Molecular Therapy
Original Article

Engineered prime editors with PAM flexibility

Jiyeon Kweon,[1,2] Jung-Ki Yoon,[3,4] An-Hee Jang,[1,2] Ha Rim Shin,[1,2] Ji-Eun See,[1,2] Gayoung Jang,[1,2] Jong-Il Kim,[5,6] and Yongsub Kim[1,2]

[1]Department of Biomedical Sciences, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; [2]Stem Cell Immunomodulation Research Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; [3]Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea; [4]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea; [5]Genome Medicine Institute, Medical Research Center, Seoul National University, Seoul, Republic of Korea; [6]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Republic of Korea

Although prime editors are a powerful tool for genome editing, which can generate various types of mutations such as nucleotide substitutions, insertions, and deletions in the genome without double-strand breaks or donor DNA, the conventional prime editors are still limited to their target scopes because of the PAM preference of the *Streptococcus pyogenes* Cas9 (spCas9) protein. Here, we describe the engineered prime editors to expand the range of their target sites using various PAM-flexible Cas9 variants. Using the engineered prime editors, we could successfully generate more than 50 types of mutations with up to 51.7% prime-editing activity in HEK293T cells. In addition, we successfully introduced the BRAF V600E mutation, which could not be induced by conventional prime editors. These variants of prime editors will broaden the applicability of CRISPR-based prime editing technologies in biological research.

INTRODUCTION

CRISPR-Cas systems have been widely repurposed for targeted genome engineering in living cells and organisms.[1] Although Cas nucleases such as Cas9 and Cas12a are widely used for gene knockout and gene correction, they induce DNA double-strand breaks (DSBs),[2–6] which are associated with undesired outcomes such as unwanted genomic mutations[7] and activation of p53 in response to DNA damage.[8,9] To overcome these limitations, cytosine and adenine base editors have been developed for targeted C:G to T:A and A:T to G:C conversion without DNA DSBs.[10–14] Recently, Anzalone et al.[15] developed prime editors (PEs), which can write new genetic information without DSBs or donor DNA. PE2 utilizes an engineered Moloney murine leukemia virus (M-MLV) reverse transcriptase (RT) fused to a *Streptococcus pyogenes* Cas9 (spCas9)-H840A nickase and prime editing guide RNA (pegRNA) to manipulate the target site. PEs allow the introduction of various types of mutations such as transition, transversion, insertion, and deletion in the various organisms;[16,17] however, the NGG protospacer adjacent motif (PAM) preference of spCas9 restricts the targetable loci of PEs in the genome.

Previously, the PAM limitation of spCas9 was overcome by engineering spCas9. Joung et al.[18,19] developed the VQR and VRQR variants for NGA PAM recognition and the VRER variant for NGCG PAM recognition by structure-based directed evolution. Hu et al.[20] developed the xCas9 variant for NG PAM via phage-assisted continuous evolution. Nishimasu et al.[21] used a rationally engineered spCas9 variant (spCas9-NG) for NG PAM recognition. Recently, Walton et al.[22] generated the SpG variant, which is capable of recognizing NG PAM and a near PAM-less spCas9 variant (SpRY) by structure-guided mutagenesis. To increase the PAM availability of PEs, we developed PE2 variants using various spCas9 variants and successfully introduced more than 50 mutations in HEK293T cells using the PE2 variants.

RESULTS

Prime editing at NGN PAM sites using PE2 variants

We constructed PE2 variants having a preference for non-canonical PAMs using structure-guided engineered spCas9 variants, including the VQR and VRQR variants for NGA PAM, and the VRER variant for NGCG PAM from Joung et al.,[18,19] the NG variant for NG PAM from Nureki et al.,[21] and both the SpG variant for NG PAM and SpRY variant for unconstrained PAM preferences from Kleinstiver et al.[22] H840A mutations were further introduced in each variant to generate nickases, and PE2 variants were constructed by changing the wild-type spCas9-H840A nickase to these nickases. We designated these variants as PE2-VQR, PE2-VRQR, PE2-VRER, PE2-NG, PE2-SpG, and PE2-SpRY.

To compare the activity of PE2 variants with that of wild-type PE2, we designed pegRNAs, which can target 35 genomic sites bearing NGN PAM. The pegRNAs had the recommended length of primer binding site (PBS) and RT template (13-nt PBS and 11- to 14-nt RT templates) and were designed to introduce three types of mutations (insertions, deletions, and substitutions) at the target genomic sites. We constructed the pegRNAs more efficiently using the single-strand DNA assembly method (Supplemental materials and methods).[23] We named the pegRNAs as follows: (1) multiple

Received 11 November 2020; accepted 19 February 2021;
https://doi.org/10.1016/j.ymthe.2021.02.022.

Correspondence: Yongsub Kim, Department of Biomedical Sciences, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea.
E-mail: yongsub1.kim@gmail.com

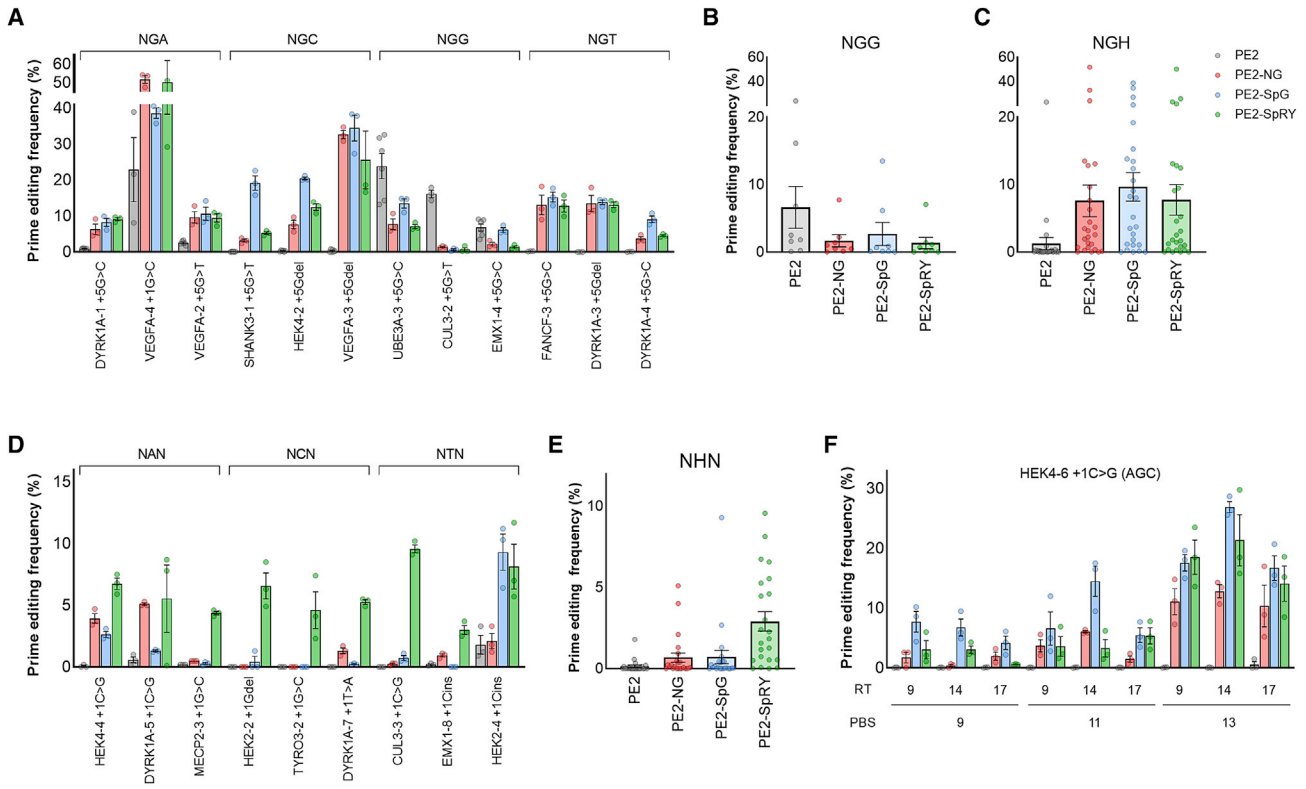


Figure 1. Prime editing in HEK293T cells by PE2 variants

(A) Representative prime editing activities at the NGA, NGC, NGG, and NGT PAM target sites. (B) Summarized prime editing activities of each PE2 variant at eight target sites with NGG PAM. (C) Summarized prime editing activities of each PE2 variant at 26 target sites with NGH PAM. (D) Representative prime editing activities at a non-NGN PAM site (NAN, NCN, and NTN PAM). (E) Summarized prime editing activities of each PE2 variant at 23 target sites with NHN PAM. (F) Prime editing activities at the HEK4-6+1C>G site with pegRNAs of various lengths of PBS and RT templates. The PAM sequence is shown in parentheses. The numerical values of prime editing activities are listed in Table S1, and information for pegRNAs is described in Table S2. Mean ± SEM of n = 3 or n = 6 independent biological replicates in (A), (D), and (F). Mean ± SEM of the average prime editing frequency for each target site in (B), (C), and (E).

pegRNAs targeting the same gene were distinguished by a number after a hyphen (e.g., FANCF-1, FANCF-2), and (2) the intended mutation type and position of each pegRNA were listed next to the number (e.g., FANCF-1+5Cins, FANCF-2+5G>C). Using these pegRNAs, we examined the prime editing activity of six PE2 variants and wild-type PE2 in HEK293T cells (Figures 1A–1C and S1) by targeted deep sequencing, and we set the threshold of editing activity to above 0.5%. As expected, wild-type PE2 induced targeted mutations predominantly at the NGG PAM sites (active in 7 of 9 sites, up to 23.8% at the UBE3A-3+5G>C site) with some recognition of the NGA PAM sites (active in 4 of 9 sites, up to 22.9% at the VEGFA-4+1G>C site) and did not induce mutations at the NGC sites (active in 0 of 8 sites) or NGT sites (active in 0 of 9 sites). In contrast, the PE2-NG, PE2-SpG, and PE2-SpRY variants could edit NGN sites (PE2-NG, active in 24 of 34 sites; PE2-SpG, active in 26 of 35 sites; PE2-SpRY, active in 24 of 35 sites). In comparison with other PE2 variants, the PE2-SpG variant showed the highest activity at NGH (where H is A, C, or T) PAM sites (Figure 1C). The PE2-VQR and PE2-VRQR variants were both active at 5 of 7 NGA PAM sites. The PE2-VQR variant showed up to 32.7% activity at the VEGFA-2+5G>T site, and the PE2-VRQR variant exhibited up to 30.2% activity at the VEGFA-1+5G>C site (Figure S2). We also confirmed that the PE2-VRER variant showed a preference for NGCG PAM compared with NGG PAM (up to 30.9% activity at the VEGFA-3+5Gdel site) (Figure S3).

Prime editing at NHN PAM sites using PE2 variants

As the spCas9-SpRY variant can edit the genome in a near-PAMless manner, we investigated whether the PAM specificity of the PE2-SpRY variant was also relaxed. We additionally designed 23 pegRNAs targeting NHN PAM sites, which could introduce three types of mutations (insertions, deletions, and substitutions) at the target sites. The PE2-SpRY variant was active in 18 of 23 NHN PAM sites and induced prime editing up to 9.6% at the CUL3-3 +1C>G site, and other PE2 variants showed lower activity (PE2, active in 2 of 23 sites; PE2-NG, active in 5 of 23 sites; PE2-SpG, active in 4 of 23 sites) across NHN PAM sites (Figures 1D and 1E). Overall, we examined a total of 58 pegRNAs targeting NNN PAM sites and found that the PE2-

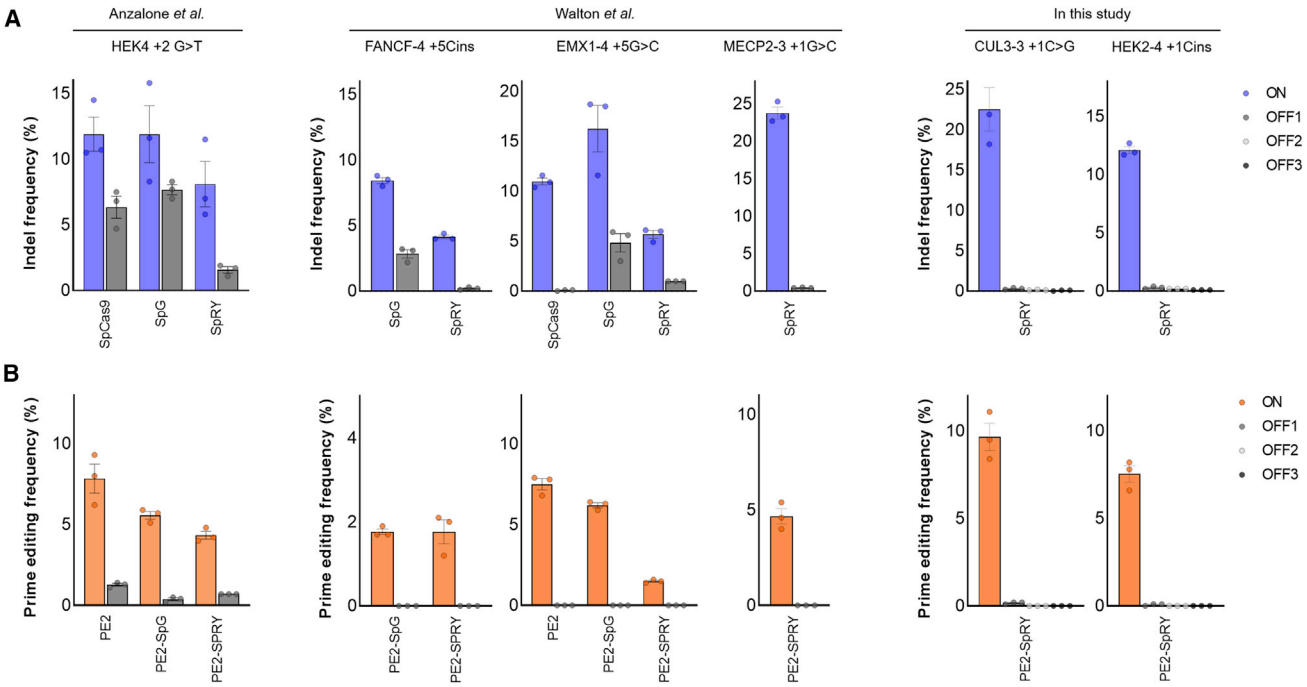


Figure 2. Off-target analysis of PE2 variants

(A and B) Six on-target sites and their 10 off-target sites were examined to have mutations with nuclease (spCas9, spCas9-SpG, spCas9-SpRY) (A) and PEs (wild-type PE2, PE2-SpG, and PE2-SpRY) (B) in HEK293T cells (detailed information is listed in Table S3). The specific ratio, which means the relative mutation frequency of on-target sites compared to that of off-target sites, is listed in Table S4. The numerical values of the mutation frequencies are described in Table S1, and information for pegRNAs is described in Table S2. Mean ± SEM of n = 3 independent biological replicates.

SpRY variant could manipulate genome in a PAM-independent manner (active in 43 of 58 sites). Collectively, these results showed that previously developed spCas9 variants could be successfully integrated into prime editing toolkits to relax the canonical PAM preferences.

Optimization of the length of PBS and RT templates of pegRNAs

Although we chose target sequences with high levels of indels generated by the spCas9 nuclease, prime editing activity varied from 0.0% to 51.7%, and many sites showed less than 0.5% activity (Table S1). As previous studies described that the prime editing activity of wild-type PE2 is affected by the lengths of the PBS and RT templates, we examined whether the composition of pegRNA could affect the prime editing activity of PE2 variants.

Two pegRNAs (targeting the HEK4-6+1C>G site with NGC PAM and the VEGFA-4+1G>C site with NGA PAM) were selected and reconstituted to have various PBS and RT template lengths. At the HEK4-6+1C>G site, although wild-type PE2 was inactive with all pegRNAs, the PE2-NG, PE2-SpG, and PE2-SpRY variants were active and showed the highest prime editing activity (12.8%, 26.9%, and 21.4%, respectively) by pegRNAs with the 13-nt PBS and 14-nt RT templates (Figure 1F). Longer PBSs were preferred for all three PE2 variants (average activity of 3.29% with a 9-nt PBS and 16.6% with a 13-nt PBS); however, there was no preference in terms of the lengths of the RT templates. These preferences were also observed at the VEGFA-4+1G>C site (Figure S4). As prime editing activity depends on the lengths of the PBS and RT templates, optimization of pegRNAs is essential to maximize the activity of the PE2 variants.

Comparison of the off-target effect of PE variants compared to their nuclease

Previously, compared with Cas9 nuclease, wild-type PE2 has lower off-target effects at already known Cas9 off-target sites.[15,24,25] Recently, off-target effects of PE2 were analyzed through the Cas9-H840 nickase-dependent manner, but since there have been no further advances in the technology to analyze PE2-dependent off-target effects, we chose several sites to validate the potential off-target effects of PE2 variants as follows: (1) HEK4+2G>T site and its off-target site, which is known as an off-target site for wild-type PE2; (2) FANCF-4+5Cins, EMX1-4+5G>C, and MECP2-3+1G>C sites and their off-target sites, which are known as off-target editing sites for the SpCas9-SpRY variant; and (3) CUL3-3+1C>G and HEK2-4 +1Cins sites, which showed high activity with the PE2-SpRY variant at NHN PAM sites, and their homologous sites (Table S3). We determined the mutation frequencies by targeted deep sequencing and calculated each specificity ratio, which means the relative mutation frequency of on-target sites compared to that of off-target sites, to compare the off-target effects of nucleases and PE variants (Figures 2A and 2B; Table S4). We found that the PEs including wild-type

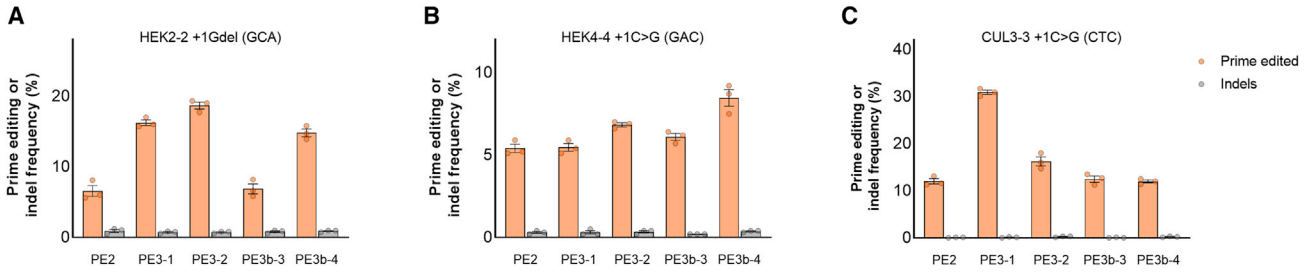


Figure 3. Prime editing in HEK293T cells by PE3 and PE3b systems

(A–C) Comparison of prime editing and indel activities at the HEK2-2+1Gdel site (A), at the HEK4-4+1C>G site (B), and at the CUL3-3+1C>G site (C). The PAM sequences are shown in parentheses. The numerical values of the mutation frequencies are described in Table S1, and the information for pegRNAs is described in Table S2. Mean ± SEM of n = 3 independent biological replicates.

PE2, PE2-SpG, and PE2-SpRY induced much lower off-target editing compared with their nucleases (spCas9, spCas9-SpG, and spCas9-SpRY). For example, at the EMX1-4 +5G>C site and its off-target site, the PE2-SpG variant showed a specificity ratio above 124, which was 36.8-fold higher than the specificity ratio of the spCas9-SpG nuclease (3.37). The high specificities of the PEs may be attributed to the requirement of additional base pairing with the PBS and RT templates to prime edit the genome.

Prime editing in PE3 and PE3b systems

Anzalone et al.[15] demonstrated an increase in prime editing activity by inducing additional nicks on the non-edited strand, which was named PE3 or PE3b. To examine whether additional nicks affect the prime editing activity of PE2 variants, we constructed additional gRNAs to incorporate PE3, which can induce additional nicks distant from pegRNA-induced nicks, or PE3b, which can induce nicks on the non-edited strand only after the resolution of the edited strand flap. At the HEK2-2+1Gdel site, PE3 improved prime editing activity up to 2.8-fold (6.6%–18.7%), and PE3b improved prime editing activity up to 2.2-fold (6.6%–14.8%) (Figure 3A). At the CUL3-3+1C>G site, PE3 increased prime editing up to 2.9-fold (12.0%–34.4%), and at the HEK4-4+1G>C site, PE3b increased prime editing up to 2.1-fold (5.4%–11.4%) (Figures 3B and 3C). Interestingly, we could not detect any increase in indels at these three sites with the PE2-SpRY variant when using different amounts of gRNAs with PE3 or PE3b (Figure S5), and we speculated that the nicking activity of spCas9-H840A variants may affect the indel frequencies of PE3 or PE3b systems. Overall, we found that PE3 and PE3b could successfully increase the prime editing activity of PE2 variants without excess indel formation.

Introduction of the BRAF V600E mutation using the PE2-SpRY variant

We performed a direct comparison of wild-type PE2 and PE2 variants for inducing the same mutations. We chose four sites bearing the NHN PAM sequence, which were active with PE2 variants, and constructed additional pegRNAs targeting those sites for wild-type PE2. We compared their prime editing activities in HEK293T cells, and targeted deep sequencing showed the advantage of PE2 variants over wild-type PE2 (Figure S6). For example, at VEGFA-3 target sites, the +5G del mutation was located 15 nt away from the nick position of wild-type PE2 and was shown to be introduced more efficiently by PE2-SpG compared with wild-type PE2 (23.8% and < 0.05%, respectively).

BRAF V600E (which is caused by a T:A to A:T transversion mutation) is a well-known driver mutation in various cancers and diseases and a therapeutic target for anti-cancer drugs.[26] Despite the broad ability of wild-type PE2 to introduce various pathogenic mutations, the distance between the pegRNA nick site and BRAF V600E target locus is 18 nt, which could not be efficiently targeted with wild-type PE2 (Figure 4A). To compare the BRAF V600E targeting efficiency of the wild-type PE2 and PE2-SpRY variant, we constructed pegRNAs with single-base resolution around the BRAF V600E target locus and examined their prime editing activity in HEK293T cells (Figure 4B). We found that PE2-SpRY improved prime editing activity up to 6.7-fold compared with the activity of wild-type PE2 (0.3%–1.8%); however, there was no significant increase in prime editing activity by the optimization of the lengths of PBS and RT templates (Figure S7). Then, we chose two active pegRNAs, BRAF-GAA and BRAF-AAA, and assessed the PE3 and PE3b systems; we found that PE3b significantly improved prime editing activity up to 11.8% with BRAF-AAA pegRNA and BRAF-PE3b-1 gRNA (Figures 4C and S8).

DISCUSSION

In summary, we generated various types of PE2 variants with PAM flexibility and analyzed the prime editing activity of the variants at more than 50 types of mutations. Compared with other PE2 variants, the PE2-SpG variant may be suitable for targeting the NGH PAM site (p values of 0.011 and 0.025 compared to the PE2-NG and PE2-SpRY variant, respectively). The PE2-SpRY variant induced prime editing without PAM restriction, albeit with reduced activity. In comparison with Cas9 nucleases or base editors, the wild-type PE2 is less dependent on a precisely positioned NGG PAM sequence for targeted genome editing due to flexibility in RT template design; however, as prime editing activities are highly dependent on the composition of pegRNAs, experimental optimization of the pegRNA is required to achieve efficient prime editing. Our PE2 variants could offer more

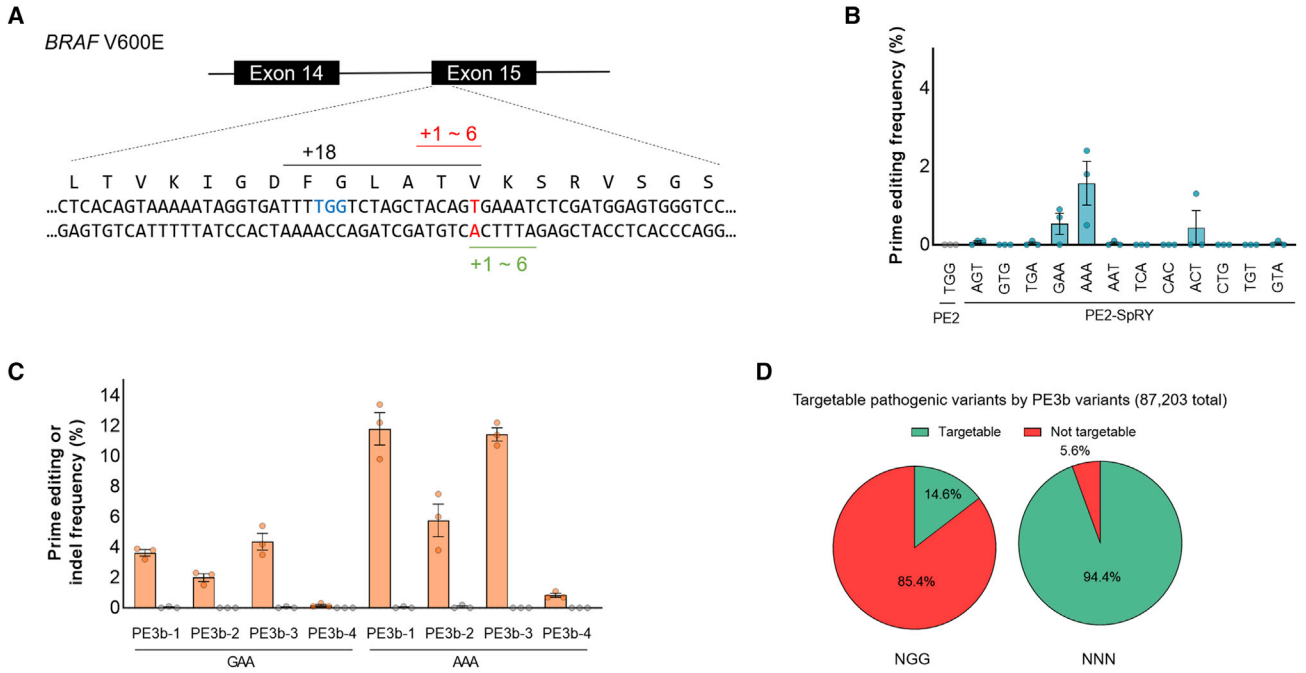


Figure 4. The BRAF V600E mutation by the PE2-SpRY variant in HEK293T cells

(A) Schematic overview of prime editing for the BRAF V600E mutation. The target thymine nucleotide is highlighted in red, and the closest PAM sequence of wild-type PE2 (TGG) is indicated in blue, which can induce nick 18 nt away from that target thymine nucleotide. The 12 pegRNAs were designed to induce nicks of 1–6 nt in both directions from the target thymine by the PE2-SpRY variant. (B) Prime editing activities of PE2 and the PE2-SpRY variant at the BRAF V600E site. (C) Prime editing activity of the PE2-SpRY variant in PE3 and PE3b systems at the BRAF V600E site. The numerical values of the mutation frequencies are described in Table S1, and the information for pegRNAs is described in Table S2. Mean ± SEM of n = 3 independent biological replicates. (D) Comparison of prime editable pathogenic variants. Among 87,203 variants in the ClinVar database, targetable pathogenic variants using PE3b pegRNAs were analyzed by a computational pipeline, which was developed by Morris et al.[27]

optimal spacer sequences regardless of PAM restriction. Using the PE2-SpRY variant, we successfully introduced the clinically significant BRAF V600E mutation with more than 10% prime editing efficiency. Furthermore, we think that the reversion of the BRAF V600E mutation to the wild-type BRAF sequence could be possible by using the PE2 variant.

Overall, the PE2 variant could enhance prime editing systems by providing more options for spacer sequences and expanding the use of the PE3b strategy with PAM flexibility. The PE2-SpRY variant could offer increased prime editable sites per human pathogenic variant (covering 94.4% of pathogenic variants) with PAM flexibility and could also increase the number of PE3b applicable sites (up to 28.8 sites per pathogenic variant) for improving prime editing efficiency without unwanted mutations (Figures S9 and S10; Table S5). Recent studies are helpful for the design of pegRNAs and analysis of PE-meditated off-target effects.[25,28] Our constructs have been shared to Addgene for the research community.

MATERIALS AND METHODS

Plasmid DNA and pegRNAs

To construct PE2 variants, we used pCMV-PE2 (Addgene plasmid #132775) plasmid DNA as vector DNA, and the coding sequences of spCas9 variants were obtained by PCR amplification from NG-ABEmax (NG variant, Addgene plasmid #124163), RTW3520 (VQR variant, Addgene plasmid #139990), RTW3160 (VRER variant, Addgene plasmid #139991), RTW3161 (VRQR variant, Addgene plasmid #139992), RTW4177 (SpG variant, Addgene plasmid #139998), and RTW4830 (SpRY variant, Addgene plasmid #139989). All of the PE2 variants used in this study were available from Addgene. The pegRNAs were cloned into the pU6-pegRNA-GG-acceptor (Addgene plasmid #132777), and the spacer sequences and RT template and PBS sequences are listed in Table S2.

Mammalian cell culture and transfection

HEK293T cells (ATCC CRL-11268) were maintained Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (Welgene). Cells were not tested for mycoplasma contamination. HEK293T cells were seeded onto a TC-treated 96-well plate (Corning Life Sciences) at 3×10^4 cells per well density 1 day before transfection. Transfection was conducted using 400 ng of plasmids and 0.6 μL of Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's protocol at approximately 60% cell confluency. Unless otherwise specified, 300 ng of PE2 and 100 ng of pegRNA-encoding plasmids were used. In the PE3 and PE3b experiments, two different doses of plasmids encoding gRNAs were used: (1) 30 ng of gRNA with 270 ng of PE2 and 100 ng of pegRNA-encoding plasmids; and (2) 100 ng of

gRNA with 200 ng of PE2 and 100 ng of pegRNA-encoding plasmids. The transfected cells were incubated at 37°C for 3 days, and genomic DNA was prepared by directed lysis of the cells using the lysis buffer (10 mM Tris-HCl with pH 7.5, 0.05% SDS, 100 μg/mL Proteinase K; QIAGEN). The cell lysate was incubated at 56°C for 30 min, followed by a 99°C 15-min incubation to inactive the Proteinase K. In off-target analysis experiments, 1.5×10^5 HEK293T cells were seeded onto TC-treated 24-well plate (Corning Life Sciences) and 1.5 μg of PE2 variants plasmids and 500 ng of pegRNA plasmids DNA were transfected, and genomic DNA was isolated using the DNeasy Blood & Tissue kit (QIAGEN) according to the manufacturer's protocol.

Targeted deep sequencing and data analysis

The target sites were amplified by a total of three rounds of PCR and subjected to the Illumina MiniSeq or iSeq 100 sequencing system as previously described.[29] Briefly, 3 μL of cell lysate or 1 μL of isolated genomic DNA was subjected to the first round of PCR, and then,1 μL of the first PCR product was used in the second round of PCR. The Illumina TruSeq HT dual index adaptor sequences were attached with index PCR primer pairs using 1 μL of second-round PCR product. The size of the PCR amplicon was confirmed in 2% agarose gel and the amplicons were subjected to 150-bp paired-end sequencing using the Illumina MiniSeq or iSeq 100 sequencing system. The paired-end reads were joined using the fastq-join tool (https://github.com/brwnj/fastq-join). Targeted deep sequencing analysis was performed using MAUND[30] (https://github.com/ibs-cge/maund), and all results were confirmed by Cas-Analyzer (http://www.rgenome.net/cas-analyzer/).[31] Substitution and indel frequencies were quantified as the percentage of total sequencing reads, and the threshold of editing activity was set to above 0.5%.

Analysis of targetable pathogenic variants in the ClinVar database

All possible pegRNAs for correcting pathogenic variants in the ClinVar database[32] (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/) were designed by using the pipeline (https://gitlab.com/sanjanalab/primeediting) developed by Morris et al.[27] with modifications. We downloaded all variants in ClinVar (released date: August 10, 2020) and selected the variants with the "pathogenic" identifier. Only single base-pair substitutions and insertions and deletions of 10 base pairs or less were calculated as targeted variants. We designed and counted the number of all possible probes for PE2 and PE3b separately with either NGG, NGN (PAM for PE2-NG and PE2-SpG), or NNN (PAM for PE2-SpRY) for each pathogenic variant. The PBS end site was located at the nick site, 3 nt downstream of the PAM, and the target editing sites were able to be placed in any position of RT template but the last. The lengths of PBS and RT templates were fixed to 13 and 16 nt, respectively. The secondary sgRNAs for PE3b were designed separately and complementary with the edited strand of DNA sequence, which can induce a nick within 3 bp downstream of the target.

Data availability

All data supporting the findings of this study are available in the article, in Supplemental information, or from the corresponding author upon request. The deep sequencing data have been deposited in the NCBI BioProject: PRJNA675696. All plasmids used in this study are available through Addgene.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.ymthe.2021.02.022.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (grants 2020R1F1A1075508, 2017M3A9B4062419, and 2018 R1A5A2020732) and the Ministry of Trade, Industry & Energy, Korea (20012445)

AUTHOR CONTRIBUTIONS

Y.K. supervised the research. J.K., A.-H.J., H.R.S., J.-E.S., and G.J. carried out experiments. J.K. and Y.K. performed data analysis. J.-K.Y. analyzed the ClinVar database. J.-K.Y. and J.-I.K. provided conceptual advice. Y.K. and J.K. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. *38*, 824–844.

2. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

3. 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.

4. Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. *31*, 230–232.

5. Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

6. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.

7. Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat. Biotechnol. *36*, 765–771.

8. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. Nat. Med. *24*, 927–930.

9. Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., et al. (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nat. Med. *24*, 939–946.

10. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

11. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*, aaf9729.

12. Hess, G.T., Frésard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat. Methods *13*, 1036–1042.

13. Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat. Methods *13*, 1029–1035.

14. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

15. 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., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149–157.

16. 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.

17. 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.

18. Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R., et al. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

19. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

20. Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., and Liu, D.R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature *556*, 57–63.

21. Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science *361*, 1259–1262.

22. Walton, R.T., Christie, K.A., Whittaker, M.N., and Kleinstiver, B.P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science *368*, 290–296.

23. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

24. Schene, I.F., Joore, I.P., Oka, R., Mokry, M., van Vugt, A.H.M., van Boxtel, R., van der Doef, H.P.J., van der Laan, L.J.W., Verstegen, M.M.A., van Hasselt, P.M., et al. (2020). Prime editing for functional repair in patient-derived disease models. Nat. Commun. *11*, 5352.

25. Kim, D.Y., Moon, S.B., Ko, J.H., Kim, Y.S., and Kim, D. (2020). Unbiased investigation of specificities of prime editing systems in human cells. Nucleic Acids Res. *48*, 10576–10589.

26. Karoulia, Z., Gavathiotis, E., and Poulikakos, P.I. (2017). New perspectives for targeting RAF kinase in human cancer. Nat. Rev. Cancer *17*, 676–691.

27. Morris, J.A., Rahman, J.A., Guo, X., and Sanjana, N.E. (2020). Automated design of CRISPR prime editors for thousands of human pathogenic variants. bioRxiv. https://doi.org/10.1101/2020.05.07.083444.

28. Kim, H.K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S., and Kim, H.H. (2021). Predicting the efficiency of prime editing guide RNAs in human cells. Nat. Biotechnol. *39*, 198–206.

29. Kweon, J., Jang, A.H., Kim, D.E., Yang, J.W., Yoon, M., Rim Shin, H., Kim, J.S., and Kim, Y. (2017). Fusion guide RNAs for orthogonal gene manipulation with Cas9 and Cpf1. Nat. Commun. *8*, 1723.

30. Kim, D., Kim, D.E., Lee, G., Cho, S.I., and Kim, J.S. (2019). Genome-wide target specificity of CRISPR RNA-guided adenine base editors. Nat. Biotechnol. *37*, 430–435.

31. Park, J., Lim, K., Kim, J.S., and Bae, S. (2017). Cas-analyzer: an online tool for assessing genome editing results using NGS data. Bioinformatics *33*, 286–288.

32. Landrum, M.J., Chitipiralla, S., Brown, G.R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang, W., Kaur, K., Liu, C., et al. (2020). ClinVar: improvements to accessing data. Nucleic Acids Res. *48* (D1), D835–D844.

YMTHE, Volume *29*

Supplemental Information

Engineered prime editors with PAM flexibility

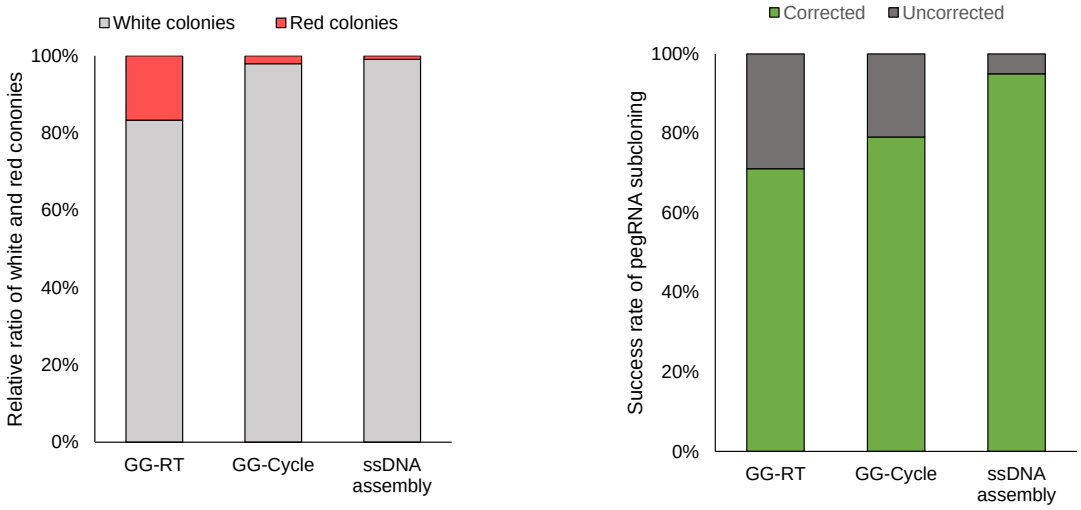
Jiyeon Kweon, Jung-Ki Yoon, An-Hee Jang, Ha Rim Shin, Ji-Eun See, Gayoung Jang, Jong-Il Kim, and Yongsub Kim

Supplementary Information

Table of contents

Note S1. Procedure for pegRNA subcloning by ssDNA assembly.

To improve the throughput of pegRNA construction, we used the ssDNA assembly method instead of the Golden Gate (GG) assembly method. In comparison with Golden Gate assembly, ssDNA assembly has no oligonucleotide annealing step in a thermocycler and does not require the phosphorylation of gRNA scaffold oligonucleotides. The ratio of red/white colonies is similar in the ssDNA assembly method (GG-RT: incubation at room temperature for 10 min; GG-Cycle: 8 cycles of 5 min at 16°C and 5 min at 37°C). We also determined the cloning efficiency by the targeted-deep sequencing of pooled transformants. In comparison with conventional methods, ssDNA assembly has a slightly higher success rate in pegRNA cloning. The detailed protocol of pegRNA cloning is described below.



1. Oligonucleotide preparation for pegRNA cloning

    a. Oligonucleotides for the spacer sequence (target-specific component)
    The 18 nt 5' overhang (5'-tggaaaggacgaaacacc-3') and 3' overhang (5'-gttttagagctagaaata-3') are attached to the desired spacer sequence as follows:
    5'- tggaaaggacgaaacaccNNNNNNNNNNNNNNNNNNNgttttagagctagaaata-3',
    where N×20 is the target spacer sequence without PAM, and the spacer sequence must begin with a guanine nucleotide. If the spacer sequence does not begin with a guanine nucleotide, an additional guanine nucleotide should be added for pegRNA transcription by the U6 promoter.

    b. Oligonucleotides for the RT template and PBS sequence (target-specific component)
    The 18 nt 5' overhang (5'-gtggcaccgagtcggtgc-3') and 3' overhang (5'-ttttttaagcttgggcc-3') are attached to the desired spacer sequence as follows:
    5'-gtggcaccgagtcggtgcNNNNNNNNNNNNNNNNNNNttttttaagcttgggcc-3',
    where Ns are the desired RT template and PBS sequence.

    c. Oligonucleotides for the pegRNA scaffold sequence (common component)
    The reverse complementary sequence of gRNA scaffold is as follows:
    5'-gcaccgactcggtgccacttttcaagttgataacggactagccttattttaacttgctatttctagctctaaaac-3'

2. Vector preparation for pegRNA cloning

    The pU6-pegRNA-GG-acceptor (addgene, plasmid #132777) plasmid DNA is digested with

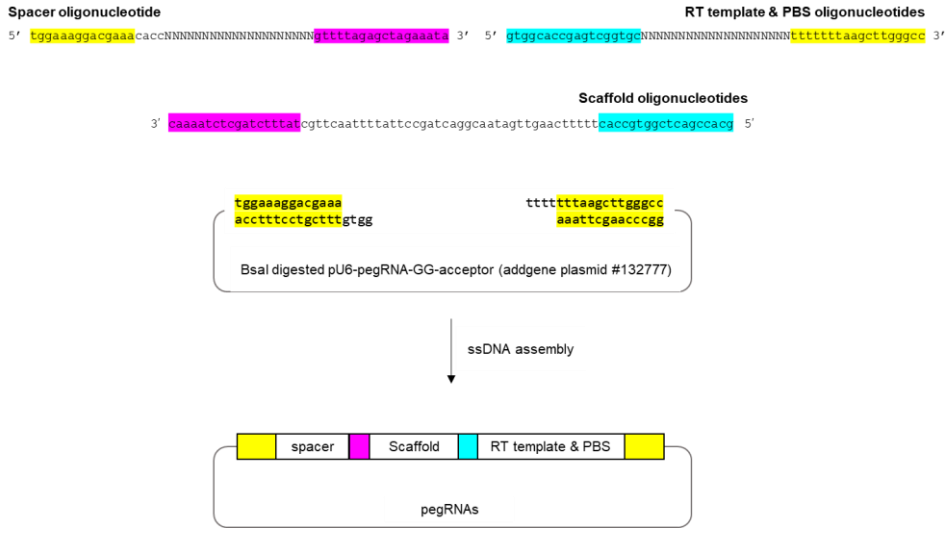
the BsaI-HFv2 (NEB R3733S) restriction enzyme as previously described.

3. pegRNA construction by ssDNA assembly
   Each component is mixed for ssDNA assembly as follows:

Vector DNA	50 ng
Oligonucleotides for spacer	1 µl of 100 nM
Oligonucleotides for RT template and PBS	1 µl of 100 nM
Oligonucleotides for gRNA scaffold	1 µl of 100 nM
2× HiFi DNA Assembly Master Mix (NEB #E2621S)	5 µl
H_2O	up to 10 µl

   The reaction mixture is incubated at 50°C for 30 min, and an appropriate amount is transformed using competent cells.

The following scheme summarizes the process of ssDNA assembly in the pegRNA subcloning protocol.



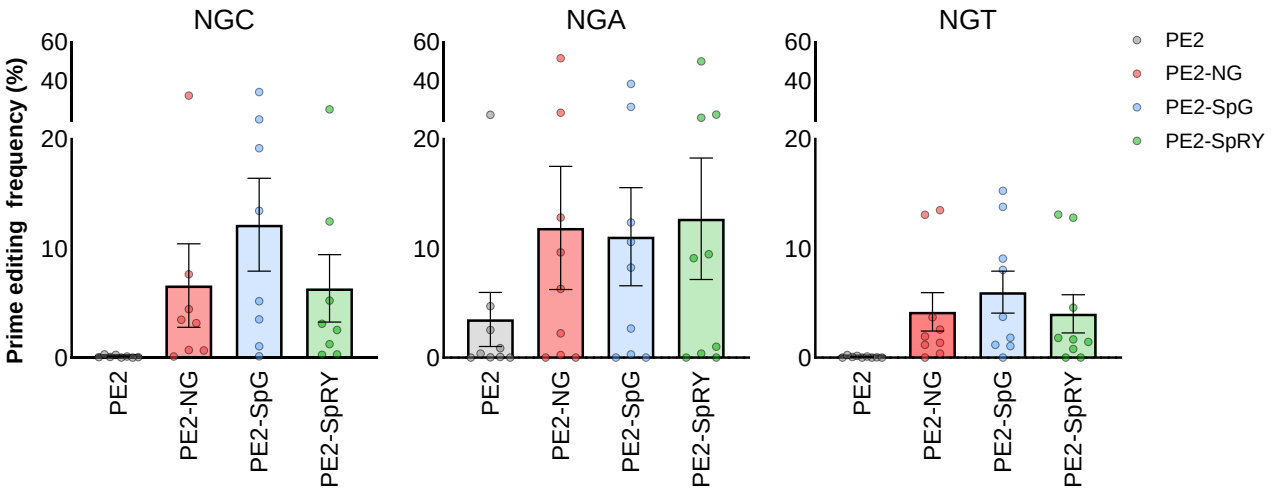


Figure S1. Prime editing activities of the PE2 variants at the NGC, NGA, and NGT PAM sites in HEK293T cells. The PE2-SpG variant showed average prime editing activities of 12.2% and 6.0% at 8 NGC PAM sites and 9 NGT PAM sites, respectively. The wild-type PE2 showed an average prime editing activity of 2.5% at 9 NGA PAM sites. The numerical values of the prime editing activity are listed in Table S1. Mean ± s.e.m. of the average prime editing frequency for each target site.

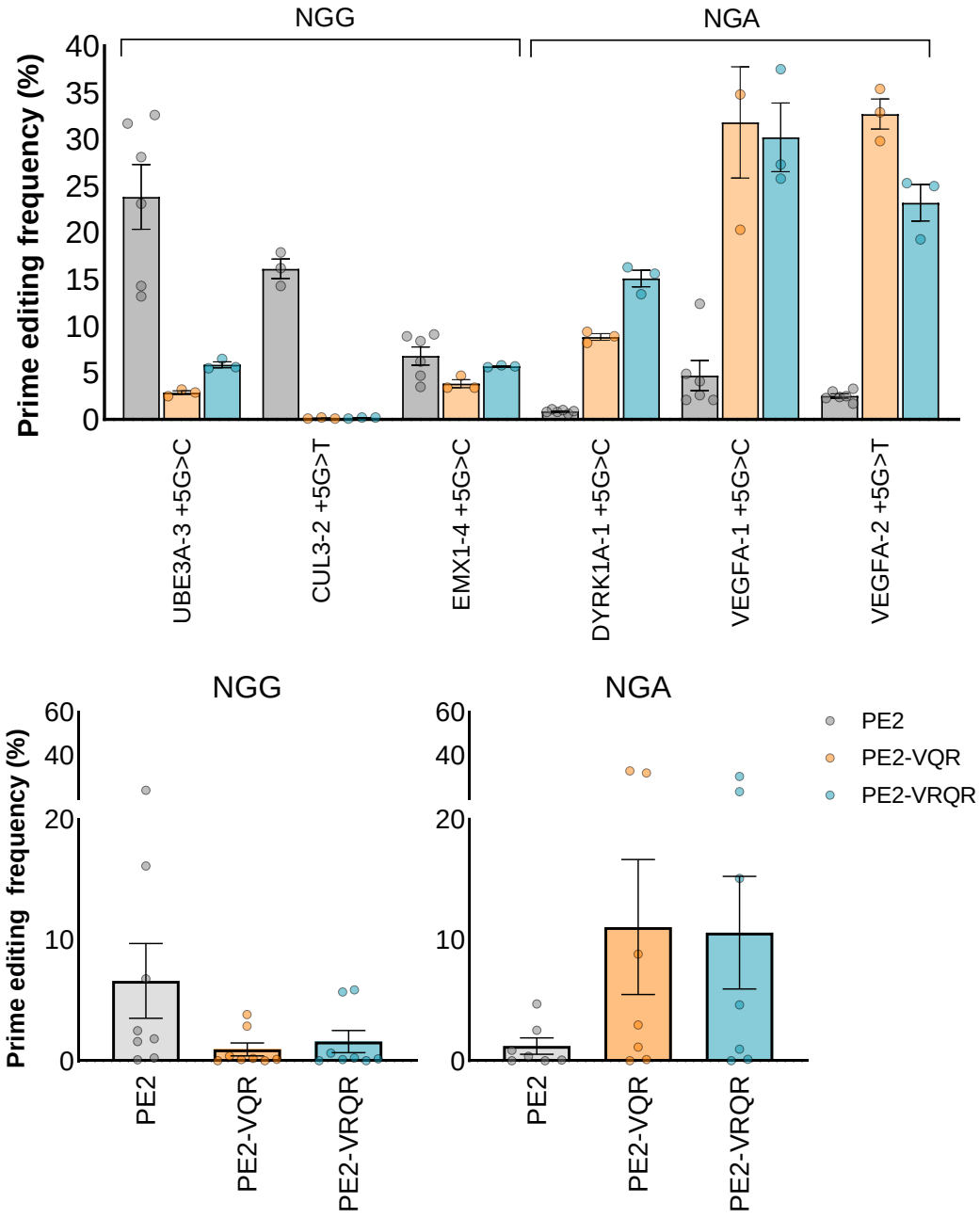


Figure S2. Prime editing activities of the PE2-VQR and PE2-VRQR variants in HEK293T cells. The PE2-VQR and PE2-VRQR variants showed average prime editing activities of 11.1% and 10.7% at 7 NGA PAM sites each. The numerical values of the prime editing activity are listed in Table S1. Mean ± s.e.m. of $n = 3$ or n=6 independent biological replicates in the top panel. Mean ± s.e.m. of the average prime editing frequency for each target site in the bottom panel.

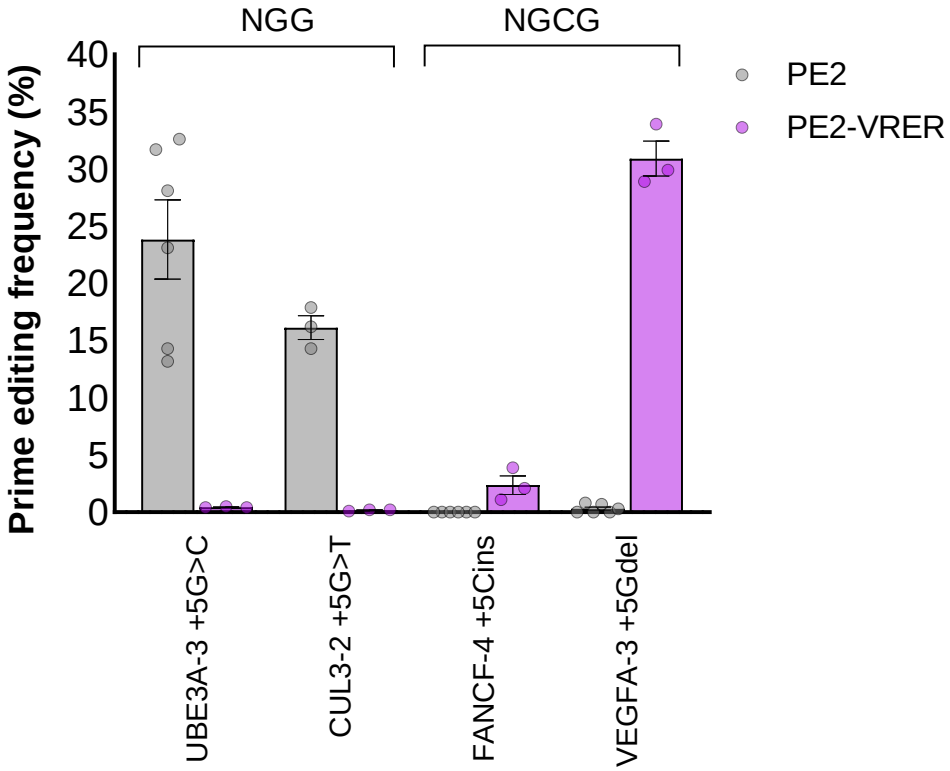


Figure S3. Prime editing activities of the PE2-VRER variant in HEK293T cells. The PE2-VRER variant showed an average prime editing activity of 30.9% at the VEGFA-3+5Gdel site with NGCG PAM. The numerical values of the prime editing activity are listed in Table S1. Mean ± s.e.m. of $n = 3$ or n=6 independent biological replicates.

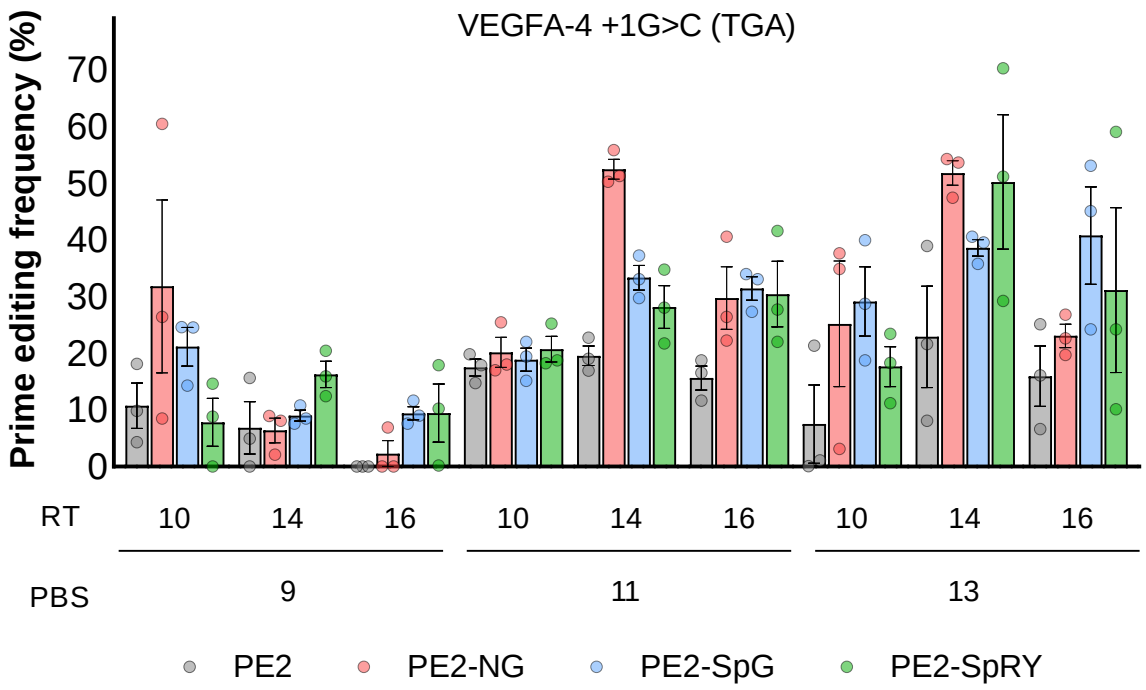


Figure S4. Prime editing activities at the VEGFA-4 +1G>C site with various PBS and RT template lengths in pegRNAs in HEK293T cells. The PAM sequence is shown in the parentheses, and the numerical values of the prime editing activity are listed in Table S1. Mean ± s.e.m. of $n = 3$ independent biological replicates.

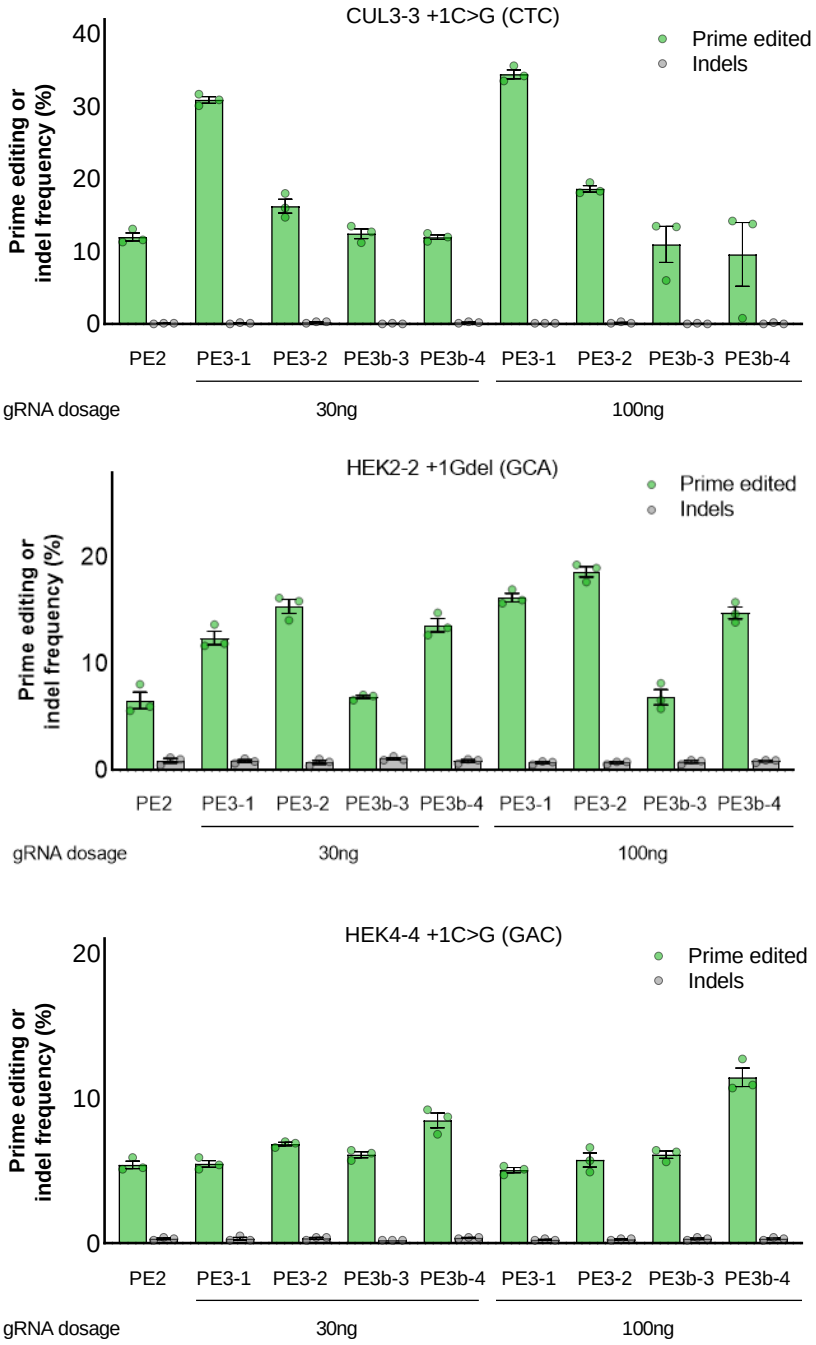


Figure S5. Prime editing using the PE3 and PE3b systems at three target sites in HEK293 cells. Two different amounts (30 ng or 100 ng) of gRNAs were used for PE3 or PE3b as described in the Materials and Methods section, and the frequencies of prime editing activity with 30 ng of gRNA were plotted as shown in Figures 3A-C. There was no significant increase in prime editing activity or indels with an increased amount of gRNA at three target sites. The PAM sequences are shown in the parentheses, and the numerical values of the prime editing activity are listed in Table S1. Mean ± s.e.m. of $n = 3$ independent biological replicates.

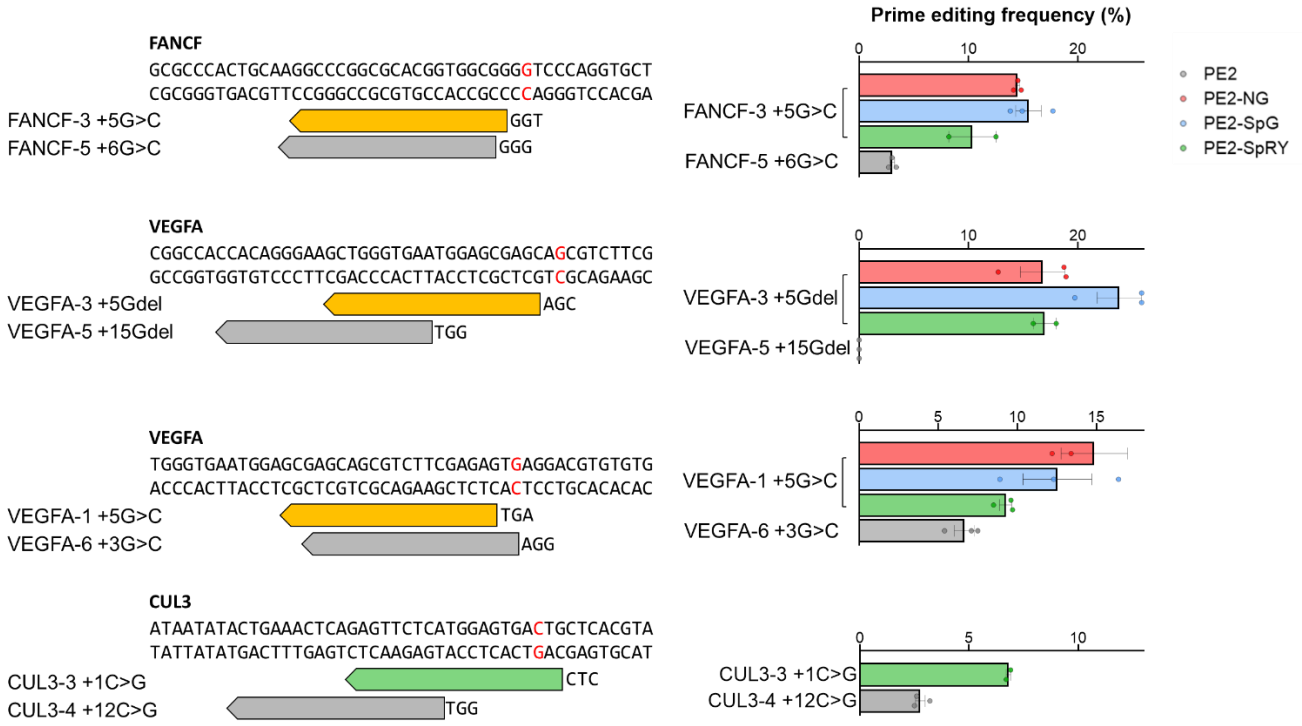


Figure S6. Comparison of prime editing activities between wild-type PE2 and PE2 variants in HEK293T cells. The pegRNAs for wild-type PE2 were designed to introduce the same mutations introduced by PE2 variants. The positions of the intended mutations are highlighted in red, and the PAM and spacer sequences of each pegRNA are shown below the target sequences. The numerical values of the prime editing activity are listed in Table S1. Mean ± s.e.m. of n = 2 or 3 independent biological replicates.

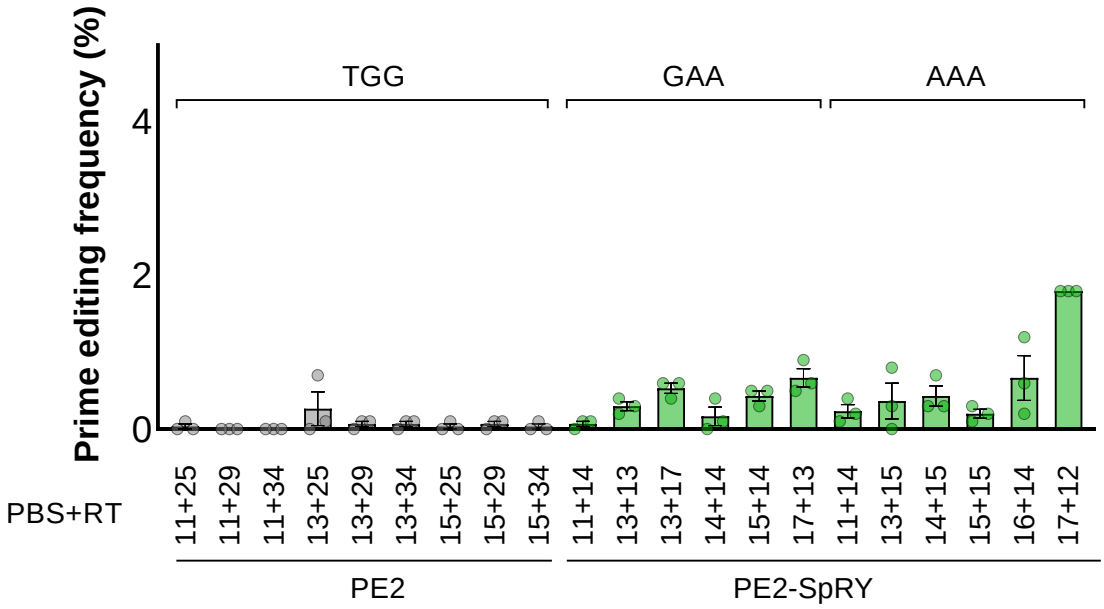


Figure S7. Prime editing activities at the BRAF V600E site with various PBS and RT template lengths in pegRNAs. The pegRNA with 13 nt PBS and 25 nt RT template showed the highest prime editing activity (average of 0.3%) for wild-type PE2, and the pegRNA with 17 nt PBS and 12 nt RT template showed 1.8% prime editing activity for PE2-SpRY. The numerical values of the prime editing activity are listed in Table S1. Prime editing was conducted using HEK293T cells. Mean ± s.e.m. of n = 3 independent biological replicates.

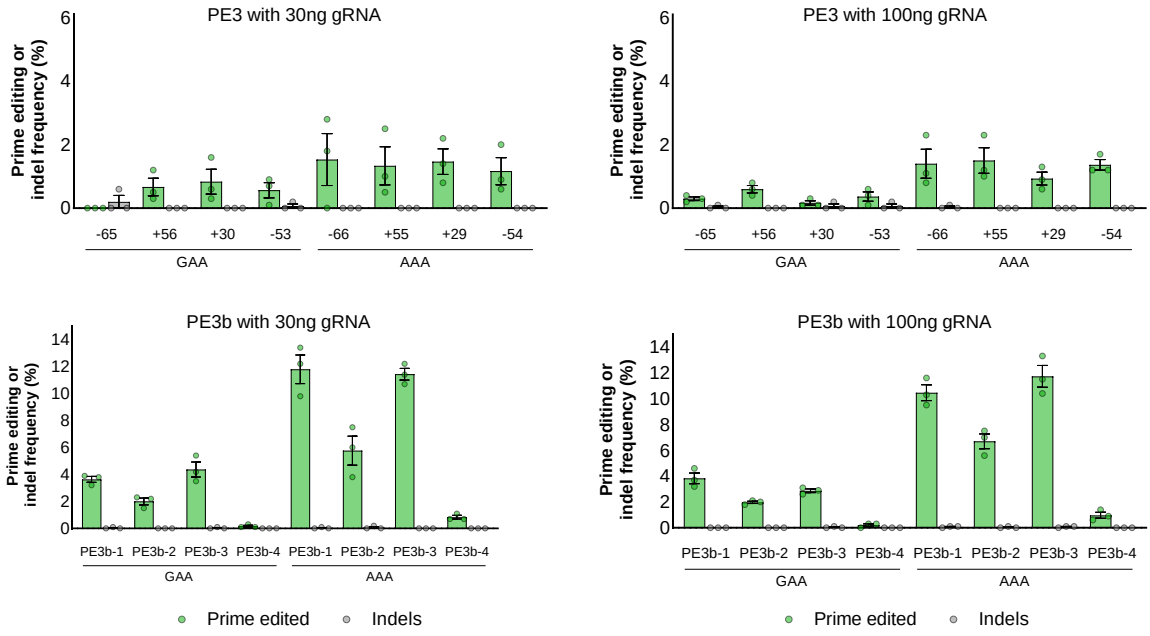


Figure S8. Prime editing using the PE3 and PE3b systems at the BRAF V600E site. BRAF-GAA and BRAF-AAA pegRNAs were evaluated with four gRNAs in the PE3 system. Each gRNA could induce nicks to the non-edited strand; the nicked positions are indicated with the locations of pegRNA-induced nicks. The numerical values of the prime editing activity are listed in Supplementary Table 1. Prime editing was conducted using HEK293T cells. Mean ± s.e.m. of n = 3 independent biological replicates.

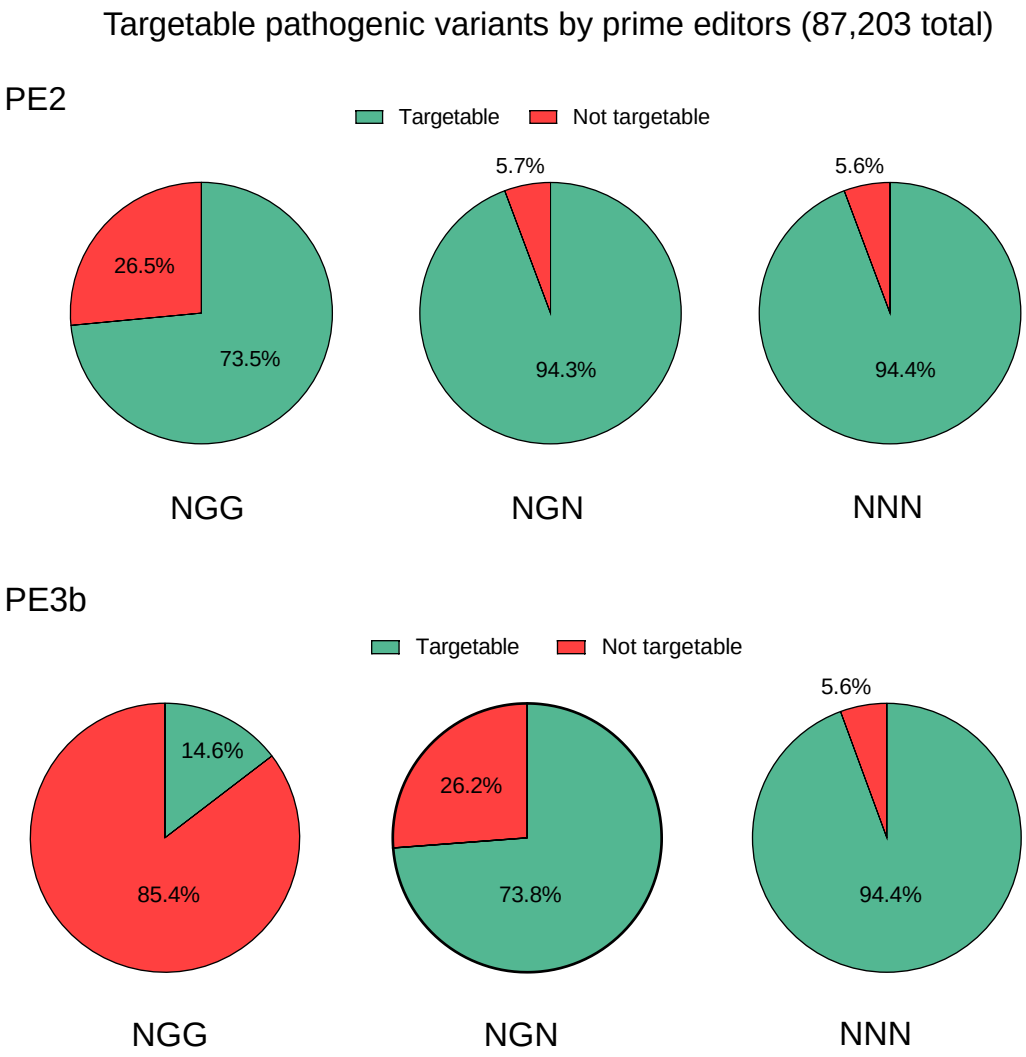


Figure S9. Fraction of human pathogenic variants that could be targeted with the PE2 or PE3b system. Overall, 73.5% of pathogenic variants were targetable for wild-type PE2 (with NGG PAM), and 94.4% of pathogenic variants were targetable for PE2-SpRY (with NNN PAM). In the case of the PE3b system, only 14.6% of variants were targetable for wild-type PE2; in contrast, targetable pathogenic variants were greatly increased up to 94.4% for PE2-SpRY. The remaining 5.6% of pathogenic variants that cannot be targeted were large deletions, insertions, or complex mutations. Total counts are listed in Supplementary Table 4.

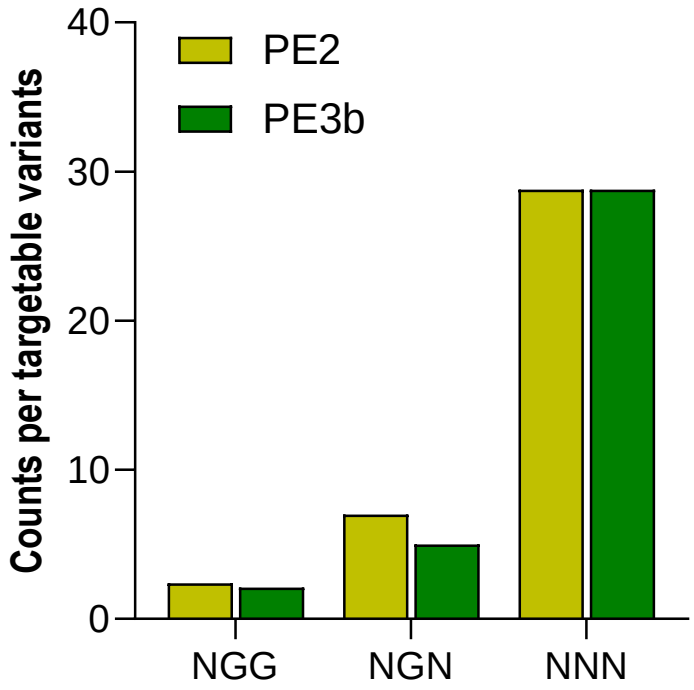


Figure S10. Number of target sites per targetable variant in the PE2 and PE3b systems. The PE2-SpRY variant increased the number of designable pegRNAs per pathogenic variant (from 2.4 to 28.8 pegRNAs in the PE2 system and from 1.6 to 28.8 pegRNAs in the PE3b system). Considering the diverse activity of prime editing, the more pegRNAs that can be designed, the higher the success rate of prime editing. Total counts are listed in Supplementary Table 4.

Table S3. List of off-target sites in this study.

pegRNA names		Chr.	Spacer sequences	PAM	PBS	RT template	Reference for off-target sites
HEK4 +2G>T	ON	Chr.20	GGCACTGCGGCTGGAGGTGG	GGG	GCGGCTGGAGG	TTGGGGTTAA	Anzalone *et al*.
	OFF	Chr.10	GGCACgaCGGCTGGAGGTGG	GGG	aCGGCTGGAGG	TTGGGGgTtg	
FANCF-4 +5Cins	ON	Chr.11	GCAGAAGGGATTCCATGAGG	TGC	AAGGGATTCCATG	AGGTCGCGCGAAGGC	Walton *et al*.
	OFF	Chr.11	GaAGAAGGGtTTCCATGAGG	AGA	AAGGGtTTCCATG	AGGaCGatacctgag	
EMX1-4 +5G>C	ON	Chr.2	GTCACCTCCAATGACTAGGG	TGG	CCTCCAATGACTA	GGGTCGGCAACCA	
	OFF	Chr.17	GTCACCTgtAATGACTAGGG	AGA	CCTgtAATGACTA	GGGaCaGtAAtgg	
MECP2-3 +1G>C	ON	Chr.X	GGGTGGTTCCATAATCTGTG	TAT	GGTTCCATAATCT	CTGTATACCTAAG	
	OFF	Chr.4	GcGTGGTTaCATAATCTGTG	GAG	GGTTaCATAATCT	CTGgAggggTgca	
CUL3-3 +1C>G	ON	Chr.2	GAGTTCTCATGGAGTGACTG	CTC	TCTCATGGAGTGA	GTGCTCACGTAAC	In this study
	OFF1	Chr.14	aAGTTtTCATGGAGTGACTG	TCA	TtTCATGGAGTGA	GTGtcacaGagta	
	OFF2	Chr.2	GtGTTaTCATGGAGTGACTG	AGA	TaTCATGGAGTGA	GTGagaAtGacAC	
	OFF3	Chr.8	GAGTTCTgATGGAGTGACTG	CTG	TCTgATGGAGTGA	GTGCTggtGccca	
HEK2-4 +1Cins	ON	Chr.5	GGGCGGGCCAGCCTGAATAG	CTG	GGGCCAGCCTGAA	CTAGCTGCAAACAA	
	OFF1	Chr.13	GGGCtGGCCAGCCTGAtTAG	AAT	tGGCCAGCCTGAt	CTAGaataggctAA	
	OFF2	Chr.22	GGGCGaGCCAGCCTG(g)AATAG	GTA	GaGCCAGCCTG(g)AA	CTAGgTaCtggCct	
	OFF3	Chr.5	aGGCtGGCCAGCCTGAATAG	CTT	tGGCCAGCCTGAA	CTAGCTtCAgcagg	

Mismatches are highlighted in blue with lowercase letters. At the HEK2-3 +1Cins OFF-2 site, () indicates a bulge position compared with the on-target site. The intended mutations of pegRNAs are shown in red.

Table S4. Specificity ratio of on- and off-target prime editing activities.

pegRNA name		Specificity ratio					
		Nucleases			Prime editors		
		wildtype	SpG	SpRY	PE2	PE2-SpG	PE2-SpRY
HEK4 +2G>T		1.88	1.55	5.17	6.18	6.19	15.18
FANCF-4 +5Cins		N/A	2.98	20.67	N/A	> 35	> 35
EMX1-4 +5G>C		164.5	3.37	5.67	> 150	> 124	> 30
MECP2-3 +1G>C		N/A	N/A	54.85	N/A	N/A	> 93
CUL3-3 +1C>G	OFF1	N/A	N/A	50	N/A	N/A	57.6
	OFF2	N/A	N/A	225	N/A	N/A	> 192
	OFF3	N/A	N/A	112.5	N/A	N/A	> 192
HEK2-4 +1Cins	OFF1	N/A	N/A	40.33	N/A	N/A	113
	OFF2	N/A	N/A	121	N/A	N/A	> 150
	OFF3	N/A	N/A	60.5	N/A	N/A	> 150

Table S5. Analysis of targetable pathogenic variants using PE variants.

Types		Number of targetable variants		number of pegRNAs per target variants	
Prime editors	PAM	Counts	Percentage (%)	Total counts	pegRNAs per targetabel variants
PE2	NGG	64,121	44.8%	156,862	2.4
	NGN	82,239	79.2%	572,093	7.0
	NNN	82,283	89.6%	2,372,052	28.8
PE3b	NGG	12,769	7.9%	26,589	2.1
	NGN	64,315	51.3%	320,763	5.0
	NNN	82,283	83.8%	2,372,052	28.8

*A total of 87,203 variants from the ClinVar database were analyzed.