

BEscreen: a versatile toolkit to design base editing libraries

Philipp G. Schneider ^{1,2}, Shuang Liu ^{1,2}, Lars Bullinger ^{1,3,4}, Benjamin N. Ostendorf ^{1,2,3,5,*}

¹Department of Hematology, Oncology, and Tumor Immunology, Charité-Universitätsmedizin Berlin, 13353 Berlin, Germany

²Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine, 10115 Berlin, Germany

³German Cancer Consortium (DKTK), Partner Site Berlin, and German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

⁴National Center for Tumor Diseases (NCT), Partner Site, 13353 Berlin, Germany

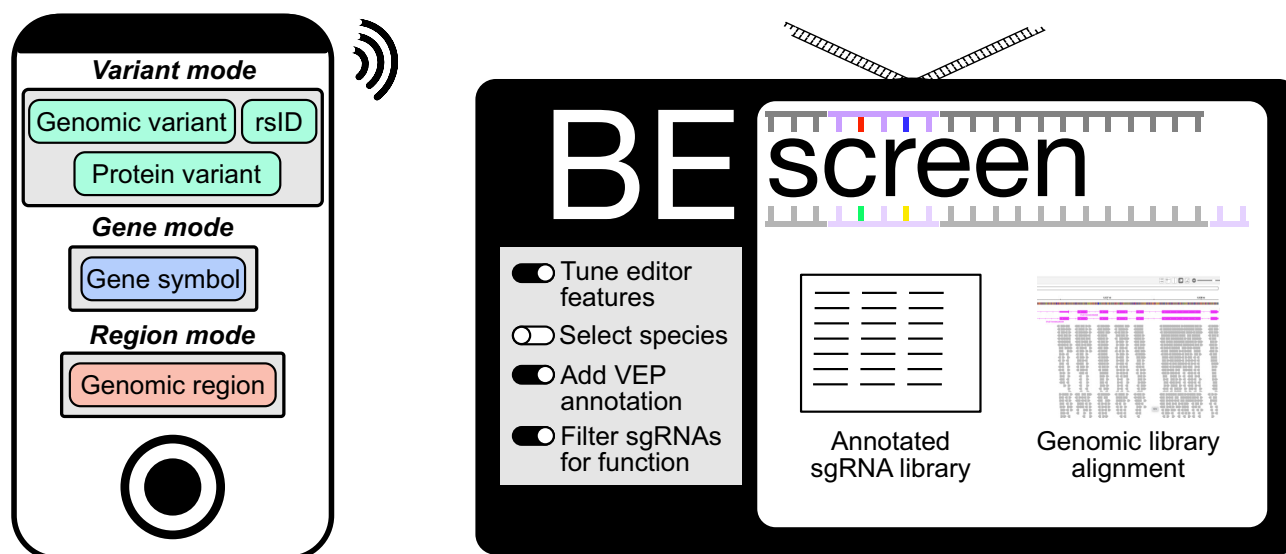
⁵Berlin Institute of Health, 10178 Berlin, Germany

*To whom correspondence should be addressed. Email: benjamin.ostendorf@charite.de

Abstract

Base editing enables the high-throughput screening of genetic variants for phenotypic effects. Base editing screens require the design of single guide RNA (sgRNA) libraries to enable either gene- or variant-centric approaches. While computational tools supporting the design of sgRNAs exist, no solution offers versatile and scalable library design enabling all major use cases. Here, we introduce BEscreen, a comprehensive base editing guide design tool provided as a web server (bescreen.ostendorflab.org) and as a command line tool. BEscreen provides variant-, gene-, and region-centric modes to accommodate various screening approaches. The variant mode accepts genomic coordinates, amino acid changes, or rsIDs as input. The gene mode designs near-saturation libraries covering the entire coding sequence of given genes or transcripts, and the region mode designs all possible guides for given genomic regions. BEscreen enables selection of guides by biological consequence, it features comprehensive customization of base editor characteristics, and it offers optional annotation using Ensembl's Variant Effect Predictor. In sum, BEscreen is a highly versatile tool to design base editing screens for a wide range of use cases with seamless scalability from individual variants to large, near-saturation libraries.

Graphical abstract



Introduction

Genetic variation is a central driver of phenotypic diversity, contributing to both physiological variability and susceptibility to a wide range of diseases [1]. The advent of next-generation sequencing has allowed for determining the association between many genetic variants with specific phenotypes in large human cohorts [2]. However, the assessment of which variants causally drive a given phenotype has remained a central challenge in the field [3].

Base editors (BEs) catalyze the conversion of nucleotides; cytosine base editors (CBEs) convert a C•G to a T•A base pair, while adenine base editors (ABEs) convert an A•T to a G•C base pair. ABEs and CBEs consist of a deaminase coupled to a catalytically impaired version of Cas9. This combination enables the use of single guide RNAs (sgRNAs) to direct the deaminase to specific sites in the genome, thus enabling the modification of specific genomic loci [4–6]. More recently, BEs have been explored that catalyze additional

Received: March 19, 2025. Revised: April 28, 2025. Editorial Decision: April 29, 2025. Accepted: May 3, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

nucleotide modifications [7, 8]. By transducing a pool of base editor-expressing cells with a library of sgRNAs to introduce multiple variants in parallel, base editing can be leveraged for the high-throughput assessment of variants for causal effects at scale [9, 10]. Recent studies have demonstrated the potential of such BE screens, leading to the identification of single nucleotide polymorphisms (SNPs) that modulate diverse phenotypes, including T cell function [11, 12], drug responses [13], and signaling [14, 15].

BE screens require computational tools to facilitate the design of sgRNA libraries. While several tools that aid in the design of BE sgRNAs exist [16–21], no tool provides broad and flexible functionality to address all major use cases for BE screens. These use cases range from interrogating specific, pre-defined variants to assessing all possible edits in given genes, transcripts, or otherwise defined genomic regions. In addition, over the last few years, a range of different BEs have been pioneered that exhibit distinct characteristics, including different editing windows, different requirements for the presence of protospacer adjacent motifs (PAMs), and others, requiring flexible options in an sgRNA design tool [5, 6, 22–25].

Here, we introduce BEscreen, a versatile toolkit that enables the design of BE sgRNA libraries for a broad range of use cases (Supplementary Fig. S1). BEscreen accepts different inputs, including genomic coordinates, protein mutations, genes, transcripts, and genomic regions. It thus aids in the design and functional annotation of sgRNAs for both variant- and gene-centric applications. Its capacity to filter sgRNAs for given biological consequences makes it uniquely suitable to design screening libraries with defined positive and negative controls. BEscreen also provides unprecedented flexibility in terms of defining PAM requirements and other base editor characteristics.

Materials and methods

BEscreen offers three modes focused on variants, genes, and genomic regions

BEscreen offers three different modes to design base editing guides for pre-defined variants, for coding sequences (CDS) of specific genes and transcripts, and for given genomic regions. Input can be provided manually or as a CSV file. An overview of the workflow using BEscreen is outlined in Fig. 1.

Variant-centric mode

BEscreen's variant mode identifies any possible sgRNA that mediates editing of a pre-defined nucleotide or amino acid variant. Nucleotide variants are accepted as a concatenated string composed of the genomic coordinates and reference and alternative bases. For example, "12_6537866_C_T" indicates a variant with a thymidine instead of a cytosine on chromosome 12 at position 6537866. Alternatively, nucleotide variants can be provided using their rsID from dbSNP (e.g. "rs1062436").

Amino acid mutations can be provided as gene or transcript name concatenated with an amino acid substitution (e.g. "GAPDH-L270F" for a leucine to phenylalanine substitution at position 207 in the default *GAPDH* transcript or "GAPDH-201-L270F" for the same variant in transcript 201 of *GAPDH*). BEscreen assesses whether the desired mutation can be introduced by editing a single base in one of the three

bases making up the codon. If so, it will convert this edit to genomic coordinates and design guides accordingly.

Gene-centric mode

BEscreen's gene mode identifies any possible sgRNA in the CDS of given genes or transcripts to generate near-saturation BE libraries. The gene mode accepts HUGO Gene Nomenclature Committee (HGNC) gene names (e.g. "GAPDH") as input with the option to specify a given transcript (e.g. "GAPDH-201" for *GAPDH* transcript 201). It then extracts the nucleotide sequence of all CDS regions in the specified gene(s) or transcript(s) and slides through this sequence searching for combinations of PAM sites and editable bases on both strands to identify and annotate guide sequences.

Region-centric mode

BEscreen's region mode accepts a genomic region as input (e.g. "12:6536490–6537490" for region 6536490–6537490 on chromosome 12). BEscreen extracts the nucleotide sequence of indicated regions and slides through the sequence searching for combinations of PAM sites and editable bases on both strands to identify all possible sgRNAs.

BEscreen is highly customizable

BEscreen offers extensive customization options for all three modes. Notably, BEscreen enables full customization of PAM site requirements and base editor characteristics, including adjusting the length of the desired guides and the position of the editing window. BEs are expected to exhibit high editing efficiency in their individually determined editing window, but edits may occur outside of this window, albeit at lower efficiency [20]. To account for this, users can define a safety region to annotate guides with possible off-target edits in regions of minor editing efficiency adjacent to the main editing window. For convenience, all parameters can be set using base editor presets [18, 20, 21, 23, 26].

In addition to customization options for input and base editor characteristics, BEscreen enables tailoring of the output. Importantly, as a part of this, BEscreen provides functionality to identify guides with given biological consequences. This feature enables convenient identification of screen controls, such as synonymously editing guides as negative controls and splice site disrupting or nonsense variant-mediating guides as positive controls.

BEscreen web server implementation

The BEscreen web server (bescreen.ostendorflab.org) is a Shiny for Python-based web app (v1.2.1; Python v3.12.3). The main third-party modules used are Polars (v1.16.0; github.com/pola-rs/polars) for data processing, pyfaidx (v0.8.1.3; github.com/mdshw5/pyfaidx) to extract genomic sequences from a FASTA reference genome, and PyRanges (v0.1.2; [27]) to parse a GTF annotation file. Optional annotation is based on Ensembl's Variant Effect Predictor (VEP; v113.3; [28]) and NCBI's BLAST (v2.16.0; [29]). igv.js (v3.1.3; [30]) is used with BAM files created by pysam (v0.22.1; github.com/pysam-developers/pysam) to display the alignment of sgRNAs to the reference genome. BEscreen provides the Ensembl reference genomes for human (GRCh38, Release 112) and mouse (GRCm39, Release 112) and is running on Ubuntu (24.04.1 LTS) and Apache2 (version 2.4.58). A command-line version is available at github.com/ostendorflab/bescreen.

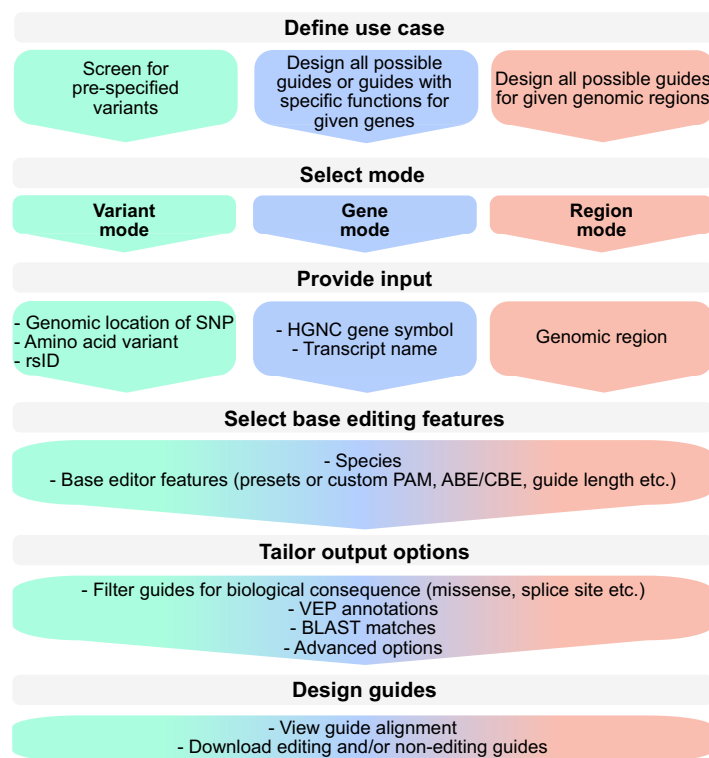


Figure 1. Use cases of BEScreen. Flowchart illustrating the use cases enabled by BEScreen's three modes (variant-, gene-, and region-based). BEScreen can be used to identify sgRNAs for specific variants (nucleotide- or amino acid-defined) as well as for genes and genomic regions. ABE, adenine base editor; CBE, cytosine base editor; HGNC, HUGO Gene Nomenclature Committee; PAM, protospacer adjacent motif; SNP, single nucleotide polymorphism; VEP, Variant Effect Predictor.

Results

BEScreen provides rich output

BEScreen's main output is a table providing detailed information of the identified guides, including the introduced variant(s), base change, guide sequence with and without PAM site, sequence of the editing window (and, if set, for the safety region), the genomic location, and the strand of guide alignment. In addition, the table details how many edits take place within the editing window, and the distance of these edits to the center of the editing window, where editing is most efficient. For the variant and gene modes, BEScreen annotates identified guides with the affected gene, transcript, and exon. In addition, BEScreen reports affected codons and their translation product pre- and post-editing with their positions. In the gene and region modes, output includes an additional table listing non-editing guides (sgRNAs predicted to introduce no edits).

Optionally, the output of BEScreen can be annotated using Ensembl's VEP [28]. To identify additional binding sites of given guides, BEScreen can identify matches of the guide sequence in the reference genome using NCBI's BLAST [29]. In addition to the tabular output, the BEScreen web server displays genomic alignment of the identified sgRNAs in an embedded version of the Integrative Genome Viewer (igv.js) [30].

Case study

To illustrate BEScreen's functionality in practice, we designed sgRNAs for variants in the *APOE* gene using the most commonly used ABE and CBE BEs. *APOE* variants are the largest monogenetic risk modifiers of Alzheimer's disease and have

more recently been implicated in modulating additional phenotypes ranging from cardiovascular disease to tumor and infection immunity [31–35].

Three highly prevalent variants of *APOE* exist, termed *APOE2*, *APOE3*, and *APOE4*, which are defined by two SNPs. The first SNP (rs429358) is a T to C substitution on chromosome 19 at position 44908684 leading to a change of a TGC codon to CGC, resulting in a cysteine to arginine substitution (*APOE*^{C130R}). The second SNP (rs7412) corresponds to a C to T substitution at position 44908822 (*APOE*^{R176C}). The most common *APOE3* variant is defined by the rs429358(T)/rs7412(C) genotype, while the *APOE4* and *APOE2* variants correspond to rs429358(C)/rs7412(C) and rs429358(T)/rs7412(T) genotypes, respectively. We used BEScreen's variant mode to identify guides that mediate the introduction of these mutations in the reference *APOE3* background (Fig. 2A). BEScreen identified one guide suitable for the rs429358 edit. Notably, inspection of the guide's characteristics in the tabular output revealed that this guide introduces a second edit changing the upstream GTG codon to GCG, resulting in an additional missense mutation. For the rs7412 SNP, BEScreen identified two suitable guides mediating the desired *APOE*^{R176C} mutation. By entering variants at the protein instead of nucleotide level (i.e. *APOE*-C130R instead of rs429358), BEScreen searches for whether additional edits in any of the other two bases making up the affected codon achieve the intended protein consequences, which here was not the case.

The implications of other, less common *APOE* variants remain largely unexplored, rendering their systematic interrogation intriguing [36]. To design an sgRNA library introducing

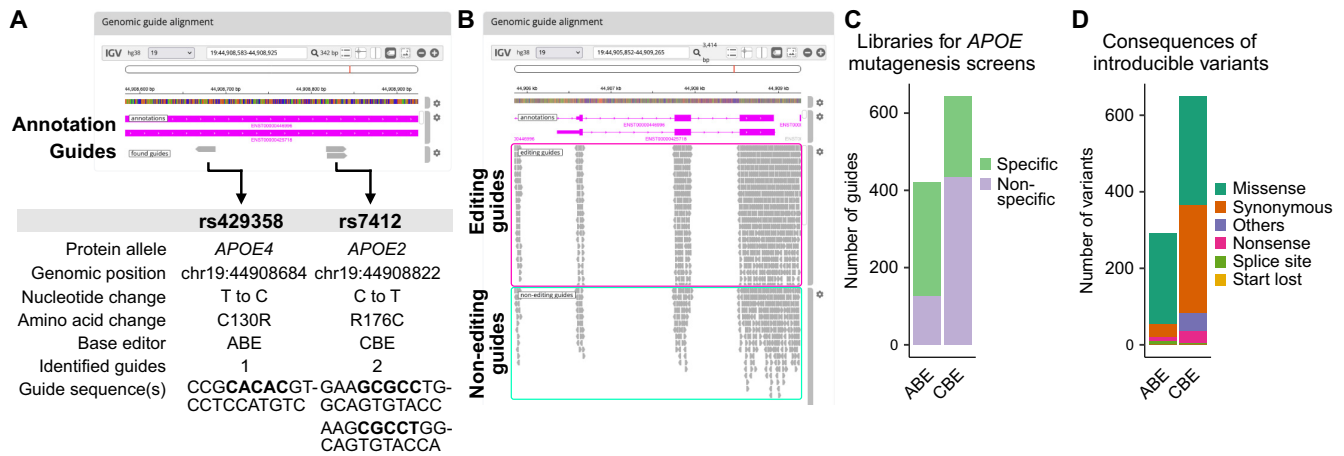


Figure 2. Case study: design of individual guides and sgRNA libraries for variants in the *APOE* gene. **(A)** Overview of the output of BEscreen used in variant mode to identify guides introducing the clinically significant rs429358 and rs7412 variants in the *APOE* gene. **(B)** Genomic alignment of a subset of the guides identified by BEscreen's gene mode to identify all possible editing and non-editing guides in the *APOE* coding sequence. **(C)** Number of editing guides identified for ABE (adenine base editor) and CBE (cytosine base editor) as shown in panel (B), indicating whether guides are predicted to be specific (mediate only one edit) or not. **(D)** Overview of the consequences mediated by editing guides shown in panel (B).

all possible edits into *APOE*, we next leveraged BEscreen's gene mode (Fig. 2B). Using “*APOE*” as input, BEscreen designed 1065 editing guides, 506 of which mediated only one edit and were thus annotated as specific (Fig. 2C). As detailed in BEscreen's output, the consequences mediated by all edits included 526 missense, 313 synonymous, 11 splice site, 48 nonsense, 2 start lost, and 2 stop lost edits (Fig. 2D). In addition, 402 binding but non-editing guides were identified. These non-editing guides and guides conferring synonymous edits are valuable as negative controls in a BE screen, while guides mediating splice site disruptions or nonsense mutations may serve as positive controls. Thus, BEscreen designed a complete library for evaluating known and unknown variants in the *APOE* gene with minimal and convenient input requirements.

Finally, we used the genomic region spanning the *APOE* gene as input for BEscreen's region mode (“19:44905791–44909393”), resulting in 3150 and 1247 editing and non-editing guides, respectively. Notably, these guides cover the entire gene region, including introns and 5' and 3' untranslated regions (UTRs) in contrast to the output of the gene mode.

Discussion

CRISPR-Cas9-coupled base editing has enabled the rapid adoption of functional screens to interrogate the causal impact of genetic variants at scale [9, 10]. BEscreen is a versatile toolkit to design sgRNAs from small-scale experiments focusing on pre-specified single nucleotide and amino acid variants to large-scale near-saturation screens of genes and transcripts or genomic regions. BEscreen is uniquely suited to design positive and negative controls by allowing for sgRNA selection by biological consequence, including synonymous or non-editing (for negative controls) and nonsense or splice site mutations (for positive controls). BEscreen features three modes that enable BE screens focused on variants, genes, or genomic regions. Future iterations of BEscreen may include more refined quality control metrics as biological characteristics of specific BEs become better characterized and will support additional

nucleotide changes as novel nucleotide editing enzymes are discovered or engineered. In sum, BEscreen provides a comprehensive tool for the design of sgRNA libraries for base editing screens.

Acknowledgements

We thank Dieter Beule for helpful discussions and technical insight. We thank members of the Ostendorf lab and the lab of Dimitrios L. Wagner for extensive testing of BEscreen and helpful feedback.

Supplementary data

Supplementary data is available at NAR online.

Funding

This work was supported by the German Cancer Aid [Max Eder program reference 70114327 to B.N.O.], a Fritz Thyssen project grant [reference 10.24.1.010MN to B.N.O.], a scholarship from Guangzhou Elite Project [to S.L.], and a fellowship of the Charité-BIH Digital Clinician Scientist Program [to B.N.O.]. Funding to pay the Open Access publication charges for this article was provided by the publications fund of the German Research Foundation and Charité.

Conflict of interest

None declared.

Data availability

The BEscreen web server is available at bescreen.ostendorflab.org. This website is free and open to all users and there is no login requirement. A command line tool for local installation is available at github.com/ostendorflab/bescreen and at doi.org/10.6084/m9.figshare.28616528.

References

1. Yu Z, Coorens THH, Uddin MM *et al.* Genetic variation across and within individuals. *Nat Rev Genet* 2024;25:548–62. <https://doi.org/10.1038/s41576-024-00709-x>
2. Young AI, Benonisdottir S, Przeworski M *et al.* Deconstructing the sources of genotype–phenotype associations in humans. *Science* 2019;365:1396–400. <https://doi.org/10.1126/science.aax3710>
3. Camp JG, Platt R, Treutlein B. Mapping human cell phenotypes to genotypes with single-cell genomics. *Science* 2019;365:1401–5. <https://doi.org/10.1126/science.aax6648>
4. Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* 2020;38:824–44. <https://doi.org/10.1038/s41587-020-0561-9>
5. Gaudelli NM, Komor AC, Rees HA *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 2017;551:464–71. <https://doi.org/10.1038/nature24644>
6. Komor AC, Kim YB, Packer MS *et al.* Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 2016;533:420–4. <https://doi.org/10.1038/nature17946>
7. Zhao D, Li J, Li S *et al.* Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* 2021;39:35–40. <https://doi.org/10.1038/s41587-020-0592-2>
8. Ye L, Zhao D, Li J *et al.* Glycosylase-based base editors for efficient T-to-G and C-to-G editing in mammalian cells. *Nat Biotechnol* 2024;42:1538–47. <https://doi.org/10.1038/s41587-023-02050-w>
9. Hanna RE, Hegde M, Fagre CR *et al.* Massively parallel assessment of human variants with base editor screens. *Cell* 2021;184:1064–1080. <https://doi.org/10.1016/j.cell.2021.01.012>
10. Cuella-Martin R, Hayward SB, Fan X *et al.* Functional interrogation of DNA damage response variants with base editing screens. *Cell* 2021;184:1081–1097. <https://doi.org/10.1016/j.cell.2021.01.041>
11. Schmidt R, Ward CC, Dajani R *et al.* Base-editing mutagenesis maps alleles to tune human T cell functions. *Nature* 2024;625:805–12. <https://doi.org/10.1038/s41586-023-06835-6>
12. Walsh ZH, Shah P, Kothapalli N *et al.* Mapping variant effects on anti-tumor hallmarks of primary human T cells with base-editing screens. *Nat Biotechnol* 2025;43:384–95. <https://doi.org/10.1038/s41587-024-02235-x>
13. Coelho MA, Strauss ME, Watterson A *et al.* Base editing screens define the genetic landscape of cancer drug resistance mechanisms. *Nat Genet* 2024;56:2479–92. <https://doi.org/10.1038/s41588-024-01948-8>
14. Coelho MA, Cooper S, Strauss ME *et al.* Base editing screens map mutations affecting interferon- γ signaling in cancer. *Cancer Cell* 2023;41:288–303. <https://doi.org/10.1016/j.ccell.2022.12.009>
15. Belli O, Karava K, Farouni R *et al.* Multimodal scanning of genetic variants with base and prime editing. *Nat Biotechnol* 2024; <https://doi.org/10.1038/s41587-024-02439-1>
16. Siegner SM, Karasu ME, Schröder MS *et al.* PnB Designer: a web application to design prime and base editor guide RNAs for animals and plants. *BMC Bioinformatics* 2021;22:101. <https://doi.org/10.1186/s12859-021-04034-6>
17. Chen C-L, Rodiger J, Chung V *et al.* SNP-CRISPR: a web tool for SNP-specific genome editing. *G3* 2020;10:489–94. <https://doi.org/10.1534/g3.119.400904>
18. Dandage R, Després PC, Yachie N *et al.* beditor: a computational workflow for designing libraries of guide RNAs for CRISPR-mediated base editing. *Genetics* 2019;212:377–85. <https://doi.org/10.1534/genetics.119.302089>
19. Hwang G-H, Park J, Lim K *et al.* Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics* 2018;19:542. <https://doi.org/10.1186/s12859-018-2585-4>
20. Rabinowitz R, Abadi S, Almog S *et al.* Prediction of synonymous corrections by the BE-FF computational tool expands the targeting scope of base editing. *Nucleic Acids Res* 2020;48:W340–7. <https://doi.org/10.1093/nar/gkaa215>
21. Wang Y, Gao R, Wu J *et al.* Comparison of cytosine base editors and development of the BEable-GPS database for targeting pathogenic SNVs. *Genome Biol* 2019;20:218. <https://doi.org/10.1186/s13059-019-1839-4>
22. Gaudelli NM, Lam DK, Rees HA *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol* 2020;38:892–900. <https://doi.org/10.1038/s41587-020-0491-6>
23. Neugebauer ME, Hsu A, Arbab M *et al.* Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat Biotechnol* 2023;41:673–85. <https://doi.org/10.1038/s41587-022-01533-6>
24. Lam DK, Feliciano PR, Arif A *et al.* Improved cytosine base editors generated from TadA variants. *Nat Biotechnol* 2023;41:686–97. <https://doi.org/10.1038/s41587-022-01611-9>
25. Walton RT, Christie KA, Whittaker MN *et al.* Unconstrained genome targeting with near-PAMless engineered CRISPR–Cas9 variants. *Science* 2020;368:290–6. <https://doi.org/10.1126/science.aba8853>
26. Sheriff A, Guri I, Zebrowska P *et al.* ABE8e adenine base editor precisely and efficiently corrects a recurrent COL7A1 nonsense mutation. *Sci Rep* 2022;12:19643. <https://doi.org/10.1038/s41598-022-24184-8>
27. Stovner EB, Sætrom P. PyRanges: efficient comparison of genomic intervals in Python. *Bioinformatics* 2020;36:918–9. <https://doi.org/10.1093/bioinformatics/btz615>
28. McLaren W, Gil L, Hunt SE *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016;17:122. <https://doi.org/10.1186/s13059-016-0974-4>
29. Camacho C, Coulouris G, Avagyan V *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>
30. Robinson JT, Thorvaldsdottir H, Turner D *et al.* igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* 2023;39:btac830. <https://doi.org/10.1093/bioinformatics/btac830>
31. Strittmatter WJ, Saunders AM, Schmechel D *et al.* Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci USA* 1993;90:1977–81. <https://doi.org/10.1073/pnas.90.5.1977>
32. Mahley RW. Apolipoprotein E: from cardiovascular disease to neurodegenerative disorders. *J Mol Med* 2016;94:739–46. <https://doi.org/10.1007/s00109-016-1427-y>
33. Corder EH, Saunders AM, Risch NJ *et al.* Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* 1994;7:180–4. <https://doi.org/10.1038/ng0694-180>
34. Ostendorf BN, Bilanovic J, Adaku N *et al.* Common germline variants of the human APOE gene modulate melanoma progression and survival. *Nat Med* 2020;26:1048–53. <https://doi.org/10.1038/s41591-020-0879-3>
35. Ostendorf BN, Patel MA, Bilanovic J *et al.* Common human genetic variants of APOE impact murine COVID-19 mortality. *Nature* 2022;611:346–51. <https://doi.org/10.1038/s41586-022-05344-2>
36. Rabinovici GD, Dubal DB. Rare APOE missense variants-can we overcome APOE ϵ 4 and Alzheimer disease risk? *JAMA Neurol* 2022;79:649–51. <https://doi.org/10.1001/jamaneurol.2022.0854>