casmini-tool: a comprehensive database for efficient and specific guide RNA design using dCasMINI

Spencer Lopp[*]
Epicrispr Biotechnologies
spencer.lopp@epic-bio.com

Robin W. Yeo[*]
Epicrispr Biotechnologies
robin.yeo@epic-bio.com

Nishant Jha
Nitro Bio
me@nishantjha.org

Xiao Yang
Epicrispr Biotechnologies
xiao.yang@epic-bio.com

Gabriella Alvarez
Epicrispr Biotechnologies

Courtney Klappenbach
Epicrispr Biotechnologies

Daniel O. Hart
Epicrispr Biotechnologies
dan.hart@epic-bio.com

Lei S. Qi
Department of Bioengineering,
Sarafan ChEM-H
Stanford University
slqi@stanford.edu

Timothy P. Daley[†]
Epicrispr Biotechnologies
tim.daley@epic-bio.com

Abstract

The dCasMINI protein is a hyper-compact, nuclease-inactivated CRISPR-Cas system engineered for transcriptional modulation and epigenetic editing [Xu et al., 2021]. The small size of dCasMini (529 amino acids), less than half the size of comparable Cas9 molecules, makes it ideal for AAV-based therapies which are frequently limited by AAV's small cargo capacity. Unlike Cas9 or Cas12a, there is no available computational tools for designing CasMINI guides. To facilitate and accelerate the development of dCasMINI-based applications, we synthesized knowledge regarding CasMINI guide design and built a website to assist researchers in designing optimal guides for dCasMINI-based experiments for transcriptional inhibition (CRISPRi) and activation (CRISPRa), which covers 99.7% of genes for CRISPRi and 99.9% of genes for CRISPRa. We experimentally characterized the importance of each nucleotide position on the guide RNA for determining its activity. Based on this information, , our tool offers more sensitively mapping off-targets and provides information about alignment mismatches in the spacer seed region, which we have experimentally determined to be critical for true binding events. The tool is freely available at casmini-tool.com.

The CRISPR-Cas revolution is upon us. Advances in CRISPR-Cas9 based therapeutics have resulted in transformational therapies for β-thalassemia [Frangoul et al., 2020], sickle cell disease [De Dreuzy et al., 2019], B-cell lymphoma [McGuirk et al., 2022], non-Hodgkin lymphoma [O'Brien et al., 2022], hereditary transthyretin amyloidosis [Gillmore et al., 2021], among others. However, the large size of the Cas9 molecule presents challenges for therapeutic delivery. Its large size (in the range of 3-4kb) prohibits its use of adeno-associated virus (AAV) delivery, which has a packaging size below 4.7kb [Wu et al., 2010]. As a consequence, the vast majority of current CRISPR-Cas9 therapies are restricted to ex-vivo or lipid nanoparticle-based delivery modalities, severely limiting their general application. The dCasMINI molecule [Xu et al., 2021] is extremely small and ideal for AAV-based delivery, as its small size of 529 amino acids (1587bp) allows researchers to package the dCasMINI DNA sequence, guide RNA, associated promoter sequences, and modulator peptides capable of gene regulation into a standard AAV vector with a maximum cargo size of 4.7kb. Furthermore, traditional Cas9-based therapeutics are limited to the treatment of diseases that are ameliorated by gene knockouts and are thus unsuitable for a whole host of genetic diseases such as those caused by haploinsufficiency. On the other hand, dCasMINI has an inactivated nuclease domain and is thus capable of upregulating or downregulating a genetic locus when tethered to the appropriate modulator peptides. Beyond its smaller size and modulatory versatility, recent research has additionally suggested that dCasMINI has a lower incidence of off-targets than Cas9 or Cas12a [Xin et al., 2022], making it an attractive Cas molecule candidate for therapeutic applications.

To assist researchers in the applications of dCasMINI-based tools, we have developed a web-based database to rapidly search for optimal dCasMINI spacer sequences for CRISPR interference (CRISPRi)

[*]These authors contributed equally to this work
[†]To whom correspondence should be addressed

and CRISPR activation (CRISPRa) at a desired genetic locus (Figure 1). To improve upon pre-existing tools such as CHOPCHOP [Labun et al., 2019] or Casilico [Asadbeigi et al., 2022], we performed extensive computational mapping of each potential spacer sequence to extend the search for off-targets, and provided the full off-target information for each guide so that researchers can determine what type and how many off-targets they can tolerate and/or test. To further characterize computationally predicted off-targets, we experimentally determined the guide seed region (high fidelity region at 5' of spacer that is intolerant to mismatches) for dCasMINI and annotate potential off-targets with information about seed region mismatches. This will allow researchers to increase the potential search space of guides by including guides with computationally predicted off-targets that have mismatches in the seed region (and thus unlikely to result in a true binding event). We expect that our casmini-tool and website portal will greatly simplify and accelerate the guide design workflow for researchers carrying out experiments using the dCasMINI platform.

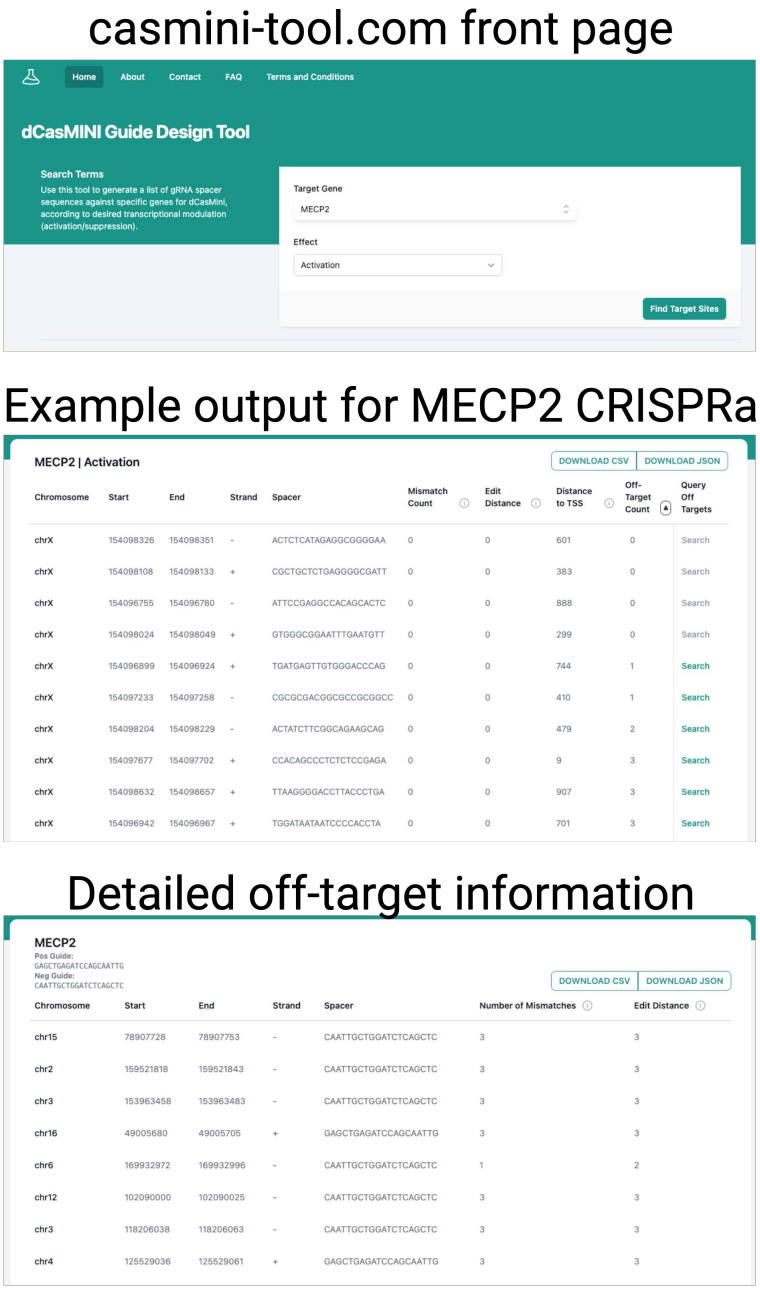


Figure 1: Example workflow for designing guides targeting the Rett syndrome associated gene MECP2 for CRISPRa.

Methods

Computational identification of on-target gRNA sequences

To identify on-target guide sequences for CRISPRi/a genome-wide, we first extracted the primary transcription start site (TSS) for each gene using the hg38 FANTOM5 database, a repository of genome-wide TSSs defined from human CAGE-seq data [fan, 2014]. Based on in-house experiments [data not shown], we defined the optimal targeting region for each regulatory modality as follows:

- CRISPRi: -200bp:+1000bp from the TSS;
- CRISPRa: -1000bp:+200bp from the TSS.

Nucleotide sequences for each of these targeting regions genome-wide were extracted using the pybedtools command *getfasta* [Dale et al., 2011]. We then identified all dCasMINI PAMs (TTTA/TTTG) in these sequences and extracted the 20 nucleotides directly downstream of each PAM to generate the list of on-target spacer sequences.

Computational mapping of putative off-targets

We sensitively mapped each 20bp on-target spacer sequence to the hg38 genome with Bowtie2 [Langmead and Salzberg, 2012], tolerating up to 3 mismatches anywhere in the spacer sequence and a maximum of twenty thousand alternative mapping locations, using the following command:

```
bowtie2 -f -D 20 -R 3 -N 0 -L 10 -i S,1,0.50 -k 20000 --score-min L,-.1,-.9
    -x GRCh38_genome -U {guide_sequences} -p 6 -S {out}
```

For each putative off-target site, we queried 5bp upstream and discarded any alignments that did not contain an appropriate PAM in the upstream region (we chose 5bp to allow for a single nucleotide bulge between the 4bp PAM and the spacer sequence). Given our experimental data on the importance of the 6bp seed region (see below), we further verified whether any mismatches were present in the first 6bp of the 20bp spacer alignment and annotated each off-target with this information.

Experimental determination of the dCasMINI seed region

To identify the high-fidelity seed region of dCasMINI spacer sequences, we synthesized a panel of gRNA-encoding gene fragments where each individual nucleotide in the spacer sequence targeting CD2 was mutated to all three other nucleotides ("single-mismatch") or deleted ("single-deletion"). Each gRNA variant was cloned into the gRNA plasmid backbone downstream of mU6 promoter. HEK293T cells were co-transfected with individual gRNA variant plasmids and our proprietary dCasMINI CRISPRa plasmid [Carosso et al., 2023] as triplicates in a 96-well plate format. Three days post-transfection, CD2 target gene activation was quantified by cell surface antibody staining of live cells using APC anti-human CD2 antibody (Biolegend, 309224) followed by flow cytometry on the cell population transfected with both gRNA and dCasMINI plasmids (performed on Cytoflex and analyzed using Flowjo software). The level of gene activation by each gRNA variant was compared and normalized to the activity of the wild-type gRNA as a control. The seed region in the spacer with minimal tolerance on single mismatch/deletion was defined with a cutoff of normalized activity below 0.2.

Results

Case study: generating CRISPRa guides against MECP2

As a case study, we queried our website tool for CRISPRa guides against MECP2, a gene frequently mutated in Rett syndrome, a severe and progressive X-linked neurodevelopmental disorder (Figure 1) [Lamonica et al., 2017]. Querying "Target Gene" = MECP2 and "Effect" = Activation on the website portal generated a list of on-target guide sequences situated -1000bp:+200bp around the TSS (since that is the appropriate targeting window for transcriptional activation) (Figure 2). On-target spacer sequences are annotated with chromosomal position and strand, mismatch count, edit distance, distance to the TSS, and off-target count. These guides are ranked in descending order based on number of off-targets to prioritize the more therapeutically relevant guide sequences. When interested in a specific spacer sequence with off-targets, the user can further query information about putative off-target binding sites by following the "Query Off Targets" search button associated with the on-target guide (Figure 1).

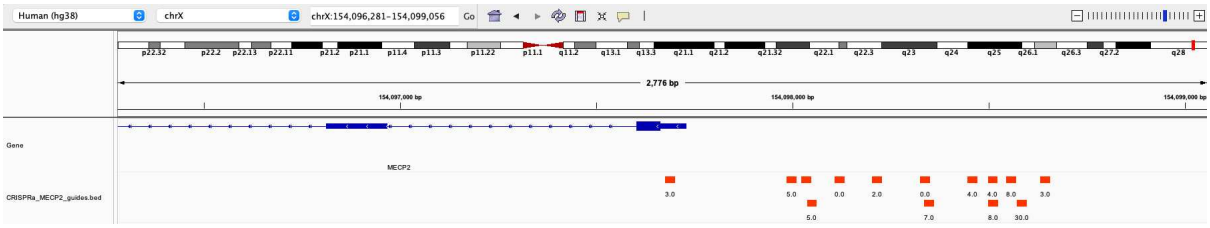


Figure 2: IGV genome browser visualization of CRISPRa guides targeting MECP2 annotated with number of potential off-targets per guide.

A comprehensive database for computational identification of on-target gRNA sequences

Our tool provides on-target dCasMINI guide sequences for 22,498 human genes suitable for CRISPRi and/or CRISPRa (based on position relative to the primary TSS). In total, we identified 420,024 spacer sequences suitable for CRISPRi targeting 22,429/22,498 = 99.7% of genes and 477,655 spacer sequences for CRISPRa targeting 22,474/22,498 = 99.9% of genes (96,738 spacer sequences were situated +/- 200bp around the TSS and thus appropriate for both CRISPRi and CRISPRa). For CRISPRi, genes have an average of 19 on-target spacer sequences suitable for transcriptional suppression (Figure 3). For CRISPRa, genes have an average of 21 on-target spacer sequences suitable for transcriptional activation (Figure 4).

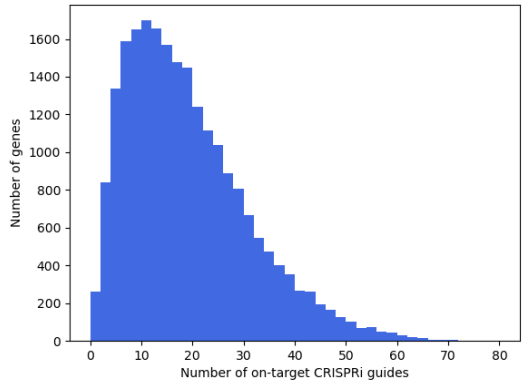


Figure 3: Histogram displaying number of on-target CRISPRi guides per gene genome-wide.

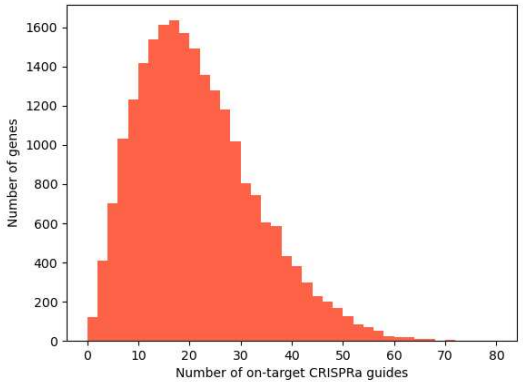


Figure 4: Histogram displaying number of on-target CRISPRa guides per gene genome-wide.

dCasMINI seed region experiment

To computationally predict off-targets, we first investigated the tolerance of dCasMINI on mismatched target sites and characterize the high-fidelity seed region on its spacer sequence [Slaymaker et al., 2016], we systematically mutated the CD2 guide spacer sequence to introduce single-base mismatches and single-base deletions at different positions, and then measured the tolerance of dCasMINI-modulator fusion protein against these spacer variants in the context of CRISPRa. We identified a conserved seed sequence in the +1 to +6 range where such single mismatches/deletions are minimally tolerated by dCasMINI, irrespective of the modulator used (Figure 5).

Computational mapping of putative off-targets

To provide comprehensive off-target information for CRISPRi/a guides, we used Bowtie2 to sensitively map each on-target spacer sequence genome-wide (tolerating up to 3 mismatches) and annotated potential off-targets with information about whether mismatches were present in the high-fidelity 6bp seed region of the spacer. We anticipate this will allow researchers to prioritize testing guides with minimal predicted off-target effects, which is critical for developing safe and efficacious therapeutic products. In total, we mapped 228,370,545 potential off-target sites associated with CRISPRi spacer sequences and 454,352,553 potential off-target sites associated with CRISPRa spacer sequences.

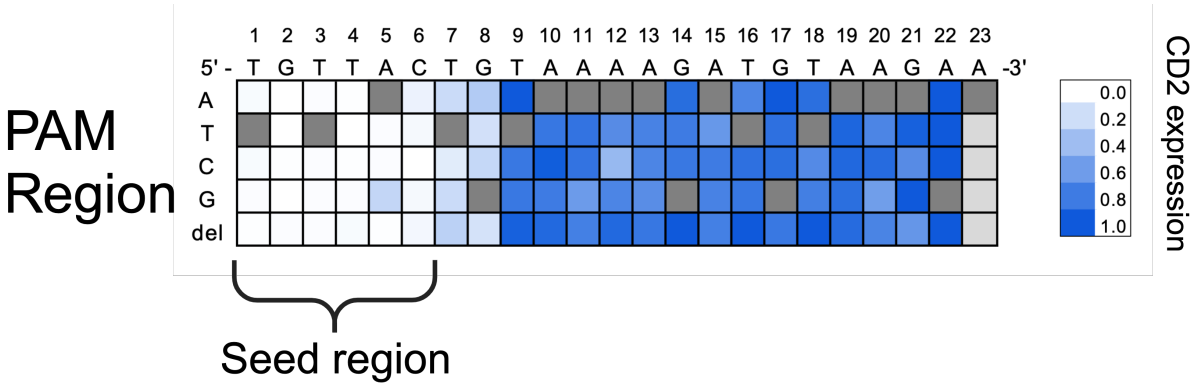


Figure 5: Heatmap illustrating single mismatch/deletion tolerance of dCasMini spacer sequences targeting endogenous CD2 for transcriptional activation, 3 day post transfection. The heatmap was generated by normalizing CD2 expression with a given single mismatch/deletion spacer to the maximum expression level of CD2 with the WT gRNA sequence. Lower expression indicates that the mismatch disrupts

Ultimately we predict users will be most interested in selecting guides with a minimal number of potential off-target binding events. Analyzing the off-target database reveals that genes have an average of 8 on-target guides with ≤ 5 potential off-target sites for CRISPRi (Figure 6) as well as CRISPRa (Figure 7). However, many of these potential off-target sites (approximately 26%) contain one or more mismatches in the spacer high-fidelity seed region which are extremely unlikely to result in true off-target binding events (Figure 5). Incorporating these data on seed region fidelity allows us to effectively discard 1/4 computationally predicted off-target sites which have been experimentally demonstrated to be false positive predictions, thus increasing the number of potentially therapeutically relevant on-target guides to test.

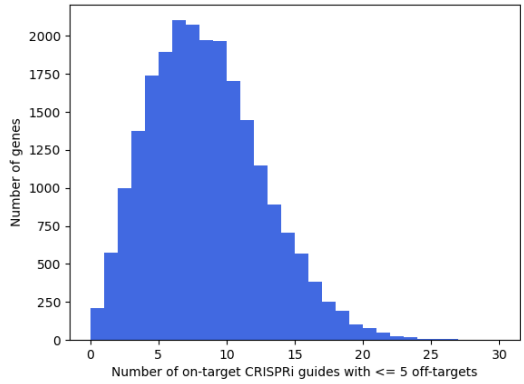


Figure 6: Histogram displaying number of on-target CRISPRi guides with ≤ 5 potential off-target sites per gene genome-wide.

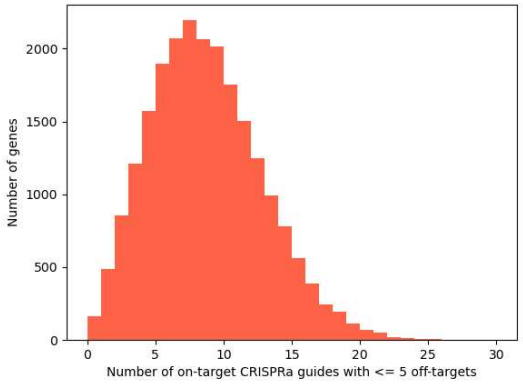


Figure 7: Histogram displaying number of on-target CRISPRa guides with ≤ 5 potential off-target sites per gene genome-wide.

Discussion

The simplicity and versatility of the CRISPR-Cas9 platform for genetic editing has increasingly resulted in its use as a therapy for the treatment of genetic diseases [Wang et al., 2016]. While extremely powerful, the traditional CRISPR-Cas9 platform suffers from a number of key drawbacks: its large size prevents it from being delivered in a single AAV and its nuclease ability renders it unsuitable for the treatment of a host of genetic disorders in which healthy genes are inactivated or more subtle downregulation is required. Since the discovery of Cas9 [Jinek et al., 2012], there has been an explosion of research into the diversification and optimization of other Cas molecules [Chavez et al., 2023]. Recently, a hyper-compact, nuclease-inactivated Cas molecule, termed dCasMINI, was engineered to be small enough for AAV delivery without compromising on-target efficacy or the incidence of off-target binding events making it an ideal

Cas molecule candidate for the development of AAV-based therapies [Xu et al., 2021]. However, different Cas molecules are governed by different guide design principles making guide sequence design a frequent bottleneck in the dCasMINI workflow for CRISPRi/a.

Here we present a tool to easily design gRNAs against any human loci of interest for CRISPRi/a experiments, along with the first presentation of mismatch data in the high-fidelity seed region of dCasMINI gRNAs. We anticipate that this web-based tool will be of value to the CRISPR-Cas research community and will allow researchers to more easily and quickly design CRISPRi/a experiments with dCasMINI. We plan to add new functionality in future versions, so please look out for updates.

Acknowledgements

The authors would like to thank Yanxia Liu (Functional Genomics, EpiCRISPR Biotechnologies) for helpful conversation and reviewing the manuscript.

Authors' Contributions

S.L. and R.W.Y. developed the backend database of on- and off-target guides. S.L. and N.J. built the frontend website. X.Y. developed and analysed the seed region experiments. D.O.H., L.S.Q, and T.P.D. supervised and directed this project. R.W.Y. and T.P.D. wrote the manuscript with input from all authors.

Author Disclosure Statement

L.S.Q. is the founder of Epicrispr Biotechnologies, and also serves as a scientific advisor for Laboratory of Genomics Research and Kytopen. S.L., R.W.Y., X.Y., D.O.H., L.S.Q., and T.P.D. hold provisional patents relating to this work, are employees of and acknowledge outside interest in Epicrispr Biotechnologies.

References

A promoter-level mammalian expression atlas. *Nature*, 507(7493):462–470, 2014.

Adnan Asadbeigi, Milad Norouzi, Mohammad Sadegh Vafaei Sadi, Mojtaba Saffari, and Mohammad Reza Bakhtiarizadeh. Casilico: A versatile crispr package for in silico crispr rna designing for cas12, cas13, and cas14. *Frontiers in Bioengineering and Biotechnology*, 10:957131, 2022.

Giovanni Carosso, Robin Yeo, T Blair Gainous, M Zaki Jawaid, Xiao Yang, Vincent Cutillas, Timothy Daley, Daniel Hart, and Lei Qi. Discovery and engineering of hypercompact epigenetic modulators for durable gene activation. 2023.

Michael Chavez, Xinyi Chen, Paul B Finn, and Lei S Qi. Advances in crispr therapeutics. *Nature Reviews Nephrology*, 19(1):9–22, 2023.

Ryan K Dale, Brent S Pedersen, and Aaron R Quinlan. Pybedtools: a flexible python library for manipulating genomic datasets and annotations. *Bioinformatics*, 27(24):3423–3424, 2011.

Edouard De Dreuzy, Jack Heath, John A Zuris, Patricia Sousa, Ramya Viswanathan, Sean Scott, Jen Da Silva, Terence Ta, Stacy Capehart, Tongyao Wang, et al. Edit-301: an experimental autologous cell therapy comprising cas12a-rnp modified mpb-cd34+ cells for the potential treatment of scd. *Blood*, 134:4636, 2019.

Haydar Frangoul, Yael Bobruff, Maria Domenica Cappellini, Selim Corbacioglu, Christine Marie Fernandez, Josu De la Fuente, Stephan A Grupp, Rupert Handgretinger, Tony W Ho, Suzan Imren, et al. Safety and efficacy of ctx001 in patients with transfusion-dependent β-thalassemia and sickle cell disease: early results from the climb thal-111 and climb scd-121 studies of autologous crispr-cas9-modified cd34+ hematopoietic stem and progenitor cells. *Blood*, 136:3–4, 2020.

Julian D Gillmore, Ed Gane, Jorg Taubel, Justin Kao, Marianna Fontana, Michael L Maitland, Jessica Seitzer, Daniel O'Connell, Kathryn R Walsh, Kristy Wood, et al. Crispr-cas9 in vivo gene editing for transthyretin amyloidosis. *New England Journal of Medicine*, 385(6):493–502, 2021.

Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. A programmable dual-rna–guided dna endonuclease in adaptive bacterial immunity. *science*, 337(6096):816–821, 2012.

Kornel Labun, Tessa G Montague, Maximilian Krause, Yamila N Torres Cleuren, Håkon Tjeldnes, and Eivind Valen. Chopchop v3: expanding the crispr web toolbox beyond genome editing. *Nucleic acids research*, 47(W1):W171–W174, 2019.

Janine M Lamonica, Deborah Y Kwon, Darren Goffin, Polina Fenik, Brian S Johnson, Yue Cui, Hengyi Guo, Sigrid Veasey, Zhaolan Zhou, et al. Elevating expression of mecp2 t158m rescues dna binding and rett syndrome–like phenotypes. *The Journal of clinical investigation*, 127(5):1889–1904, 2017.

Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with bowtie 2. *Nature methods*, 9 (4):357–359, 2012.

Joseph P McGuirk, Constantine S Tam, Nicolaus Kröger, Peter A Riedell, Hemant S Murthy, Phoebe Joy Ho, Joseph E Maakaron, Edmund K Waller, Farrukh T Awan, Paul J Shaughnessy, et al. Ctx110 allogeneic crispr-cas9-engineered car t cells in patients (pts) with relapsed or refractory (r/r) large b-cell lymphoma (lbcl): Results from the phase 1 dose escalation carbon study. *Blood*, 140(Supplement 1):10303–10306, 2022.

Susan O'Brien, Loretta J Nastoupil, James Essell, Ly Dsouza, Doug Hart, Emiri Matsuda, Tarah Satterfield, Tonia Nesheiwat, Ashley Hammad, Franco Davi, et al. A first-in-human phase 1, multicenter, open-label study of cb-010, a next-generation crispr-edited allogeneic anti-cd19 car-t cell therapy with a pd-1 knockout, in patients with relapsed/refractory b cell non-hodgkin lymphoma (antler study). *Blood*, 140(Supplement 1):9457–9458, 2022.

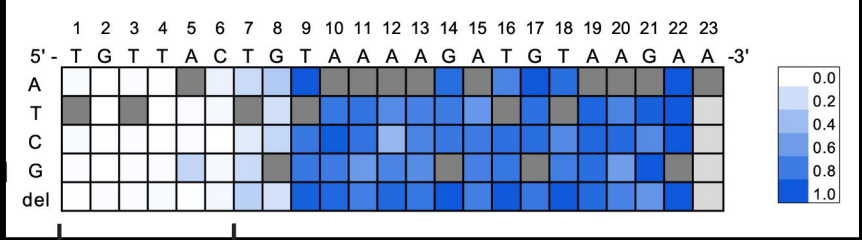
Ian M Slaymaker, Linyi Gao, Bernd Zetsche, David A Scott, Winston X Yan, and Feng Zhang. Rationally engineered cas9 nucleases with improved specificity. *Science*, 351(6268):84–88, 2016.

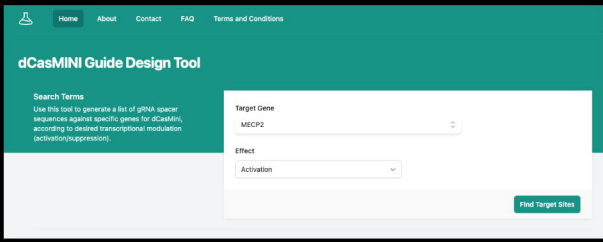
Haifeng Wang, Marie La Russa, and Lei S Qi. Crispr/cas9 in genome editing and beyond. *Annual review of biochemistry*, 85:227–264, 2016.

Zhijian Wu, Hongyan Yang, and Peter Colosi. Effect of genome size on aav vector packaging. *Molecular Therapy*, 18(1):80–86, 2010.

Changchang Xin, Jianhang Yin, Shaopeng Yuan, Liqiong Ou, Mengzhu Liu, Weiwei Zhang, and Jiazhi Hu. Comprehensive assessment of miniature crispr-cas12f nucleases for gene disruption. *Nature communications*, 13(1):5623, 2022.

Xiaoshu Xu, Augustine Chemparathy, Leiping Zeng, Hannah R Kempton, Stephen Shang, Muneaki Nakamura, and Lei S Qi. Engineered miniature crispr-cas system for mammalian genome regulation and editing. *Molecular cell*, 81(20):4333–4345, 2021.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
5' - T G T T A C T G T A A A A G A T G T A A G A A -3'
A
T
C
G
del
0.0
0.2
0.4
0.6
0.8
1.0



MECP2 | Activation

DOWNLOAD CSV | DOWNLOAD JSON

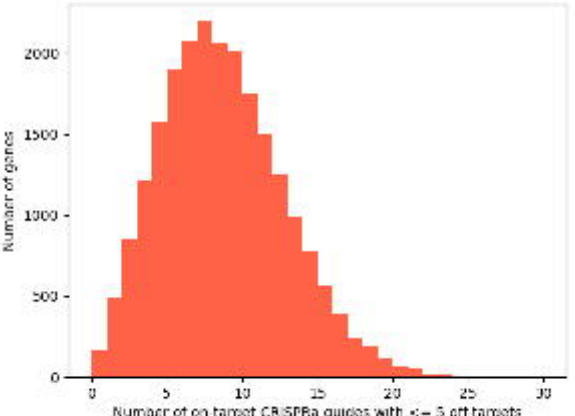
Chromosome	Start	End	Strand	Spacer	Mismatch Count	Edit Distance	Distance to TSS	Off-Target Count	Query Off Targets
chrX	154098326	154098351	-	ACTCTCATAGAGGCGGGGAA	0	0	601	0	Search
chrX	154098108	154098133	+	CGCTGCTCTGAGGGGCGATT	0	0	383	0	Search
chrX	154096755	154096780	-	ATTCCGAGGCCACAGCACTC	0	0	888	0	Search
chrX	154098024	154098049	+	GTGGGCGGAATTTGAATGTT	0	0	299	0	Search
chrX	154096899	154096924	+	TGATGAGTTGTGGGACCCAG	0	0	744	1	Search
chrX	154097233	154097258	-	CGCGCGACGGCGCCGCGGCC	0	0	410	1	Search
chrX	154098204	154098229	-	ACTATCTTCGGCAGAAGCAG	0	0	479	2	Search
chrX	154097677	154097702	+	CCACAGCCCTCTCTCCGAGA	0	0	9	3	Search
chrX	154098632	154098657	+	TTAAGGGGACCTTACCCTGA	0	0	907	3	Search
chrX	154096942	154096967	+	TGGATAATAATCCCCACCTA	0	0	701	3	Search

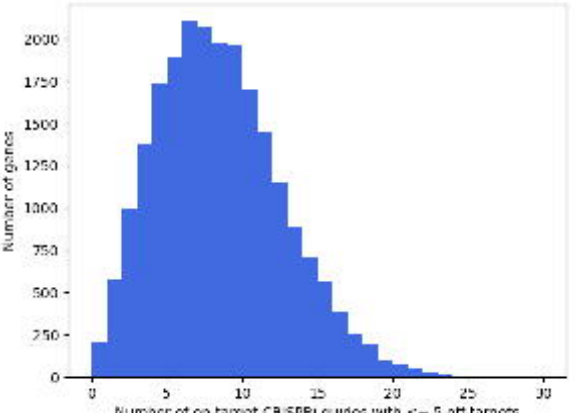
MECP2

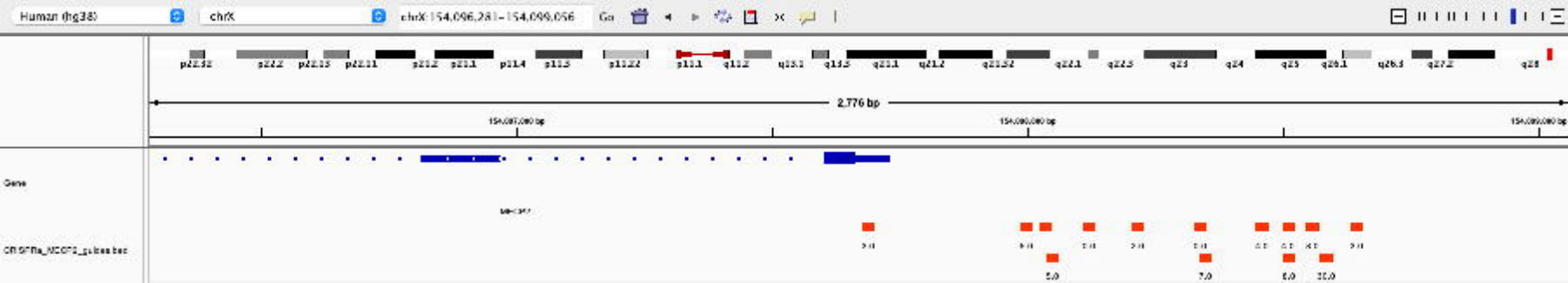
Pos Guide:
GAGCTGAGATCCAGCAATTG
Neg Guide:
CAATTGCTGGATCTCAGCTC

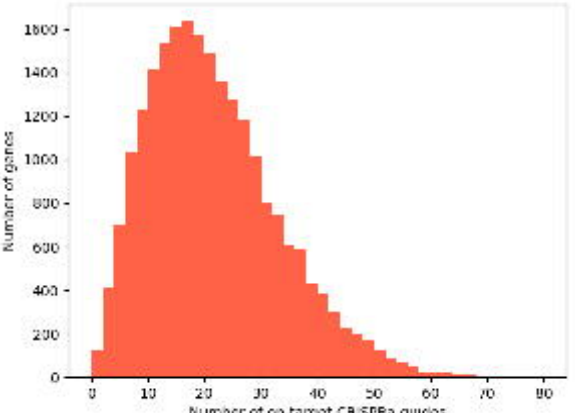
DOWNLOAD CSV | DOWNLOAD JSON

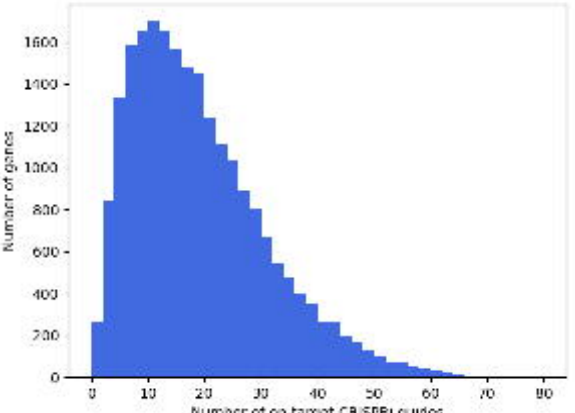
Chromosome	Start	End	Strand	Spacer	Number of Mismatches	Edit Distance
chr15	78907728	78907753	-	CAATTGCTGGATCTCAGCTC	3	3
chr2	159521818	159521843	-	CAATTGCTGGATCTCAGCTC	3	3
chr3	153963458	153963483	-	CAATTGCTGGATCTCAGCTC	3	3
chr16	49005680	49005705	+	GAGCTGAGATCCAGCAATTG	3	3
chr6	169932972	169932996	-	CAATTGCTGGATCTCAGCTC	1	2
chr12	102090000	102090025	-	CAATTGCTGGATCTCAGCTC	3	3
chr3	118206038	118206063	-	CAATTGCTGGATCTCAGCTC	3	3
chr4	125529036	125529061	+	GAGCTGAGATCCAGCAATTG	3	3


2000
1500
1000
500
0
Number of genes
0
5
10
15
20
25
30
Number of on target CRISPRa guides with <= 5 off targets


2000
1750
1500
1250
1000
750
500
250
0
Number of genes
0
5
10
15
20
25
30
Number of on target CRISPRi guides with <= 5 off targets


Human (hg38)
chrX
chrX:154,096,281–154,099,056
Go
2,776 bp
Gene
5.0
7.0
6.0
10.0


1800
1400
1200
1000
800
600
400
200
0
Number of genes
0
10
20
30
40
50
60
70
80
Number of on target CRISPRa guides


1600
1400
1200
1000
800
600
400
200
0
Number of genes
0
10
20
30
40
50
60
70
80
Number of on target CRISPRi guides