

dbGuide: a database of functionally validated guide RNAs for genome editing in human and mouse cells

Alexander A. Gooden[†], Christine N. Evans[†], Timothy P. Sheets[†], Michelle E. Clapp and Raj Chari^{✉*}

Laboratory Animal Sciences Program, Frederick National Lab for Cancer Research, Frederick, MD 21702, USA

Received August 14, 2020; Revised September 03, 2020; Editorial Decision September 14, 2020; Accepted September 22, 2020

ABSTRACT

With the technology's accessibility and ease of use, CRISPR has been employed widely in many different organisms and experimental settings. As a result, thousands of publications have used CRISPR to make specific genetic perturbations, establishing in itself a resource of validated guide RNA sequences. While numerous computational tools to assist in the design and identification of candidate guide RNAs exist, these are still just at best predictions and generally, researchers inevitably will test multiple sequences for functional activity. Here, we present *dbGuide* (<https://sgrnascorer.cancer.gov/dbguide>), a database of functionally validated guide RNA sequences for CRISPR/Cas9-based knockout in human and mouse. Our database not only contains computationally determined candidate guide RNA sequences, but of even greater value, over 4000 sequences which have been functionally validated either through direct amplicon sequencing or manual curation of literature from over 1000 publications. Finally, our established framework will allow for continual addition of newly published and experimentally validated guide RNA sequences for CRISPR/Cas9-based knockout as well as incorporation of sequences from different gene editing systems, additional species and other types of site-specific functionalities such as base editing, gene activation, repression and epigenetic modification.

INTRODUCTION

CRISPR/Cas9-based genome editing has been an indispensable technology for understanding the biology of living organisms (1,2). As a result, a tremendous effort has been invested in aiding researchers in designing and executing CRISPR-based experiments by providing critical resources

such as reagents and protocols as well as software tools that allow a user to rapidly identify guide RNA sequences with specific predicted on-target and off-target characteristics (3–18) (Table 1). However, these sequences are essentially predictions, often requiring assessment of multiple candidate guide RNAs.

Over the last 7 years, there has been an explosion of literature utilizing CRISPR/Cas9 to knockout specific genes or genomic regions, of which the majority is in human or mouse model systems. The culmination of all of this work has provided an opportunity to mine this information for guide RNA sequences shown to be functional at either a genotypic and/or phenotypic level. Variation in reporting standards, typographical errors and relegation of sequences into supplementary materials are among a few reasons why curation of this information has been non-trivial.

To this end, we present *dbGuide*, a database of functionally validated guide RNA sequences for CRISPR/Cas9 mediated knockout experiments in human or mouse cells (<http://sgrnascorer.cancer.gov/dbguide>). We have manually curated guide RNA sequences from over 1000 peer-reviewed articles with each sequence having direct reference to the original publication. We are also making results of targeted amplicon sequencing for ~2000 unique sgRNA sequences tested individually in human (293T) or mouse (NIH-3T3 or P19) cultured cells publicly searchable in our database. In total, these efforts encompass nearly 6000 unique guide RNA sequences for which some level of validated activity exists. To our knowledge, this represents the largest database of functionally validated guide RNA sequences. We expect this to be a continually growing resource through multiple mechanisms. We have provided a downloadable template to encourage researchers to submit their newly published/validated sequences and our computational framework will also allow for the inclusion of sequences for CRISPR systems other than for *Streptococcus pyogenes*, modalities such as base editing, gene activation/repression and epigenetic modifications, and sequences used in other species.

*To whom correspondence should be addressed. Tel: +1 301 846 711; Fax: +1 301 846 6165; Email: raj.chari@nih.gov

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

Table 1. List of sgRNA design tools

Name	URL	Species	Design?	Validated guides
CCTop (12)	https://crispr.cos.uni-heidelberg.de/	Many	Yes	No
CHOPCHOP (6)	http://chopchop.cbu.uib.no/	Many	Yes	No
CRISPOR (3)	http://crispor.tefor.net/	Many	Yes	No
CRISPRseek (16)	http://bioconductor.org/packages/release/bioc/html/CRISPRseek.html	Many	Yes	No
CRISPRz (13)	https://research.nhgri.nih.gov/CRISPRz/	Zebrafish	No	Yes
CRISPR-ERA (7)	http://crispr-era.stanford.edu/index.jsp	Many	Yes	No
CRISPR MultiTargeter (11)	http://www.multicrispr.net/	Many	No	No
Cas-Designer (8)	http://www.rgenome.net/cas-designer/	Many	Yes	No
DeepHF (14)	http://www.deephf.com/index/#/Predict	Human	No	No
E-CRISP (4)	http://www.e-crisp.org/E-CRISP/	Many	Yes	No
EuPaGDT (9)	http://grna.ctegd.uga.edu/	Many	Yes	No
inDelphi (18)	https://indelphi.giffordlab.mit.edu/	Human, Mouse	Yes	No
sgRNAcas9 (15)	http://www.biotoools.com/software.html	Many	Yes	No
sgRNA Designer (17)	https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design	Many	No	No
sgRNA Scorer 2.0 (32)	https://sgrnascorer.cancer.gov/	Many	Yes	No
WU-CRISPR (5)	http://crispr.wustl.edu/	Many	Yes	No
COMMERCIAL				
Addgene	https://www.addgene.org/crispr/reference/grna-sequence/	Many	No	Yes
CRISPR gRNA Design Tool (Atum)	https://www.atum.bio/eCommerce/cas9/input	Many	Yes	No
Benchling	www.benchling.com	Many	Yes	No
CRISPR Design Tool (Horizon Discovery)	https://horizondiscovery.com/en/products/tools/CRISPR-Design-Tool	Many	No	No
CRISPR sgRNA Design Tool (GenScript)	https://www.genscript.com/grna-design-tool.html	Many	No	No
CRISPR Cas9 Guide RNA Checker (IDT DNA)	https://www.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE	Many	Yes	No
True Designer Genome Editor (ThermoFisher Scientific)	https://www.thermofisher.com/us/en/home/life-science/genome-editing/genear-crispr/genear-crispr-search-and-design-tool.html	Many	Yes	No

MATERIALS AND METHODS

Master list of human and mouse sgRNA sequences

In addition to functionally validated sgRNA sequences, in order to provide further utility to the database, computationally designed sequences were also obtained from a variety of sources, primarily focused on protein-coding genes (17,19–26). These sources are listed in Supplementary Table S1. CSV-delimited files containing all sgRNA sequences present in our database for human (https://sgrnascorer.cancer.gov/downloads/hg_guide_info.csv.gz) and mouse (https://sgrnascorer.cancer.gov/downloads/mm_guide_info.csv.gz) can be downloaded directly from the dbGuide home page.

Published sgRNA sequences

A broad search of the PubMed database for ‘CRISPR OR Cas9’ was performed and yielded over 15 000 indexed citations. Subsequently, review articles, publications not using human or mouse cells, publications not using *S. pyogenes* Cas9 and publications not performing knock-out experiments were excluded. In total, guide RNA sequences were sourced from a total of 1322 peer-reviewed articles (Supplementary Table S2) from which the guide RNA sequences used could be ascertained. Although it is likely some published sequences have been missed, a data template is pro-

vided for which these missed sequences can be submitted and incorporated into subsequent database updates.

Targeted amplicon sequencing data

Quantitative editing data for nearly 2000 sgRNA sequences were generated internally from either transfection of Cas9/sgRNA plasmids (1 µg) or Cas9 protein (4 µg)/*in vitro* transcribed (IVT) (2.25 µg) RNA into mouse (NIH-3T3 or P19) or human (HEK293T) cells in a 24-well format. Cas9 protein was produced using plasmid Addgene-62731, a gift from Niels Geijsen (Addgene plasmid # 62731; <http://n2t.net/addgene:62731>; RRID:Addgene.62731) (27). IVT RNA was produced using a similar protocol as previously published (28). Cas9/RNP complexing and transfection using Lipofectamine 2000 were also performed similarly to as previously described (29). Polymerase chain reaction from genomic DNA was performed and amplicons were sequenced using the Illumina MiSeq V2 300 cycle kit using the PE (2 × 150) format. List of all sgRNAs sequences tested are listed in Supplementary Table S3.

Mapping information and on/off-target scoring metrics

For all sgRNA sequences obtained, genomic locations of the corresponding target sites were obtained/verified using UCSC BLAT against either the hg38 or mm10 reference genome sequences downloaded from the UCSC

A dbGuide

Welcome to the dbGuide database, a resource containing validated guide RNA sequences that have been used in gene editing experiments in human and mouse cells. We currently have aggregated information from over 1000 publications using the CRISPR/Cas9 technology in human or mouse cells as well as data generated by our lab assessing editing efficiency by Illumina sequencing. Each guide RNA listed that was sourced from a primary publication has a direct link to the publication.

To keep this resource current, we will be actively updating this database on a monthly basis and to help facilitate this, we invite researchers to submit data into our database. Our data template file can be found: [here](#). Once completed, please e-mail the file to sgnascorer@nih.gov. All contributions to this database will be given proper citation to the original publication. For further instructions/help on using our database and how to interpret results, please go [here](#).

Species: Human(hg38) Mouse(mmm10)
 Search with either a HUGO gene symbol (e.g. KRAS), Ensembl transcript (e.g. ENST00000256078), Ensembl gene (e.g. ENSG00000133703), or genomic coordinates in BED format (e.g. chr12:25205246-25250929)

If you would like to subscribe to updates please enter your e-mail address:

B

Select all Deselect all Copy CSV View Data Design New Search

Search:

Select	Gene-Symbol	guide_rna	crispr_system	position	num_publications	ind_pmids	n
<input type="checkbox"/>	BAP1	ACCGAAATCTTCCACGAGCA	SpCas9	chr3:52409712-52409734	1	32004714;	
<input type="checkbox"/>	BAP1	CCTCATCGCAGGTGTCAAGG	SpCas9	chr3:52409597-52409619	1	32004714;	
<input type="checkbox"/>	BAP1	TGGCAGTGGTCCCGACCGC	SpCas9	chr3:52406322-52406344	0		
<input type="checkbox"/>	BAP1	GCATGAGTTGCACAAGATT	SpCas9	chr3:52408049-52408071	0		
<input type="checkbox"/>	BAP1	CAAACCCCCAGTACCTGTG	SpCas9	chr3:52405750-52405772	0		
<input type="checkbox"/>	BAP1	GGAAATAATCCATATACA	SpCas9	chr3:52408584-52408606	0		
<input type="checkbox"/>	BAP1	CATCCTCAGCCAGGTGCAGC	SpCas9	chr3:52403443-52403465	0		
<input type="checkbox"/>	BAP1	GGTCTCCACTAGGGCCCTGG	SpCas9	chr3:52406897-52406919	0		
<input type="checkbox"/>	BAP1	CCTTGGATTGGCTGATGGA	SpCas9	chr3:52403256-52403278	0		
<input type="checkbox"/>	BAP1	GTTTGCTCCAGCACCAGCG	SpCas9	chr3:52405799-52405821	0		

Showing 1 to 10 of 427 entries

Previous 1 2 3 4 5 ... 43 Next

C

Select	Gene-Symbol	guide_rna	crispr_system	position	ast	total_nhej_rate	oof_rate	nun
<input checked="" type="checkbox"/>	BAP1	CCCCCGGGGAAGATGAATAA	SpCas9	chr3:52409868-52409890	34.46	13.8836	72.8279	
<input checked="" type="checkbox"/>	BAP1	ACCTGCTCGTGGGAAGATT	SpCas9	chr3:52409713-52409735	28.65	14.4078	83.0508	
<input checked="" type="checkbox"/>	BAP1	GCTATGCTCATCACAGCC	SpCas9	chr3:52407217-52407239	24.31	18.0146	85.1351	
<input checked="" type="checkbox"/>	BAP1	AGCCCTTATTATCTCCCG	SpCas9	chr3:52409866-52409888	24.001	18.9231	72.3684	
<input checked="" type="checkbox"/>	BAP1	CGACCGCAGGATCAAGTATG	SpCas9	chr3:52406309-52406331	26.16	22.739	81.25	
<input checked="" type="checkbox"/>	BAP1	GTTACGGAGCAGCTCAGCA	SpCas9	chr3:52408027-52408049	17.71	24.1525	93.2331	
<input checked="" type="checkbox"/>	BAP1	AGGCCTCTCCACCTCCTCG	SpCas9	chr3:52409723-52409745	36.5	26.6631	68.4423	
<input checked="" type="checkbox"/>	BAP1	ACCGAAATCTCCACGAGCA	SpCas9	chr3:52409712-52409734	24.62	29.012	82.3857	
<input checked="" type="checkbox"/>	BAP1	CCTCATCGCAGGTGTCAAGG	SpCas9	chr3:52409597-52409619	20.34	29.9072	86.8852	
<input type="checkbox"/>	BAP1	TGGCAGTGGTCCCGACCGC	SpCas9	chr3:52406322-52406344	25.1	None	None	

Showing 1 to 10 of 427 entries 9 rows selected

Previous 1 2 3 4 5 ... 43 Next

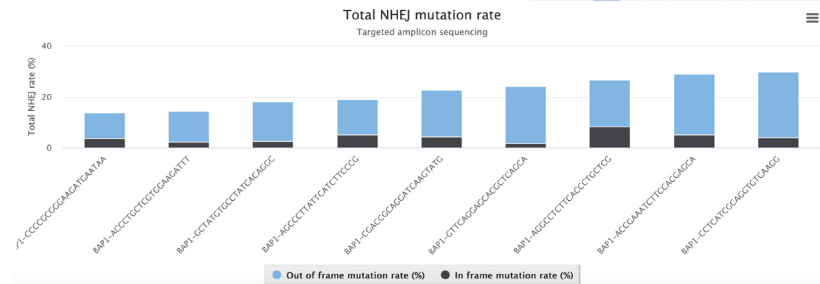


Figure 2. Screen shots of the key user interfaces. (A) Opening window of the *dbGuide* database. The user must first select whether to search within the human/mouse genome and then can specify a HGNC gene symbol, chromosomal coordinate in BED format, or ENSEMBL gene/transcript ID. (B) Window depicting results of a search by gene symbol. Selectable rows of sgRNA sequences are returned with various pre-computed metrics and PubMed identifiers, if the sequence had been used in a peer-reviewed publication. (C) Graphical representation of sgRNA sequences for which targeted amplicon sequencing data were generated. Stacked bar plots display the percentage of NHEJ mutation events that results in an ‘in-frame’ (black) or ‘out of frame’ (blue) amino acid change.

mal position (in BED format), gene symbol, *Ensembl* gene or transcript ID, or an sgRNA spacer sequence (without the PAM). In addition, a link to an excel spreadsheet template is provided for researchers wishing to contribute to the database.

Search results when searching by gene/coordinate/accession

Upon entering a valid search term, the user will be shown a table of results detailing all of the guide RNA sequences identified based on the search criteria (Figure 2B). The table displays the following columns:

Gene-Symbol: most recent official gene symbol for a gene,
guide_rna: sgRNA spacer sequence (without PAM),
crispr_system: Cas9 ortholog, currently restricted to Sp-Cas9,

position: genomic location of the sgRNA target site in either human (hg38) or mouse (mm10) assembly

in_protein_coding_exon: whether the guide RNA targets a protein coding exon,

num_transcripts: the number of GENCODE transcripts targeted,

transcript_id_list: list of GENCODE transcripts (by ENSEMBL ID),

sgrnascorer: predicted activity using the sgRNA Scorer 2.0 algorithm (−3 to 3). Higher the value, greater the predicted activity,

rule_set_2: predicted activity using the Rule Set 2 algorithm (0–1). Higher the value, greater the predicted activity

guide_scan_off: guidescan off-target score (0–1). Higher the value, higher the specificity

forecast: Favored Outcomes of Repair Events at Cas9 targets (FORECasT) score for predicting mutational outcomes (in frame indel %, lower values preferred for knock-outs),

total_nhej_rate: if amplicon sequencing data exists, value between 0 and 100, or ‘None’ otherwise

oof_rate: out of frame mutation rate which is the proportion of mutated reads which would lead to an out of frame mutation

num_publications: number of published articles using the selected guide RNA sequence

ind_pmids: list of articles by PubMed ID which used selected guide RNA sequence

num_screens: number of pooled CRISPR screens in which the guide RNA sequences was enriched or depleted

screen_pmids: list of articles by PubMed ID of CRISPR-based pooled screens in which this guide was enriched or depleted

sources: list of sources from which the guide RNA sequence has been curated from status

status: if the sgRNA sequence was functionally validated in a publication or targeted amplicon sequencing data exists, it is termed ‘validated’. Otherwise, it is termed ‘design’. This field allows the user to rapidly filter the list of returned sequences by typing ‘validate’ in the upper-right search box.

Viewing sequence editing data

In addition to sorting by total NHEJ rate, one can view the data in a graphical format. By selecting sgRNA sequences

for which data exists, the user can click the ‘View Data’ which will show a stacked bar plot breaking down the total NHEJ rate between ‘in frame’ and ‘out of frame’ percentage (Figure 2C). This can be specifically important for knockout experiments where a higher out of frame (OOF) mutation percentage gives higher probability of protein loss.

Downloading sequences for use in experiments

Upon identifying guide RNA sequences of interest, the user can export the sequences in a ‘ready to order’ format by clicking on the ‘Design’. A tab-delimited text file is generated listing oligonucleotides needed for either ligation into a plasmid or *in vitro* transcription as well as a PDF protocol describing what compatible vectors can be used and how they can be obtained. It is highly recommended to select and download multiple guide RNA sequences when performing knockout experiments to ensure any functional consequences observed may not be due to spurious off-target activity from a single guide RNA.

DATA AVAILABILITY

Sorted BAM files for amplicon sequencing data are available from NCBI under Project Identifier PRJNA664634.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to acknowledge Troy Taylor, William Gillette, Jane Jones and Dominic Esposito for kindly providing Cas9 protein. We would also like to acknowledge the Biomedical Informatics and Data Science (BIDS) and Frederick Research Computing Environment (FRCE) for providing assistance in web hosting and computational support.

FUNDING

Frederick National Laboratory for Cancer Research’s Technology and Training Fellowship (to T.P.S.) . Funding for open access charge: Frederick National Laboratory for Cancer Research.

Conflict of interest statement. None declared.

REFERENCES

- Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
- 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.
- Haeussler,M., Schönig,K., Eckert,H., Eschstruth,A., Mianné,J., Renaud,J.-B., Schneider-Maunoury,S., Shkumatava,A., Teboul,L., Kent,J. *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., Kerr,G. and Boutros,M. (2014) E-CRISP: fast CRISPR target site identification. *Nat. Methods*, **11**, 122–123.

5. Hiranniramol, K., Chen, Y., Liu, W. and Wang, X. (2020) Generalizable sgRNA design for improved CRISPR/Cas9 editing efficiency. *Bioinformatics*, **36**, 2684–2689.
6. Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H. and Valen, E. (2019) CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.*, **47**, W171–W174.
7. Liu, H., Wei, Z., Dominguez, A., Li, Y., Wang, X. and Qi, L.S. (2015) CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. *Bioinformatics*, **31**, 3676–3678.
8. Park, J., Bae, S. and Kim, J.-S. (2015) Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics*, **31**, 4014–4016.
9. Peng, D. and Tarleton, R. (2015) EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. *Microb. Genomics*, **1**, e000033.
10. Pliatsika, V. and Rigoutsos, I. (2015) ‘Off-Spotter’: very fast and exhaustive enumeration of genomic lookalikes for designing CRISPR/Cas guide RNAs. *Biol. Direct*, **10**, 4.
11. Prykhodzhiy, S.V., Rajan, V., Gaston, D. and Berman, J.N. (2015) CRISPR multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. *PLoS One*, **10**, e0119372.
12. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. and Mateo, J.L. (2015) CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One*, **10**, e0124633.
13. Varshney, G.K., Zhang, S., Pei, W., Adomako-Ankomah, A., Fohtung, J., Schaffer, K., Carrington, B., Maskeri, A., Slevin, C., Wolfsberg, T. *et al.* (2016) CRISPRz: a database of zebrafish validated sgRNAs. *Nucleic Acids Res.*, **44**, D822–D826.
14. Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F. *et al.* (2019) Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat. Commun.*, **10**, 4284.
15. Xie, S., Shen, B., Zhang, C., Huang, X. and Zhang, Y. (2014) sgRNAs9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS One*, **9**, e100448.
16. Zhu, L.J., Holmes, B.R., Aronin, N. and Brodsky, M.H. (2014) CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One*, **9**, e108424.
17. 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.
18. 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.
19. Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J. and Root, D.E. (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.*, **32**, 1262–1267.
20. Sanjana, N.E., Shalem, O. and Zhang, F. (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods*, **11**, 783–784.
21. Hart, T., Chandrashekar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S. *et al.* (2015) High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*, **163**, 1515–1526.
22. 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.
23. Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S. and Sabatini, D.M. (2017) Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic ras. *Cell*, **168**, 890–903.
24. Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S. and Sabatini, D.M. (2015) Identification and characterization of essential genes in the human genome. *Science*, **350**, 1096–1101.
25. Ztelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O.M., Mupo, A., Grinkevich, V., Li, M., Mazan, M. *et al.* (2016) A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell Rep.*, **17**, 1193–1205.
26. Henriksson, J., Chen, X., Gomes, T., Ullah, U., Meyer, K.B., Miragaia, R., Duddy, G., Pramanik, J., Yusa, K., Lahesmaa, R. *et al.* (2019) Genome-wide CRISPR screens in T helper cells reveal pervasive crosstalk between activation and differentiation. *Cell*, **176**, 882–896.
27. D’Astolfo, D.S., Pagliero, R.J., Pras, A., Karthaus, W.R., Clevers, H., Prasad, V., Lebbink, R.J., Rehmann, H. and Geijsen, N. (2015) Efficient intracellular delivery of native proteins. *Cell*, **161**, 674–690.
28. Richardson, C.D., Ray, G.J., Bray, N.L. and Corn, J.E. (2016) Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. *Nat. Commun.*, **7**, 12463.
29. Kim, K., Park, S.W., Kim, J.H., Lee, S.H., Kim, D., Koo, T., Kim, K.-E., Kim, J.H. and Kim, J.-S. (2017) Genome surgery using Cas9 ribonucleoproteins for the treatment of age-related macular degeneration. *Genome Res.*, **27**, 419–426.
30. Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
31. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
32. Chari, R., Yeo, N.C., Chavez, A. and Church, G.M. (2017) sgRNA Scorer 2.0: a species-independent model to predict CRISPR/Cas9 activity. *ACS Synth. Biol.*, **6**, 902–904.
33. Allen, F., Crepaldi, L., Alsinet, C., Strong, A.J., Kleshchevnikov, V., De Angeli, P., Páleniková, P., Khodak, A., Kiselev, V., Kosicki, M. *et al.* (2018) Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.*, **37**, 64–72.
34. Perez, A.R., Pritykin, Y., Vidigal, J.A., Chhangawala, S., Zamparo, L., Leslie, C.S. and Ventura, A. (2017) GuideScan software for improved single and paired CRISPR guide RNA design. *Nat. Biotechnol.*, **35**, 347–349.
35. Magoč, T. and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.
36. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
37. Köster, J. and Rahmann, S. (2018) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, **34**, 3600.