIU *et al.*, 2016 Breaking-Cas-  
340 interactive design of guide RNAs for CRISPR-Cas experiments for ENSEMBL genomes.  
341 *Nucleic Acids Res* **44**: W267-271.
- 342 PARK, J., S. BAE and J. S. KIM, 2015 Cas-Designer: a web-based tool for choice of CRISPR-Cas9  
343 target sites. *Bioinformatics* **31**: 4014-4016.
- 344 PEREZ, A. R., Y. PRITYKIN, J. A. VIDIGAL, S. CHHANGAWALA, L. ZAMPARO *et al.*, 2017 GuideScan software  
345 for improved single and paired CRISPR guide RNA design. *Nat Biotechnol* **35**: 347-349.
- 346 PICKAR-OLIVER, A., and C. A. GERSBACH, 2019 The next generation of CRISPR-Cas technologies and  
347 applications. *Nat Rev Mol Cell Biol* **20**: 490-507.
- 348 PLIATSIKA, V., and I. RIGOUTSOS, 2015 "Off-Spotter": very fast and exhaustive enumeration of  
349 genomic lookalikes for designing CRISPR/Cas guide RNAs. *Biol Direct* **10**: 4.
- 350 RABINOWITZ, R., R. DARNELL and D. OFFEN, 2019 CrisPam – a tool for designing gRNA sequences to  
351 specifically target a variant allele using CRISPR. *Cyotherapy* **21**: e6.
- 352 SANSON, K. R., R. E. HANNA, M. HEGDE, K. F. DONOVAN, C. STRAND *et al.*, 2018 Optimized libraries for  
353 CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**: 5416.

- 354 SONG, W., S. A. GARDNER, H. HOVHANNISYAN, A. NATALIZIO, K. S. WEYMOUTH *et al.*, 2016 Exploring the  
355 landscape of pathogenic genetic variation in the ExAC population database: insights of  
356 relevance to variant classification. *Genet Med* **18**: 850-854.
- 357 STEMMER, M., T. THUMBERGER, M. DEL SOL KEYER, J. WITTBRODT and J. L. MATEO, 2015 CCTop: An  
358 Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One* **10**:  
359 e0124633.
- 360 TANGE, O., 2018 GNU Parallel 2018, ISBN 9781387509881, DOI  
361 <https://doi.org/10.5281/zenodo.1146014>.
- 362 TYCKO, J., M. WAINBERG, G. K. MARINOV, O. URSU, G. T. HESS *et al.*, 2019 Mitigation of off-target  
363 toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat Commun* **10**:  
364 4063.
- 365 WANG, F., L. JIANG, Y. CHEN, N. A. HAELTERMAN, H. J. BELLEN *et al.*, 2015 FlyVar: a database for  
366 genetic variation in *Drosophila melanogaster*. *Database (Oxford)* **2015**.
- 367 ZHU, H., and C. LIANG, 2019 CRISPR-DT: designing gRNAs for the CRISPR-Cpf1 system with  
368 improved target efficiency and specificity. *Bioinformatics* **35**: 2783-2789.
- 369 ZHU, L. J., B. R. HOLMES, N. ARONIN and M. H. BRODSKY, 2014 CRISPRseek: a bioconductor package  
370 to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS*  
371 *One* **9**: e108424.
- 372
- 373

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

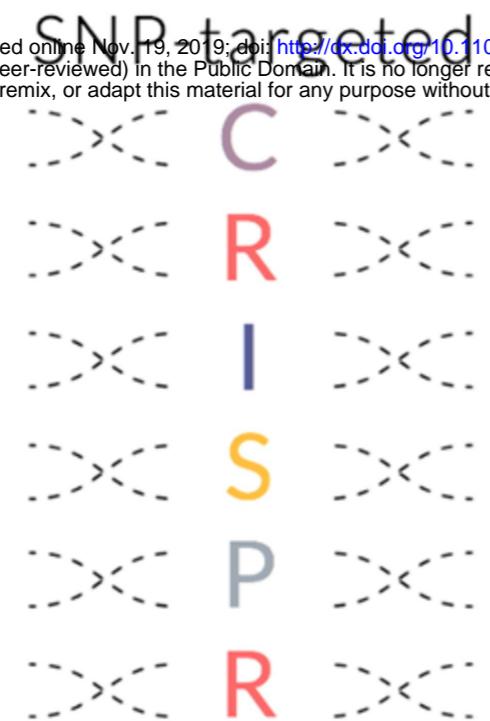
Note: We limited our survey to CRISPR design tools that do not require registration or user login.

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

bioRxiv preprint first posted online Nov. 19, 2019; doi: <https://dx.doi.org/10.1101/847277>. The copyright holder has placed this preprint (which was not peer-reviewed) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.



## Designs



Are you interested in targeting SNP or INDEL variants in your experiment? Look no further! We can help you find CRISPR guide designs (both -NGG and -NAG) in Fly, Human, Mouse, Zebrafish, and Rat genomes. Each guide will be given an efficiency and off-target score based on 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

Option to group nearby variants →

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.

No file selected.  
 Create guides where all SNPs are targeted within the 23-mer

Create guides!

Upload variant info

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

## 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)

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
<input type="checkbox"/>	NA (indel1)	2L	TCAGTTG>G	gttgCAGGATCTCAGGAACAGG	1.04	5.27	GTTgCAGGATCTCAGGAACAGG	1.04	5.27	7310399	7310377	-	7310396	16
<input type="checkbox"/>	NA (indel1)	2L	TCAGTTG>G	AGTTtcagtgcAGGATCTCAGG	0.11	5.82	AGTTTCAGTTgcAGGATCTCAGG	0.11	5.82	7310406	7310384	-	7310396	9
<input type="checkbox"/>	NA (indel1)	2L	TCAGTTG>G	GTAGTTTCAGTTtcagtgcAGG	9.85	3.60	CATCAGGTAGTTTCAGTTgcAGG	0.23	4.50	7310414	7310392	-	7310396	1
<input type="checkbox"/>	NA (indel2)	2L	G>AACGGG	AGAgAACCCAGTCAAGAGACAGG	0.11	4.20	cgggAACCCAGTCAAGAGACAGG	1.7	4.39	7310846	7310824	-	7310843	16
<input type="checkbox"/>	NA (indel2)	2L	G>AACGGG	AAAACGACAAAAAAAAAAAGAg	0	3.59	GACAAAAAAAGAaacggg	0	3.55	7310865	7310843	-	7310843	-3
<input type="checkbox"/>	NA (indel2)	2L	G>AACGGG	AAAACGACAAAAAAAAAAAGA	0	4.14	CGACAAAAAAAGAaacgg	0	3.90	7310866	7310844	-	7310843	-3
<input type="checkbox"/>	NA (snp1)	2L	G>C	AACAGCTAGgATCCAAGATCAGG	0	5.56	AACAGCTAGcATCCAAGATCAGG	1.02	7.56	8471339	8471361	+	8471348	10
<input type="checkbox"/>	NA (snp1)	2L	C>G	cCTAGCTGTTAACGCCACAGG	0.01	4.80	gCTAGCTGTTAACGCCACAGG	1.02	4.60	8471348	8471326	-	8471348	19
<input type="checkbox"/>	NA (snp2)	3L	T>G	CGTGCAACTGAAAACGCCtCTGG	0.24	5.70	CGTGCAACTGAAAACGCCgCTGG	1.14	5.54	14275672	14275694	+	14275690	1
<input type="checkbox"/>	NA (snp2)	3L	T>G	GTGCAACTGAAAACGCCtCTGG	0.15	5.82	GTGCAACTGAAAACGCCgCTGG	1.21	5.72	14275673	14275695	+	14275690	2
<input type="checkbox"/>	NA (snp2)	3L	T>G	ACTGAAAACGCCtCTGGGAATGG	1.14	3.90	ACTGAAAACGCCgCTGGGAATGG	2.24	4.80	14275678	14275700	+	14275690	7
<input type="checkbox"/>	NA (snp2)	3L	T>G	CTGAAAACGCCtCTGGGAATGG	0.41	9.34	CTGAAAACGCCgCTGGGAATGG	2.15	8.95	14275679	14275701	+	14275690	8
<input type="checkbox"/>	NA (snp3)	3L	T>C	CGTGCAACTGAAAACGCCtCTGG	0.24	5.70	CGTGCAACTGAAAACGCCcCTGG	2.22	5.63	14275672	14275694	+	14275690	1
<input type="checkbox"/>	NA (snp3)	3L	T>C	GTGCAACTGAAAACGCCtCTGG	0.15	5.82	GTGCAACTGAAAACGCCcCTGG	2.05	5.61	14275673	14275695	+	14275690	2
<input type="checkbox"/>	NA (snp3)	3L	T>C	ACTGAAAACGCCtCTGGGAATGG	1.14	3.90	ACTGAAAACGCCcCTGGGAATGG	1.42	4.03	14275678	14275700	+	14275690	7
<input type="checkbox"/>	NA (snp3)	3L	T>C	CTGAAAACGCCtCTGGGAATGG	0.41	9.34	CTGAAAACGCCcCTGGGAATGG	2.14	9.14	14275679	14275701	+	14275690	8

Showing 1 to 16 of 16 entries

Previous 1 Next

