

CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing

Kornel Labun¹, Tessa G. Montague², Maximilian Krause¹, Yamila N. Torres Cleuren¹, Håkon Tjeldnes¹ and Eivind Valen^{1,3,*}

¹Computational Biology Unit, Department of Informatics, University of Bergen, 5008 Bergen, Norway, ²Mortimer B. Zuckerman Mind Brain Behavior Institute, Department of Neuroscience, Columbia University, New York, NY 10027, USA and ³Sars International Centre for Marine Molecular Biology, University of Bergen, 5008 Bergen, Norway

Received February 26, 2019; Revised April 16, 2019; Editorial Decision April 26, 2019; Accepted May 06, 2019

ABSTRACT

The CRISPR–Cas system is a powerful genome editing tool that functions in a diverse array of organisms and cell types. The technology was initially developed to induce targeted mutations in DNA, but CRISPR–Cas has now been adapted to target nucleic acids for a range of purposes. CHOPCHOP is a web tool for identifying CRISPR–Cas single guide RNA (sgRNA) targets. In this major update of CHOPCHOP, we expand our toolbox beyond knockouts. We introduce functionality for targeting RNA with Cas13, which includes support for alternative transcript isoforms and RNA accessibility predictions. We incorporate new DNA targeting modes, including CRISPR activation/repression, targeted enrichment of loci for long-read sequencing, and prediction of Cas9 repair outcomes. Finally, we expand our results page visualization to reveal alternative isoforms and downstream ATG sites, which will aid users in avoiding the expression of truncated proteins. The CHOPCHOP web tool now supports over 200 genomes and we have released a command-line script for running larger jobs and handling unsupported genomes. CHOPCHOP v3 can be found at <https://chopchop.cbu.uib.no>

INTRODUCTION

The use of CRISPR–Cas is now ubiquitous in modern molecular biology. First introduced as a tool for introducing repair-induced mutations in the genome, the emergence of catalytically dead or fused versions of the effector proteins has transformed CRISPR–Cas into a general purpose tool for targeting. For instance, CRISPR–Cas has been used to introduce new sequences into the genome (1–3), activate (4,5) or repress (6) transcription, as an enrichment tool for

sequencing (7,8), for targeted hypermutation (9), as a diagnostic tool (10), to perform whole-organism lineage tracing (11), to target RNA molecules for destruction (12) or editing (13), and to track transcripts in live cells (14).

All CRISPR–Cas applications use a sgRNA to direct the CRISPR effector (Cas) protein to its target. In theory, CRISPR–Cas targeting only requires complementarity between the sgRNA and its nucleic acid target, but a number of studies have shown that efficient targeting follows more complex rules (15–21). For instance, the position of specific nucleotides in the target sequence, the accessibility of the target site, and the sequence of its flanking regions can all influence efficiency. The targeting efficiencies of Cas9 and Cas12a/Cpf1 have been measured in large-scale studies and combined with machine learning-based methods to optimize cutting (15,21).

There are other factors that can influence or prevent the generation of a null mutant. For instance, introducing a frameshift mutation too close to the start codon can permit translation initiation at a downstream ATG, leading to unintentional protein production. In other cases, targeting exons that are only present in a subset of isoforms can prevent null mutation generation. Finally, CRISPR–Cas gene editing can produce confounding phenotypes due to transcription adaptation or genetic compensation. For instance, a recent study showed that degradation of mutant mRNAs could result in the upregulation of related genes (22). In these situations, deleting the promoter can be a more robust method to produce knockouts.

Many CRISPR–Cas applications require the generation of a frameshift mutation to disrupt gene function, which requires a DNA repair event in which the number of inserted or deleted nucleotides is not a multiple of three. Surprisingly, it has been shown that double-strand break (DSB) repairs are not random: Cas9-induced DSBs using the same sgRNA often give rise to the same mutations (23). Recently, this has been incorporated into models that predict whether DSB repairs will give rise to a frameshifting mutation (24).

*To whom correspondence should be addressed. Tel: +47 55584074; Email: eivind.valen@gmail.com

In summary, the CRISPR–Cas system has been adapted for a wide selection of uses, and numerous factors influence each of these modes, necessitating the existence of intuitive software for target selection. This new update of the CHOPCHOP web tool incorporates new CRISPR–Cas targeting modes and predicts frameshift mutation frequency in an improved, user-friendly interface.

IMPROVEMENTS IN THE NEW RELEASE

Like previous versions, CHOPCHOP handles input from (i) gene and transcript identifiers, (ii) genomic coordinates and (iii) pasted sequences, and provides results in a number of output formats. The interface is simple (Figure 1) and requires the user to make four selections in the default mode: (a) target gene/isoform/region, (b) organism, (c) CRISPR effector (e.g. Cas9, CasX or Cas13), and (d) purpose (e.g. knockout, knockdown, repression). Advanced users can adjust the default settings by clicking the ‘Options’ button (Figure 1).

While CHOPCHOP queries typically run within a few seconds, heavy traffic can cause congestion during peak hours. We have therefore introduced a queuing system that ensures all users are prioritized. The new system retains results and calculations for up to 48 h after the initial query, permitting quick access (via caching) if an identical search is made later on and sharing results with collaborators. To increase speed for larger queries, CHOPCHOP also supports pre-filtering of sgRNA targets by (i) GC content (with a default of 10–90%) and (ii) the existence of self-complementarity within the sgRNA. This can greatly reduce computation time.

Several adjustments have been made to improve the results page visualizations. Notably, in-frame downstream ATGs are now colored in the isoform visualization (Figure 1) to help users avoid downstream translation initiation. In addition, the results table can now be sorted using any criteria.

In addition to these improvements, the latest CHOPCHOP introduces the following new major features.

Targeting the transcriptome

The most widespread use of CRISPR–Cas is to introduce mutations into DNA. However, CRISPR–Cas systems have now been engineered to target RNA. For instance, CRISPR–Cas13 has been harnessed for transcript knockdown (25), live-cell transcript imaging (14) and RNA base editing (13).

CHOPCHOP now permits CRISPR–Cas13 targeting, and implements this functionality by searching for off-targets across the complete transcriptome rather than the genome. An important aspect of RNA targeting is to avoid regions of high structure that can reduce the accessibility of Cas13 (12,14,26). CHOPCHOP calculates RNA accessibility using RNAfold from the ViennaRNA package (27) according to published recommendations (14). Briefly, accessibility is calculated in windows of 70 nucleotides, obtaining the probability that a given nucleotide position in the transcript is unpaired. For each target, we take the mean probability of structure across each position targeted by the sgRNA (14).

Similar to protospacer adjacent motifs (PAMs) in DNA-targeting modes, we support any 5′ or 3′ protospacer flanking sequences (PFS) for RNA targeting. The Cas13a PFS - an H at the 3′ end of the target - is the default. As with the PAM, the PFS should be present in the sequence of the DNA or RNA target, but not in the sgRNA. CHOPCHOP provides the appropriate sgRNA sequence, ensuring that the user does not include the PFS when ordering the sgRNA oligonucleotide.

New modes for targeting the genome

CHOPCHOP v3 expands the number of modes for DNA targeting. These include: (i) Nanopore enrichment mode. Targeted sequencing is a method used to attain high quality sequencing reads in a specific region of interest. PCR-based methods for enriching genomic regions have some limitations, for instance the maximum length of the region that can be enriched. By contrast, CRISPR–Cas provides a powerful method to excise genomic regions prior to amplification and sequencing with Oxford Nanopore Technologies (ONT) (7,8). The nanopore enrichment mode in CHOPCHOP allows users to identify pairs of gRNAs flanking large regions (up to 40kb) by excluding low-efficiency guides. The mode filters all sgRNAs with predicted self-complementarity, as this can have inhibitory effects on global Cas9 activity (8,28) and ONT personal communication); (ii) Knock-in mode. This identifies the same sgRNAs as the knockout mode, but designs homology arms up to 2 kb, which, based on recent studies (1–3) can be used for targeted insertions; (iii) Activation/repression modes. These modes are designed for use with Cas9 fusion proteins with the intention of activating or repressing a gene. Specifically, the modes target the promoter region and its flanking sites according to the guidelines specified in (29–31) in order to bring the activating/repressing domain into close proximity with the transcription start site.

For the Cas9 knockout mode, we now create a prediction of DSB repair outcomes (24). The model estimates the probability that a given sgRNA will result in a frameshift mutation. In addition, we have added efficiency scores for Cas12a/Cpf1 (21) and updated the ‘Doench’ efficiency score to the newest version (15), which is now the default scoring metric for Cas9 genomic targeting.

Expansion of genomes and targeting

CHOPCHOP now supports over 200 genomes and includes gene annotations for genomic targets, as well as three transcriptomes for RNA knockdown (human, mouse and zebrafish). While previous versions of CHOPCHOP required the selection of a specific isoform for targeting, this new version allows the user to target the entire gene. In its default ‘intersection’ mode, CHOPCHOP v3 only searches for sgRNAs present in every isoform (Figure 1). This mode can be disabled by selecting the ‘union’ mode, which will display all sgRNAs in all transcripts, as well as a column indicating whether the sgRNA targets a constitutive exon.

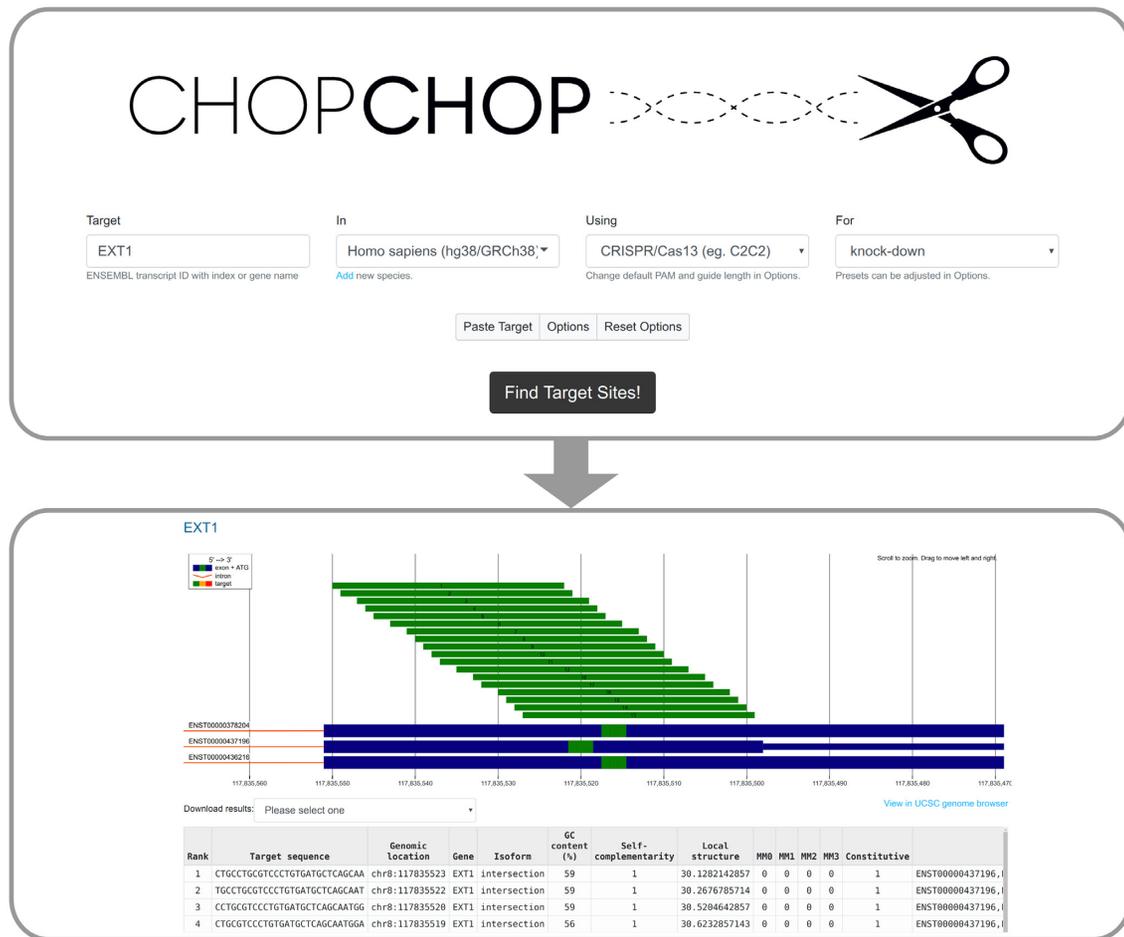


Figure 1. The workflow of CHOPCHOP when targeting RNA for knock-down. The CHOPCHOP homepage (upper box) require four types of input: (i) target, (ii) species, (iii) CRISPR effector and (iv) the purpose of the experiment. Default options will be adequate for most users, but advanced options can be revealed using the ‘Options’ button. The results of the search (lower box) are displayed along with all isoforms of the target gene. The target color indicates the quality of each sgRNA or nickase pair (green [best] to red [worst]). Below the graphic representation an interactive table allows for exploring each guide in greater detail.

Command-line version

In addition to the web interface, we also provide the code for the command-line version of CHOPCHOP, which can be run locally and is suited for larger queries or screens. This tool includes all of the functionality of the web interface in addition to extra functionality for larger experiments, such as the ability to design control sgRNAs that do not match any sequence in the genome. The command-line version of CHOPCHOP is compatible with ampliCan, a tool for sequencing-based assessment of mutations (32).

DISCUSSION AND FUTURE DEVELOPMENTS

Since its inception, the CRISPR field has undergone constant and rapid innovation, requiring the parallel development of bioinformatic tools that accommodate the new findings and technologies. In just a few years, CRISPR–Cas has become a powerful targeting tool for silencing and activating both DNA and RNA in a range of contexts, each of which requires the application of specific rules. The latest release of CHOPCHOP addresses this challenge by adding new functionalities that reflect the ever-expanding CRISPR

toolbox. So far, we have accommodated the favorite species of over 200 research groups, and as we continue to improve the functionality of CHOPCHOP, we will continue to accommodate new transcriptomes and genomes.

In conclusion, this major update expands the CHOPCHOP toolbox, retaining its position as one of the most easy-to-use, versatile CRISPR–Cas targeting tools available.

DATA AVAILABILITY

The CHOPCHOP (version 3) server is available at <https://chopchop.cbu.uib.no>; the python code for local installation is available at <https://bitbucket.org/valenlab/chopchop>.

ACKNOWLEDGEMENTS

We would like to thank James Graham from Oxford Nanopore Technologies for his recommendations on the long-read enrichment mode.

FUNDING

Bergen Research Foundation (to E.V.); University of Bergen core funding (to K.L.); Norwegian Research Council [250049 to E.V.]; National Defense Science and Engineering Graduate (NDSEG) Fellowship (to T.G.M.). Funding for open access charge: Bergens Forskningsstiftelse (Bergen Research Foundation).

Conflict of interest statement. None declared.

REFERENCES

- Nakamae, K., Nishimura, Y., Takenaga, M., Nakade, S., Sakamoto, N., Ide, H., Sakuma, T. and Yamamoto, T. (2017) Establishment of expanded and streamlined pipeline of PITCh knock-in - a web-based design tool for MMEJ-mediated gene knock-in, PITCh designer, and the variations of PITCh, PITCh-TG and PITCh-KIKO. *Bioengineered*, **8**, 302–308.
- Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.-I.T. and Yamamoto, T. (2016) MMEJ-assisted gene knock-in using TALENs and CRISPR–Cas9 with the PITCh systems. *Nat. Protoc.*, **11**, 118–133.
- Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z. *et al.* (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, **540**, 144–149.
- Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H. and Joung, J.K. (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods*, **10**, 977–979.
- Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.L., Glass, K.A., Ousterout, D.G., Leong, K.W. *et al.* (2013) RNA-guided gene activation by CRISPR–Cas9-based transcription factors. *Nat. Methods*, **10**, 973–976.
- Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442–451.
- Gabrieli, T., Sharim, H., Fridman, D., Arbib, N., Michaeli, Y. and Ebenstein, Y. (2018) Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res.*, **46**, e87.
- Gilpatrick, T., Lee, I., Graham, J.E., Raimondeau, E., Bowen, R., Heron, A., Sedlazeck, F.J. and Timp, W. (2019) Targeted Nanopore Sequencing with Cas9 for studies of methylation, structural variants and mutations. bioRxiv doi: <https://doi.org/10.1101/604173>, 11 April 2019, preprint: not peer reviewed.
- Halperin, S.O., Tou, C.J., Wong, E.B., Modavi, C., Schaffer, D.V. and Dueber, J.E. (2018) CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature*, **560**, 248–252.
- Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W., Essletzbichler, P., Dy, A.J., Joung, J., Verdine, V., Donghia, N., Daringer, N.M., Freije, C.A. *et al.* (2017) Nucleic acid detection with CRISPR–Cas13a/C2c2. *Science*, **356**, 438–442.
- McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F. and Shendure, J. (2016) Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*, **353**, aaf7907.
- Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L. *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, **353**, aaf5573.
- Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J. and Zhang, F. (2017) RNA editing with CRISPR–Cas13. *Science*, **358**, 1019–1027.
- Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev, A. *et al.* (2017) RNA targeting with CRISPR–Cas13. *Nature*, **550**, 280–284.
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.*, **34**, 184–191.
- Chari, R., Mali, P., Moosburner, M. and Church, G.M. (2015) Unraveling CRISPR–Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods*, **12**, 823–826.
- Xu, H., Xiao, T., Chen, C.-H., Li, W., Meyer, C.A., Wu, Q., Wu, D., Cong, L., Zhang, F., Liu, J.S. *et al.* (2015) Sequence determinants of improved CRISPR sgRNA design. *Genome Res.*, **25**, 1147–1157.
- Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K. and Giraldez, A.J. (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR–Cas9 targeting in vivo. *Nat. Methods*, **12**, 982–988.
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–832.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
- Kim, H.K., Min, S., Song, M., Jung, S., Choi, J.W., Kim, Y., Lee, S., Yoon, S. and Kim, H.H. (2018) Deep learning improves prediction of CRISPR–Cpf1 guide RNA activity. *Nat. Biotechnol.*, **36**, 239–241.
- El-Brolosy, M., Rossi, A., Kontarakis, Z., Kuenne, C., Guenther, S., Fukuda, N., Takacs, C., Lai, S.-L., Fukuda, R., Gerri, C. *et al.* (2019) Genetic compensation is triggered by mutant mRNA degradation. *Nature*, **568**, 193–197.
- Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Akhmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C. *et al.* (2014) Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One*, **9**, e98186.
- Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K. and Sherwood, R.I. (2018) Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*, **563**, 646–651.
- Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N. and Hsu, P.D. (2018) Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell*, **173**, 665–676.
- Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S. *et al.* (2017) Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Mol. Cell*, **65**, 618–630.
- Lorenz, R., Bernhart, S.H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
- Thyme, S.B., Akhmetova, L., Montague, T.G., Valen, E. and Schier, A.F. (2016) Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat. Commun.*, **7**, 11750.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
- Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B. and Jaenisch, R. (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.*, **23**, 1163–1171.
- La Russa, M.F. and Qi, L.S. (2015) The new state of the art: Cas9 for gene activation and repression. *Mol. Cell Biol.*, **35**, 3800–3809.
- Labun, K., Guo, X., Chavez, A., Church, G., Gagnon, J.A. and Valen, E. (2018) Accurate analysis of genuine CRISPR editing events with ampliCan. *Genome Res.*, **29**, 843–847.