

pegIT - a web-based design tool for prime editing

Mads Valdemar Anderson¹*, Jakob Haldrup, Emil Aagaard Thomsen, Jonas Holst Wolff and Jacob Giehm Mikkelsen¹*

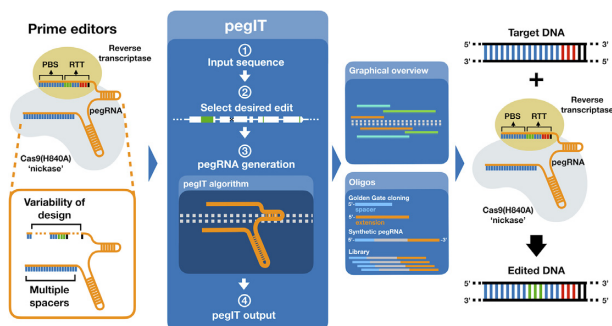
Department of Biomedicine, Aarhus University, Aarhus C, 8000, Denmark

Received March 09, 2021; Revised April 12, 2021; Editorial Decision April 26, 2021; Accepted May 11, 2021

ABSTRACT

Prime editing (PE) is a novel CRISPR-derived genome editing technique facilitating precision editing without double-stranded DNA breaks. PE, mediated by a Cas9-reverse transcriptase fusion protein, is based on dual-functioning prime editing guide RNAs (pegRNAs), serving both as guide molecules and as templates carrying the desired edits. Due to such diverse functions, manual pegRNA design is a subject to error and not suited for large-scale setups. Here, we present pegIT, a user-friendly web tool for rapid pegRNA design for numerous user-defined edits, including large-scale setups. pegIT is freely available at <https://pegit.giehmilab.dk>.

GRAPHICAL ABSTRACT



INTRODUCTION

CRISPR/Cas9 (1,2) is widely used in molecular biology and biomedical research. The technology was initially developed for disruption of single genes through indels originating from non-homologous end-joining (NHEJ), leading to creation of genome-wide knockout libraries (3). In addition, precision edits mediated by addition of templates for homology-directed repair (HDR) have innovated therapeutic gene editing, whereas the development of catalytically impaired and fusion Cas9 variants has further expanded

the CRISPR toolbox. Hence, newer genome editing methods exploit the DNA-binding properties of CRISPR effectors to direct DNA-modifying enzymes to the desired target locus. Fusion of nucleobase deaminase enzymes to catalytically impaired Cas9 variants has resulted in base editors (4–6) that can install the four transition mutations (A→G, G→A, C→T and T→C) and the two transversion mutations (C→G and G→C) with higher efficacy through editing mechanisms that are less likely to introduce unwanted mutations than repair pathways involving HDR.

More recently, described as a ‘search-and-replace’ genome editing technology, prime editing (PE) enables targeted insertions, deletions and all possible transitions and transversions without double-stranded breaks (DSBs) or the need for exogenous donor DNA templates. Prime editors (PEs) consist of a Cas9-H840A nickase (Cas9n) fused to a reverse transcriptase (RT). The Cas9n/RT complex is programmed by a prime editor guide RNA (pegRNA), which consists of three elements: (i) a conventional protospacer that specifies the genomic target locus, (ii) a primer binding site (PBS) that serves as the template for the RT and (iii) the RT template that contains the desired genomic edit (Figure 1). The Cas9 nickase, guided by the target-specific pegRNA, nicks the PAM-containing strand to form a single-strand DNA (ssDNA) flap. The PBS, which is complementary to the ssDNA flap, hybridizes and initiates reverse transcription of the RT template, thus incorporating nucleotides that contain the desired edit. For increased efficiency, an additional nicking single guide RNA (nsgRNA) can be added, a method termed PE3. Use of nsgRNAs that are specific for the edited strand prevents concurrent nicking and hence, formation of DSBs and indels, a strategy termed PE3b. Prime editing has enormous potential for treatment of genetic diseases, generation of transgenic plants and mice (7–9), and multiple other applications.

Compared to contemporary CRISPR/Cas9 technology and gene editing using base editors, which both rely on sgRNAs for precise targeting, prime editing is based on guiding RNA molecules containing the edit sequence. Hence, whereas previous CRISPR technologies require only design of oligos for the protospacer, pegRNAs also require design

*To whom correspondence should be addressed. Tel: +45 2361 7253. Email: giehm@biomed.au.dk

Correspondence may also be addressed to Mads Valdemar Anderson. Email: madsanderson@gmail.com

Present address: Mads Valdemar Anderson, Global Biopharm Rare Disease Research, Novo Nordisk, Måløv, 2760, Denmark.

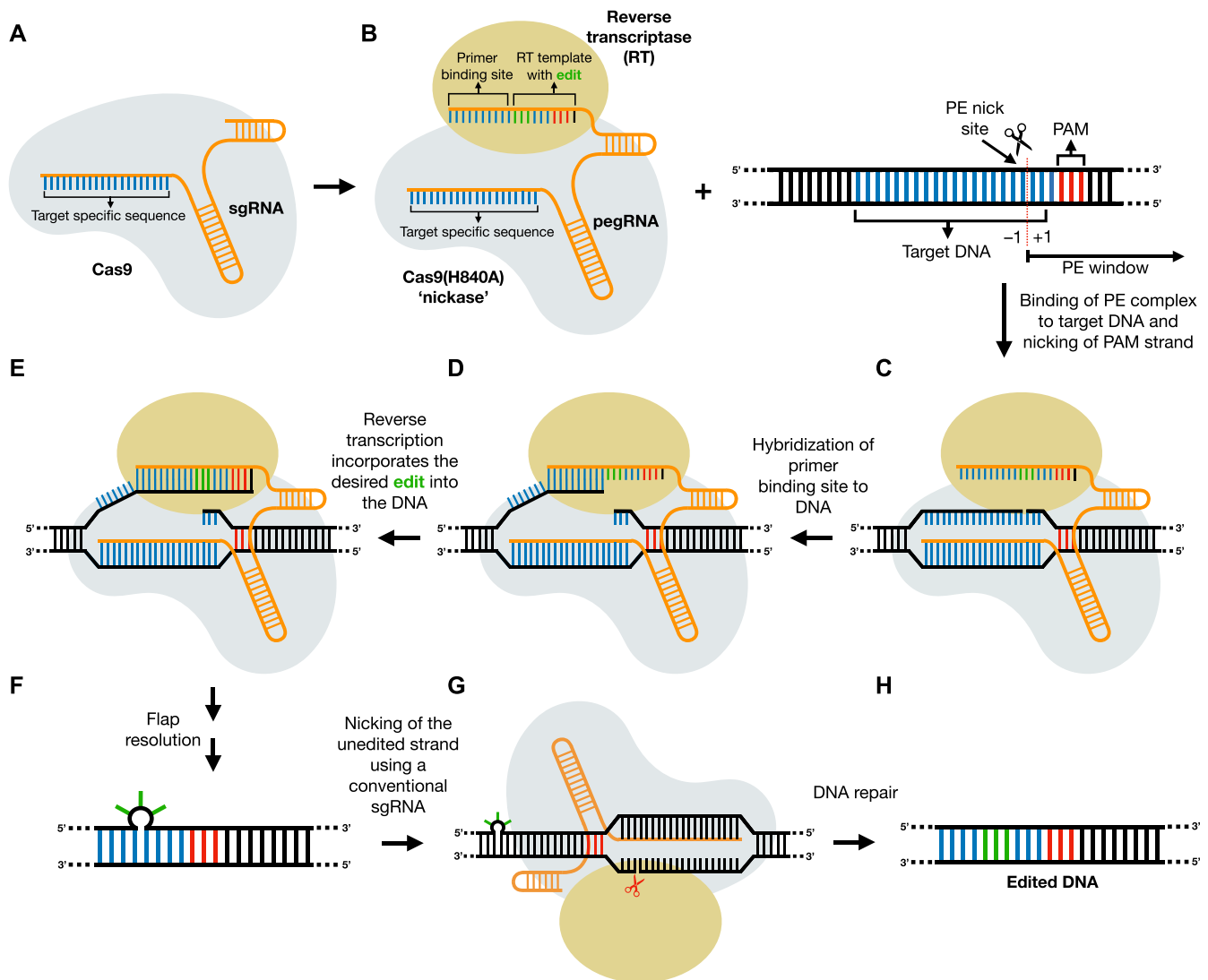


Figure 1. Schematic representation of prime editing. (A) Prime editors (PEs) consist of a fusion protein of Cas9 nickase and a reverse transcriptase (RT). This complex is coupled with a prime editing guide RNA (pegRNA), which directs the PE complex to the target sequence and encodes the desired genetic modification. (B) The PE complex binds to the target locus and nicks the PAM strand. (C) Hybridization of the 3' end of the pegRNA and 3' end of the PAM strand; the primer binding site (PBS) allows initiation of reverse transcription by the RT. (D) The RT reverse-transcribes the pegRNA sequence into the 3' PAM strand. (E) The PE complex dissociates, and cellular repair mechanisms resolve DNA flaps to introduce the edit. (F) The PE complex dissociates, and cellular repair mechanisms resolve DNA flaps to introduce the edit as a heteroduplex. (G) Nicking of the unedited strand biases repair towards incorporation of the edit, resulting in an edited homoduplex (H).

of the extension sequence consisting of a primer binding site (PBS) and RT template (RTT), which encodes the desired edit. With added functional properties of the pegRNA, this further complicates the design process.

The broad adoption of genome editing tools requires tools for rapid design of experiments, and other pegRNA design tools have recently become available. With pegFinder (10) and PrimeDesign (11), users can design pegRNAs by manually pasting the wild-type and edited sequence, either as a separate or a combined input. This requires the generation of intended edit and flanking sequences, which may increase the risk of user-introduced errors and reduce throughput. Furthermore, pegFinder does not support batch designs, and PrimeDesign only supports the wild-type NGG PAM SpCas9 nickase domain. Additionally,

these tools can only design oligos for the cloning method described by Anzalone and colleagues (7). We aimed to develop a tool that integrates reference genomes with a simple editing nomenclature to enable rapid, efficient design of pegRNAs for intended edits. Also, we aimed at generating a tool with a modular structure allowing rapid adjustments for new prime editing variants and CRISPR scoring rules.

We have developed pegIT (<http://pegit.giehlmlab.dk>), a web-tool and command-line utility for automated design of pegRNAs, for a range of user-defined edits. The user can select genes and transcripts for organisms available in the database or provide a user-defined sequence. The user can then choose between various kinds of edits, and design pegRNAs to introduce or repair the specific variant. Additionally, to facilitate design of experiments, pegIT designs

primers for PCR amplification of the target locus. Even though prime editing has been reported to induce a lower rate of off-target editing compared to Cas9, off-target editing can occur (7), and such potential off-target sites of pegRNA and nsgRNA spacers are also reported in the software (Figure 2).

MATERIALS AND METHODS

Inputs for pegIT

pegIT can either design pegRNAs for user-defined edits or for variants in ClinVar. For design of pegRNAs for ClinVar variants, the user can search for the desired variant and design pegRNAs to install or repair the variant. The ClinVar search is enabled through the BioPython package (12)

For design of user-defined edits, pegIT comprises a database of genes and transcripts for which the user can easily design pegRNAs to introduce or repair the edit(s) of interest. By selecting an organism and searching for the gene of interest, the corresponding transcripts are revealed, and the desired sequence can be selected. Alternatively, the user can paste a custom DNA sequence or provide genomic coordinates. When sequence identifiers or genomic coordinates are used as input, the corresponding genomic region is retrieved using twoBitToFa (13). Once a sequence has been selected, the user is presented with a form to enter the desired type of edit. When using custom sequences, we recommend pasting at least 200 nucleotides on either side of the desired edit.

Specifying the desired edit

Once a sequence has been selected, the user is presented with a form to enter the desired type of edit. pegIT currently supports insertions, deletions, and substitutions, and amino acid alterations can be specified for protein-coding transcripts without prior knowledge of the codon sequences. Finally, a catalogue of common epitope tags may be incorporated at specific positions if desired. Based on the type of edit, the user inputs the desired alteration. To assist input, the selected sequence is visualized. Here, the user can search for DNA sequences and select the specific position(s) to edit. Furthermore, exons and translations are presented in the sequence view, enabling easy use of the amino acid input format. Additional options include the option to design pegRNAs to repair the chosen edit and/or introduce silent mutations in the PAM for the corresponding pegRNA to increase editing efficiency (7).

Alternatively, for users desiring to design pegRNAs for many edits, it is possible to upload a text file specifying the desired edits. Finally, the user selects the desired PE nuclease domain, e.g. SpCas9 and cloning vector. Following selection of edits, pegIT returns pegRNAs and nsgRNAs to introduce or repair the specified variants.

Selection of pegRNA spacers and design of extension sequences

Once an edit has been specified, pegIT identifies potential spacers that can locate the PE complex in proximity to the

target sequence. When sequence identifiers or genomic coordinates are used as input, the corresponding genomic region is retrieved using twoBitToFa (13). The sense wild-type sequence is searched for spacers that nick upstream of the desired change, and the antisense strand is searched for spacers that nick downstream. pegIT returns up to 5 spacers (default), or a user-specified number of spacers, for further design of extension sequences. pegRNA spacers are scored based on whether the PAM is disrupted (i) by the edit, (ii) by the distance to the edit, and in the case of SpCas9, (iii) by the on-target score of the spacer.

With default settings, extension sequences are designed based on rules established previously (7). If specified by the user, pegIT first tries to introduce silent mutations to disrupt the PAM sequence of the spacer. pegIT starts with a PBS of 13 nucleotides, if the G/C content deviates from 40 to 60%, an extra nucleotide is recursively added up to a maximum length of 20 nucleotides. If no identified PBS lies within this G/C content range, pegIT selects the PBS sequence by scoring based on G/C content deviation from 50% with sequence length as a tie solver, preferring shorter sequences. The reverse transcriptase template (RTT) is designed to have a minimum of 10 homologous bases downstream of the last edited nucleotide. For larger edits, this is extended to be twice the length of altered nucleotides, up to a maximum of 34 bases. Based on the choice of PE nuclease domain and cloning vector, choice and design of pegRNAs is adapted to accommodate these, e.g. for SpCas9, RTTs that result in extension sequences starting with a 'C' are discarded, as these have been reported to result in lower efficiency, presumably by disrupting the sgRNA structure (7). For expression by U6 promoters, pegRNAs comprising a TTTT motif are discarded to ensure efficient pegRNA production by reducing the risk of premature termination during pegRNA synthesis.

Selection of nsgRNAs

Nicking the opposite strand can result in higher editing efficiencies. To identify such nsgRNAs, all spacers in the edited sequence within 100 nucleotides of the edit are identified. Utilizing nsgRNAs that are specific for the edited sequence minimizes concurrent nicking and introduction of double-stranded breaks (7), a strategy termed PE3b. pegIT prioritizes nsgRNAs that are mismatched in the wild-type sequence. For SpCas9, nsgRNAs with mismatches are scored using CFD (14) against the wild-type sequence. The best candidate is then selected based on lowest score on wild-type sequence; nonspecific sgRNAs are scored based on having a distance >50 from the pegRNA nicking site, followed by their on-target score (14).

Off-target search

The pegRNA and nsgRNA spacer sequences are mapped to the reference genome of the selected organism using Bowtie (15) with up to three mismatches, using the command 'bowtie -v 3 -a -best -y'. The DNA sequences corresponding to the matched hits including 3 nucleotides downstream are retrieved using twoBitToFa (13) and filtered to discard hits that are not adjacent to a NAG or NGG PAM.

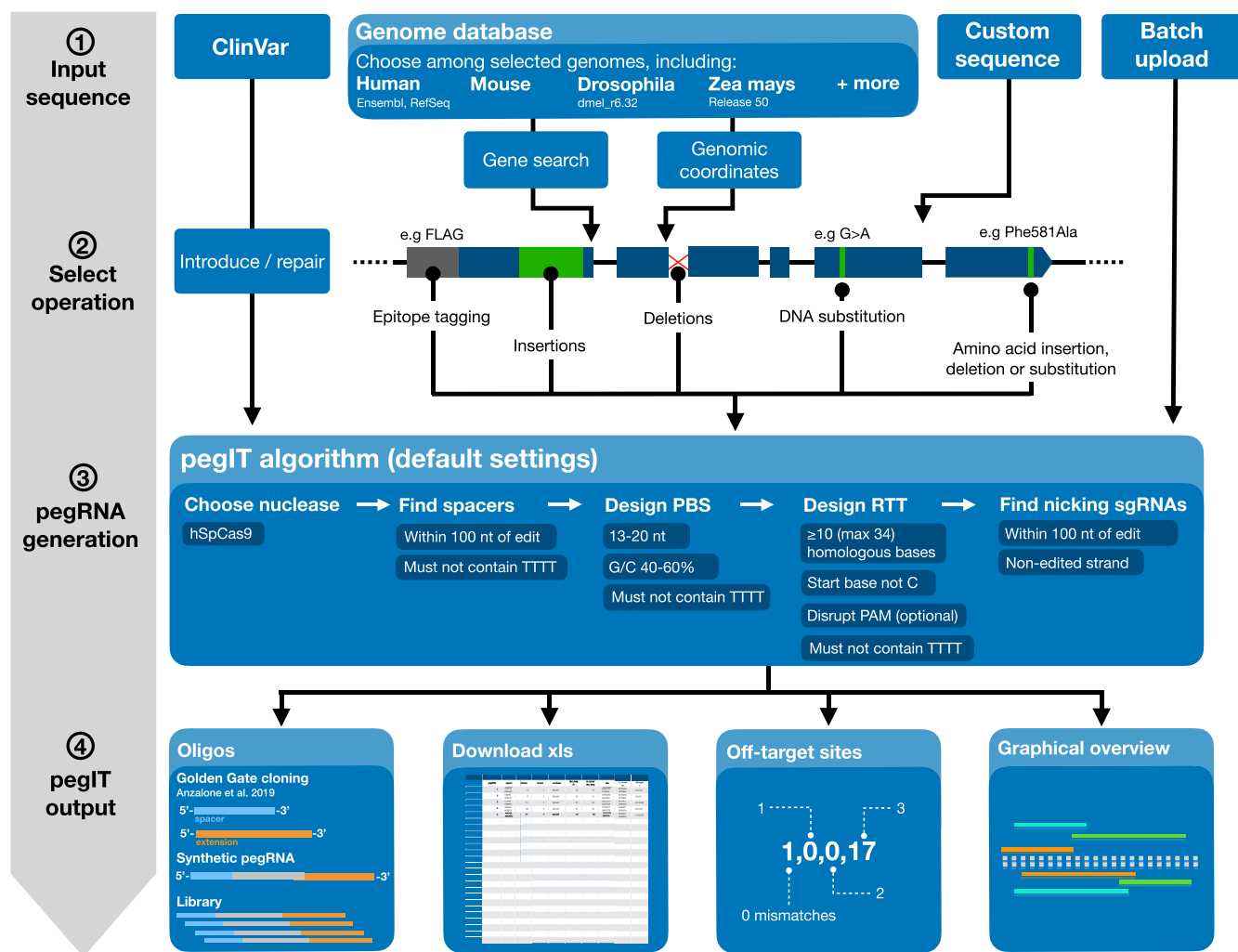


Figure 2. pegIT concept and work chart. The user selects a DNA sequence from the database or supplies a custom sequence and a specification of the desired edit to perform. Alternatively, variants from ClinVar can be retrieved, and the user can select to introduce or repair the specific variant. pegIT further supports upload of a text file specifying DNA sequences and desired edits for designing multiple pegRNAs. pegIT then identifies candidate spacers for pegRNA design, designs primer binding site and reverse transcriptase template extension sequences, and identifies nsgRNAs for PE3. The pegIT algorithm is here shown with the default settings; however, customization of settings for the pegIT algorithm is available through an advanced options tab. Following pegRNA design, pegIT provides an easy-to-use graphical overview of the designed pegRNAs and nsgRNAs as well as oligo sequences for various cloning protocols and synthetic pegRNA sequences.

Primer design

Primers are designed using Primer3 (16). With default parameters, primer size 18–25nt, optimum 22nt, product size 150–300 bp, primer T_m 57–63°C, optimum 60°C. To check primer specificity, 3 primer pairs, forward-forward, forward-reverse, and reverse-reverse for each primer pair are mapped using Bowtie in paired mode, with up to 3 mismatches and size between one third of the minimum product size and 3 times the largest product size. With default settings, the command ‘bowtie -v 3 -k 50 -best -I 50 -X 900 -y’ is used.

Results page

On the results page the user can inspect the designed pegRNAs. For jobs with multiple edits, a list of processed edits is shown. The highest ranking pegRNA design can be inspected directly in this table, and alternative pegRNA de-

signs can be examined by clicking the Details button. Single edit jobs are sent directly to the detail page. When the pegRNAs have been designed, it is possible to download an excel file containing the results. On the detail page, pegRNA spacers and their extensions are visualized for rapid validation. Further information is provided in the accompanying table, including the distance between Cas9 nicking and the desired edit, and whether the pegRNA PAM is disrupted by the edit or by silent mutations (optionally). Oligos for cloning the suggested pegRNA can be obtained directly from this table. As prime editing is a new technology, the optimal design rules for pegRNAs are not yet completely established, and locus-specific optimization may be required. To facilitate easy testing of different pegRNA extensions and nsgRNAs, pegIT provides a list of alternative nsgRNA candidates and combinations of PBS and RTT sequences, as well as oligonucleotides for cloning. These can be found in the view for the respective pegRNA or in the

downloadable excel file. The excel file contains a summary sheet listing the best ranked designs for each edit. Alternative pegRNA spacers and extensions can be found in the separate sheet for each edit.

To optimize guide RNA design, pegIT reports on-target scores previously developed for SpCas9 knockout experiments for both pegRNAs and nsgRNAs (14). For selection of PE3b nsgRNAs, pegIT also reports the off-target score for the wild-type allele for SpCas9 (14). Furthermore, the number and position of predicted binding sites for the pegRNAs and nsgRNAs in the selected organism are predicted with up to three mismatches. A detailed overview of predicted binding sites can be found in the respective detail views.

Command-line version

In addition to the webserver, the source code can be downloaded and run locally both as a command-line or as web application, which is suited for larger queries, sensitive data, or in-house cell lines. All functionality available on the web server is also available in the command-line version.

DISCUSSION

Since the adaptation of CRISPR/Cas9 for genome editing (1,17–19), our understanding of factors governing design of sgRNAs for efficient genome editing has rapidly evolved. Similarly, base editors (4) have undergone rapid development since their initial inception. To accommodate the anticipated development of the prime editing technology, pegIT was designed in a modular fashion. Nucleases, cloning plasmids, design rules and edits are independently designed as ‘plugins’, enabling easy adjustment for optimized pegRNA scoring rules and novel prime editing variants. This allows maintenance and further development of pegIT. Users can request the addition of edits, nuclease domains, organisms and cloning rules. In future updates, besides addition of new organisms, novel PE variants, and design and cloning rules, edits for phenotype changes could also be envisioned, e.g. by incorporation of edits for tuning gene expression based on models for prediction of translation rates (20).

Altogether, pegIT provides a novel platform for rapid and user-friendly design of pegRNAs and nsgRNAs for the implementation of prime editing in basic experimentation and therapy.

DATA AVAILABILITY

The source code for the pegIT software is available at <https://github.com/dkmva/pegit> and <https://github.com/dkmva/pegit-front>.

FUNDING

Lundbeck Foundation [R230-2016-2986, R324-2019-1832]. Funding for open access charge: Aarhus University.

Conflict of interest statement. None declared.

REFERENCES

- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–87.
- Rees, H.A. and Liu, D.R. (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.*, **19**, 770–788.
- Kurt, I.C., Zhou, R., Iyer, S., Garcia, S.P., Miller, B.R., Langner, L.M., Grünwald, J. and Joung, J.K. (2021) CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.*, **39**, 41–46.
- Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M.A., Rosser, S.J., Bi, C. and Zhang, X. (2021) Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.*, **39**, 35–40.
- Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.
- Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., Wang, Y., Anzalone, A.V., Raguram, A., Doman, J.L. *et al.* (2020) Prime genome editing in rice and wheat. *Nat. Biotechnol.*, **38**, 582–585.
- Liu, Y., Li, X., He, S., Huang, S., Li, C., Chen, Y., Liu, Z., Huang, X. and Wang, X. (2020) Efficient generation of mouse models with the prime editing system. *Cell Discov.*, **6**, 27.
- Chow, R.D., Chen, J.S., Shen, J. and Chen, S. (2020) A web tool for the design of prime-editing guide RNAs. *Nat. Biomed. Eng.*, **5**, 190–194.
- Hsu, J.Y., Grünwald, J., Szalay, R., Shih, J., Anzalone, A.V., Lam, K.C., Shen, M.W., Petri, K., Liu, D.R., Joung, J.K. *et al.* (2021) PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat. Commun.*, **12**, 1034.
- Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
- Karolchik, D., Hinriks, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. and Kent, W.J. (2004) The UCSC table browser data retrieval tool. *Nucleic Acids Res.*, **32**, 493–496.
- 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.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res.*, **40**, e115.
- 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 (New York, N. Y.)*, **339**, 819–823.
- 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.
- Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. (2013) RNA-programmed genome editing in human cells. *eLife*, **2**, e00471.
- Sample, P.J., Wang, B., Reid, D.W., Presnyak, V., McFadyen, I.J., Morris, D.R. and Seelig, G. (2019) Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.*, **37**, 803–809.