

Supporting information

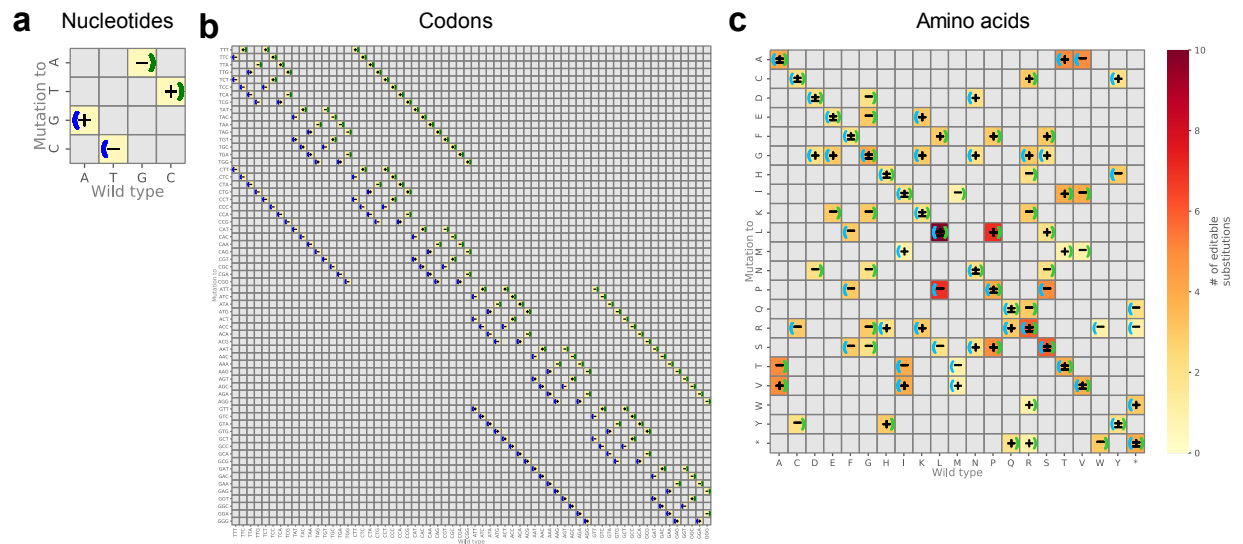
beditor: A computational workflow for
designing libraries of guide RNAs for
CRISPR-mediated base editing

Supporting figures 2

Supporting tables 8

References..... 18

Supporting figures



Flg S1 : Possible substitutions by current ABE and CBE.

a Nucleotide level substitutions.

b Codon level substitutions.

b Amino acid level substitutions.

Shown on the heatmaps are the cumulative number of substitutions that can be edited with either ABE or CBE. Left and right brackets indicate that the substitution is carried out by ABE and CBE respectively. +, - and \pm indicate substitutions for which guide RNA is designed on the +, - and both the strands respectively. Shown in gray are substitutions that are absent in the input data. The symbol * represents a non-sense mutation.

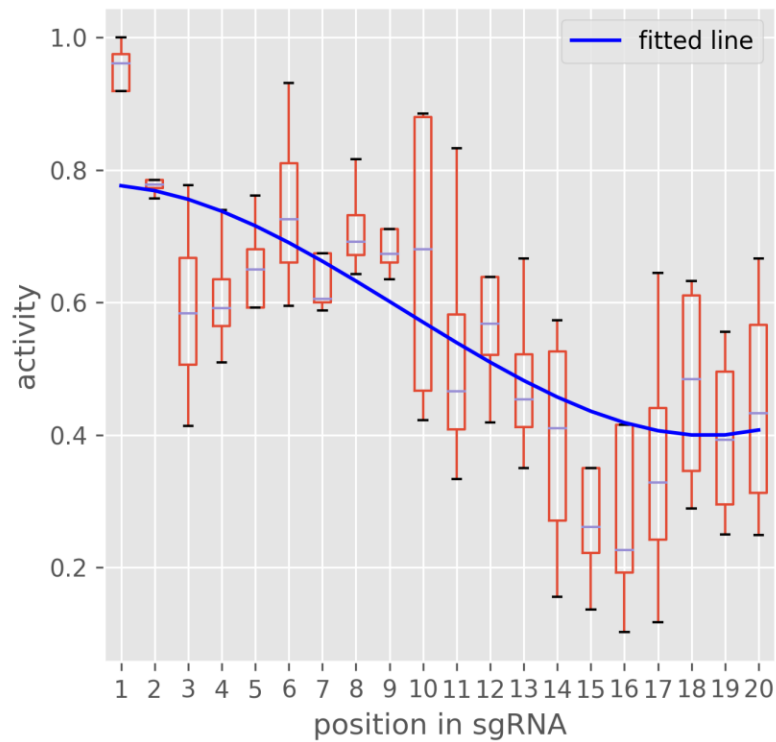


Fig S2: Fitting a third degree polynomial equation to the mismatch tolerance data from Doench et. al (3).

The gRNA activity values at each mismatch position are shown as boxplots (red). The third degree polynomial equation fitted line is shown in blue.

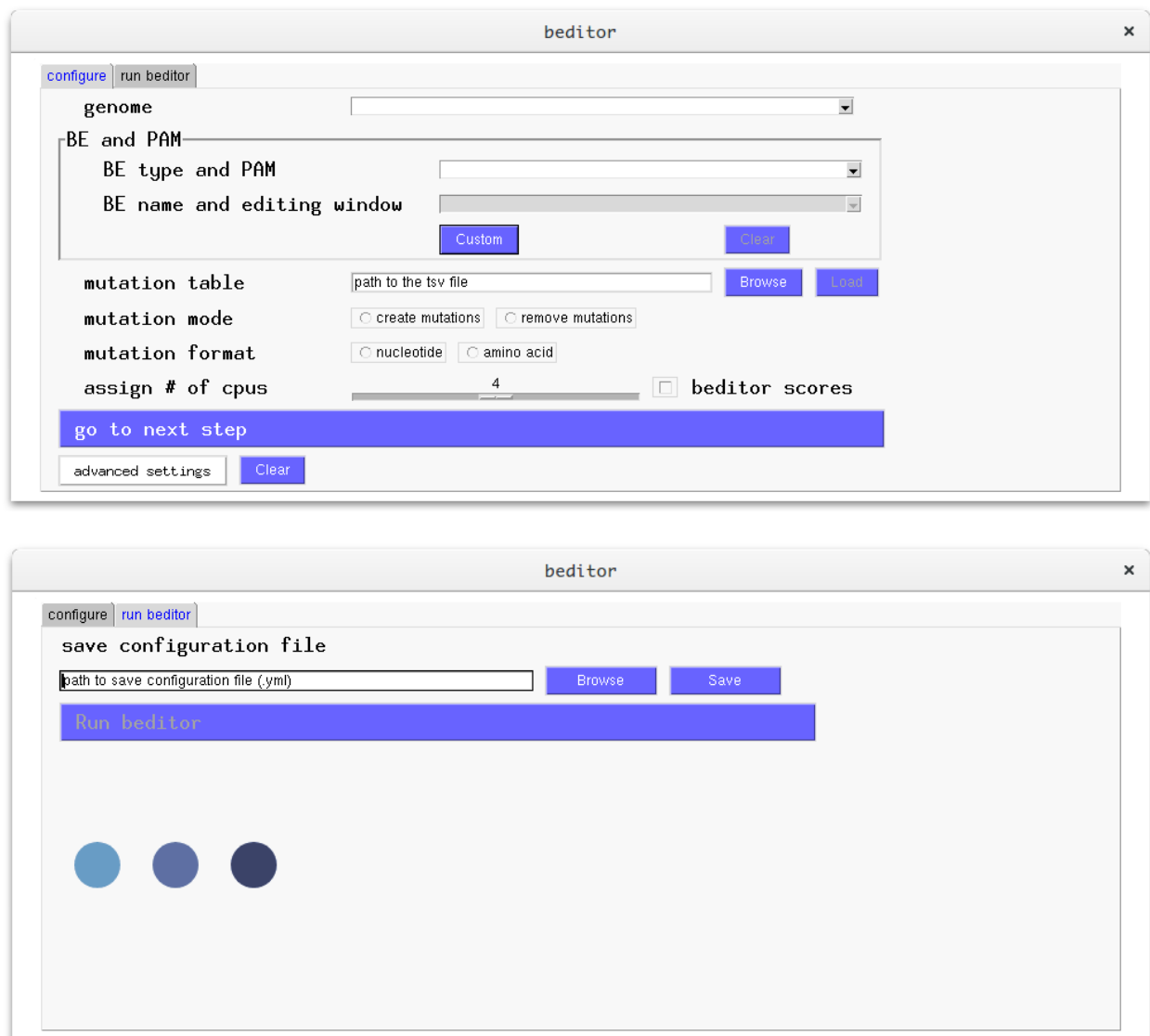


Fig S3: Graphical User Interface (GUI) of *beditor*.

Top: In the first tab of the GUI, the parameters for the analysis workflow can be provided using a series of options. These basic parameters include the species name, the name of the base editor and the PAM sequence, an input list of mutations, the format of mutations (amino acids and nucleotide) and mode of mutagenesis (model or correct), the number of cores (processors) and If the *beditor* scores are to be calculated for the gRNAs.

Bottom: the second tab of the GUI provides option to save the parameters of the analysis workflow as a YAML file and then subsequently run the program.

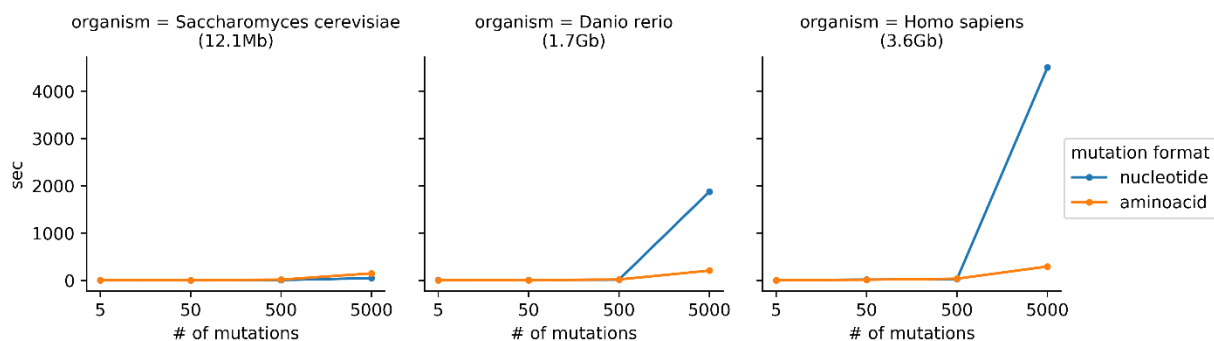


Fig S4: The time taken to design gRNA libraries depends on the number of mutations and the size of the genome.

Sets of 5-5000 nucleotide and amino acid mutations were tested with genomes of 3 species. 6 parallel processors (cores) were used for the analysis. The analysis was carried out using test_beditor.py script from test_beditor repository (https://github.com/rraadd88/test_beditor). Note that the time does not include the time taken for installation of the genomes (i.e. downloading and indexing of the genomes).

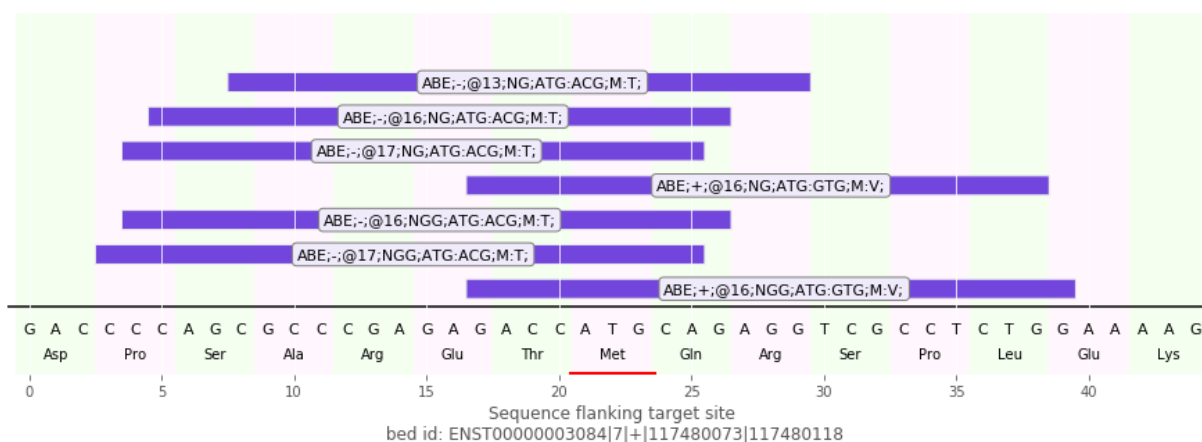


Fig S5: Representative visualizations of alignment between a gRNA and target genomic sequence.

The gRNA are shown in purple and with annotations denoting its identity (name of the base editor, target strand, distance from PAM sequence, PAM, reference codon, mutated codon, wild-type amino acid and mutated amino acid) is shown on the guide RNA. The target site is indicated in red color. Reading frames and genomic coordinates of the target DNA are shown below.

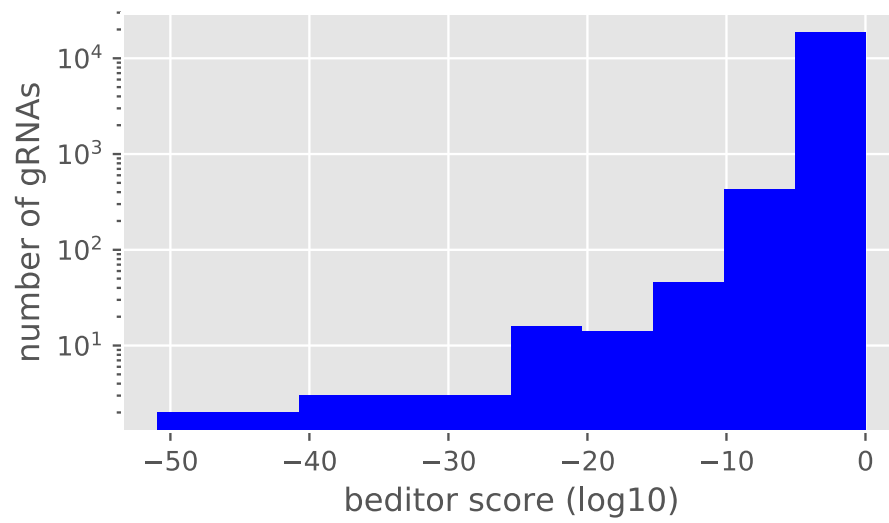


Fig S6: Distribution of beditor scores for the gRNA library designed for the demonstrative analysis of clinically associated human nucleotide mutations in 'model' mode.

Supporting tables

Table S1: Comparison of features of *beditor* with other existing tools.

Feature	<i>beditor</i>	iSTOP[1]	BE-analyzer[2]	Benchling [3]
Number of species whose genomes are directly compatible	>125	~70	~40	>100
Searches sites editable by base editors	Yes	Yes	Yes	Yes only BE [4]
Designs gRNAs against predefined set of nucleotide mutations	Yes	No	No	No
Designs gRNAs against predefined set of amino acid mutations	Yes	No	No	No
Designs gRNAs for mutational scanning using a predefined amino acid substitution matrix	Yes	No	No	No
Designs gRNAs against non-sense mutations	Yes	Yes	No	Yes
Designs control gRNAs	Yes	No	No	No
Generation of gRNA libraries for genome-wide screening of mutations	Yes	No	1 target site at a time	1 target site at a time
Multithreading/ parallel computing	Yes	No	No	No
Allows addition of any novel/custom PAM recognition sequence	Yes	No	No	No
Allows addition of any novel/custom base editors	Yes	No	No	No
Estimation of editing efficiency of gRNAs	Yes	No	No	Yes
Integrates independent measures of off-target effects	Yes (CFD scores[5])	No	No	Yes

Table S2: Penalties used for the calculation of the *beditor* score.

Penalty	Value
P_{min}	0.9
P_{max}	0.1
G_g	0.5
G_{ig}	0.9
A	0.5

Table S3: Default PAM recognition sequences supported by beditor.

PAM	Description	position	length of guide sequence	Reference
NG	xCas9	3'	20	[6]
NGA	SpCas9 mutant	3'	20	[7]
NGCG	SpCas9 mutant	3'	20	[7]
NGG	SpCas9	3'	20	[8,9]
NGGNG	Cas9 <i>S. Thermophilus</i>	3'	20	[10,11]
NGK	xCas9	3'	20	[6]
NGN	xCas9	3'	20	[6]
NNAGAA	Cas9 <i>S. Thermophilus</i>	3'	20	[10,11]
NNGRRT	SaCas9	3'	21	[12]
NNNNACA	CjCas9	3'	20	[13]
NNNNGMTT	Cas9 <i>N. Meningitidis</i>	3'	20	[14]
NNNRRT	KKH SaCas9	3'	21	[7,12]
TATV	AsCpf1 mutant	5'	23	[15]
TTN	Cpf1 <i>F. Novicida</i>	5'	23	[16]
TTTN	Cpf1 <i>Acidaminococcus</i> / <i>Lachnospiraceae</i>	5'	23	[16]
TYCV	TYCV AsCpf1 mutant	5'	23	[15]

Table S4: Input parameters used for the demonstrative analysis of clinically associated human genetic variants

Variable	Input	Description
Host		Name of host organism
genomerelease	93	Ensembl genome release
genomeassembly		Genome assembly version
dinp	din.tsv	File path of input tab-separated file
mutation_format	[aminoacid, nucleotide]	Whether the input data consists of amino acid or nucleotide mutations.
reverse_mutations	[FALSE, TRUE]	FALSE if design guide RNAs to 'model' mutations, TRUE to 'correct' mutations.
Mutations	mutations	Information about mutated amino acid is taken from the input file
mutation_type	N	Type of mutations to process, N: non-synonymous, S: synonymous, else: both
keep_mutation_nonsense	FALSE	Whether to process non-sense mutations
pams	[NGG, NG]	List of PAMs to use
bes	[Target-AID, ABE]	List of Base Editors to use
max_subs_per_codon	1	Maximum number of nucleotides that can be edited in the target codon.
mismatches_max	2	Maximum number of mismatches allowed in the alignment gRNAs against reference genome.
Cores	5	Number of processors to use for parallel processing
Chunksize	200	Number of mutations to process per processor

Table S5: Summary statistics of demonstrative analysis with custom base editors and PAM recognition sequences.

	nucleotide	amino acid	nucleotide	amino acid
Correct/model mutation	correct	correct	model	model
Total number of mutations in the input data	1000	1000	1000	1000
Total number of mutations edited	263	117	246	105
Total number of guides designed	817	562	791	403
% editability	26.3	11.7	24.6	10.5
number of guides per mutation	3.10646	4.80342	3.21545	3.8381

Table S6: Summary statistics of demonstrative analysis of representative set of species.

Species	metrics	nucleotide	amino acid	nucleotide	amino acid
	Correct/model mutation	correct	correct	model	model
	Total number of mutations in the input data	1000.00	1000.00	1000.00	1000.00
<i>Bos taurus</i>	Total number of mutations edited	76.00	45.00	69.00	36.00
	Total number of guides designed	123.00	83.00	121.00	62.00
	% editability	7.60	4.50	6.90	3.60
	number of guides per mutation	1.62	1.84	1.75	1.72
<i>Danio rerio</i>	Total number of mutations edited	45.00	39.00	57.00	43.00
	Total number of guides designed	71.00	59.00	74.00	72.00
	% editability	4.50	3.90	5.70	4.30
	number of guides per mutation	1.58	1.51	1.30	1.67
<i>Equus caballus</i>	Total number of mutations edited	80.00	53.00	55.00	43.00
	Total number of guides designed	136.00	94.00	92.00	81.00
	% editability	8.00	5.30	5.50	4.30
	number of guides per mutation	1.70	1.77	1.67	1.88
<i>Felis catus</i>	Total number of mutations edited	94.00	55.00	81.00	50.00
	Total number of guides designed	168.00	103.00	160.00	80.00
	% editability	9.40	5.50	8.10	5.00
	number of guides per mutation	1.79	1.87	1.98	1.60
<i>Gallus gallus</i>	Total number of mutations edited	93.00	39.00	74.00	46.00
	Total number of guides designed	180.00	71.00	139.00	77.00
	% editability	9.30	3.90	7.40	4.60
	number of guides per mutation	1.94	1.82	1.88	1.67
<i>Macaca fascicularis</i>	Total number of mutations edited	45.00	51.00	52.00	48.00
	Total number of guides designed	89.00	83.00	90.00	75.00
	% editability	4.50	5.10	5.20	4.80
	number of guides per mutation	1.98	1.63	1.73	1.56
<i>Mus musculus</i>	Total number of mutations edited	81.00	46.00	74.00	47.00
	Total number of guides designed	147.00	93.00	133.00	83.00
	% editability	8.10	4.60	7.40	4.70
	number of guides per mutation	1.81	2.02	1.80	1.77
<i>Pan paniscus</i>	Total number of mutations edited	73.00	57.00	62.00	54.00
	Total number of guides designed	123.00	89.00	98.00	90.00
	% editability	7.30	5.70	6.20	5.40
	number of guides per mutation	1.68	1.56	1.58	1.67
<i>Saccharomyces cerevisiae</i>	Total number of mutations edited	36.00	30.00	36.00	35.00
	Total number of guides designed	52.00	45.00	46.00	52.00

	% editability	3.60	3.00	3.60	3.50
	number of guides per mutation	1.44	1.50	1.28	1.49
<i>Sus scrofa</i>	Total number of mutations edited	68.00	50.00	55.00	45.00
	Total number of guides designed	108.00	106.00	86.00	82.00
	% editability	6.80	5.00	5.50	4.50
	number of guides per mutation	1.59	2.12	1.56	1.82

Table S7: Summary statistics for case study analysis of clinically relevant human SNPs.

Mutation format	nucleotide	amino acid	nucleotide	amino acid
Correct/model mutation	correct	correct	model	model
Total number of mutations in the input data	61083	81819	61083	81819
Total number of mutations edited	13709	19996	13867	20420
Total number of guides designed	23432	35390	22587	33311
% editability	22.44	24.43	22.70	24.95

Table S8: Default pairs of base editors and PAM recognition sequences supported by beditor.

Base editor	nucleotide wt	nucleotide mutation	window start	window end	guide length	PAM	PAM position	reference
BE1	C	T	4	8	20	NGG	downstream	[17]
BE2	C	T	4	8	20	NGG	downstream	[17]
BE3	C	T	4	8	20	NGG	downstream	[17]
HF-BE3	C	T	4	8	20	NGG	downstream	[17]
BE4/BE4max	C	T	4	8	20	NGG	downstream	[18,19]
BE4-Gam	C	T	4	8	20	NGG	downstream	[18]
YE1-BE3	C	T	5	7	20	NGG	downstream	[20]
EE-BE3	C	T	5	6	20	NGG	downstream	[20]
YE2-BE3	C	T	5	6	20	NGG	downstream	[20]
YEE-BE3	C	T	5	6	20	NGG	downstream	[20]
VQR-BE3	C	T	4	11	20	NGAN	downstream	[20]
VRER-BE3	C	T	3	10	20	NGCG	downstream	[20]
SaBE3	C	T	3	12	21	NNGRRT	downstream	[20]
SaBE4	C	T	3	12	21	NNGRRT	downstream	[18]
SaBE4-Gam	C	T	3	12	21	NNGRRT	downstream	[18]
SA(KKH)-BE3	C	T	3	12	21	NNNRRT	downstream	[20]
Cas12a-BE	C	T	10	12	23	TTTV	upstream	[21]
Target-AID	C	T	2	4	20	NGG	downstream	[22]
Target-AID	C	T	2	4	20	NG	downstream	[23]
xBE3	C	T	4	8	20	NG	downstream	[6]
eA3A-BE3	C	T	4	8	20	NGG	downstream	[24]
A3A-BE3	C	T	4	8	20	NGG	downstream	[25]
BE-PLUS	C	T	4	14	20	NGG	downstream	[26]
ABE7.9	A	G	5	8	20	NGG	downstream	[17]
ABE7.10	A	G	4	7	20	NGG	downstream	[17]
ABE7.10*	A	G	4	8	20	NGG	downstream	[17]
xABE7.10	A	G	4	7	20	NG	downstream	[6]
ABESa	A	G	6	12	21	NNGRRT	downstream	[6,27]
VQR-ABE	A	G	4	6	20	NGA	downstream	[6,27]
VRER-ABE	A	G	4	6	20	NGCG	downstream	[6,27]

Sa(KKH)-ABE	A	G	6	12	21	NNNRRT	downstream	[6,27]
--------------------	---	---	---	----	----	--------	------------	--------

References

1. Billon P, Bryant EE, Joseph SA, Nambiar TS, Hayward SB, Rothstein R, et al. CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Mol Cell* [Internet]. Elsevier Inc.; 2017;67:1068–1079.e4. Available from: <http://dx.doi.org/10.1016/j.molcel.2017.08.008>
2. Hwang G, Park J, Lim K, Kim S, Yu J, Eils R, et al. Web-based design and analysis tools for CRISPR base editing. *bioRxiv* [Internet]. 2018;5:373944. Available from: https://www.biorxiv.org/content/early/2018/07/22/373944?rss=1&utm_source=dlvr.it&utm_medium=twitter
3. Benchling Inc. Benchling for Academics · Benchling. Benchling. 2018.
4. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. Nature Publishing Group; 2016;533:420.
5. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* [Internet]. Nature Publishing Group; 2016;34:184–91. Available from: <http://dx.doi.org/10.1038/nbt.3437>
6. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* [Internet]. Nature Publishing Group; 2018;556:57–63. Available from: <http://dx.doi.org/10.1038/nature26155>
7. Kleinstiver BP, Prew MS, Tsai SQ, Topkar V V., Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. 2015;
8. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31:827–32.
9. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* (80-). 2012;
10. Magadán AH, Dupuis MÈ, Villion M, Moineau S. Cleavage of phage DNA by the *Streptococcus thermophilus* CRISPR3-Cas system. *PLoS One*. 2012;
11. Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 2010;
12. Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar V V., Zheng Z, et al. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol*. 2015;
13. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein

complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*. 2012;

14. Hou Z, Zhang Y, Propson NE, Howden SE, Chu L, Sontheimer EJ, et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci USA*. 2013;

15. Gao L, Cox DBT, Yan WX, Manteiga JC, Schneider MW, Yamano T, et al. Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol* [Internet]. Nature Publishing Group; 2017;35:789–92. Available from: <http://dx.doi.org/10.1038/nbt.3900>

16. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. 2015;

17. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* [Internet]. Nature Publishing Group; 2017;551:464–71. Available from: <http://www.nature.com/doi/10.1038/nature24644>

18. Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv*. 2017;

19. Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol*. 2018;

20. Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* [Internet]. Nature Publishing Group; 2017;35:371–6. Available from: <http://dx.doi.org/10.1038/nbt.3803>

21. Li X, Wang Y, Liu Y, Yang B, Wang X, Wei J, et al. Base editing with a Cpf1 – cytidine deaminase fusion. 2018;36.

22. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. 2016;8729:1–14.

23. Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* (80-) [Internet]. 2018;9129:eaas9129. Available from: <http://www.sciencemag.org/lookup/doi/10.1126/science.aas9129>

24. Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol* [Internet]. Nature Publishing Group; 2018; Available from: <http://dx.doi.org/10.1038/nbt.4199>

25. Gehrke JM, Cervantes OR, Clement MK, Pinello L, Joung JK, Unit MP, et al. High-precision CRISPR-Cas9 base editors with minimized bystander and off-target mutations. bioRxiv [Internet]. 2018;273938. Available from: <https://www.biorxiv.org/content/early/2018/03/01/273938>
26. Jiang W, Feng S, Huang S, Yu W, Li G, Yang G, et al. BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. Cell Res. 2018;
27. Hua K, Tao X, Zhu JK. Expanding the base editing scope in rice by using Cas9 variants. Plant Biotechnol J. 2019;