SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3488.

BICZÓK ZSUZSANNA

Patobiokémia

című program

Programvezető: Dr. Csala Miklós, egyetemi tanár

Témavezető: Dr. Welker Ervin, tudományos tanácsadó

Investigation of the effects of altered and split pegRNAs on intramolecular complementarity and prime editing outcomes

PhD thesis

Zsuzsanna Biczók

Semmelweis University Doctoral College

Molecular Medicine Division

Supervisor: Ervin Welker, D.Sc

Official reviewers: Bernadett Szikriszt, Ph.D
Tamás Mészáros, Ph.D

Head of the Complex Examination Committee: Attila Tordai, D.Sc

Members of the Complex Examination Committee: Máté Varga, Ph.D
András Váradi, D.Sc
Hedvig Tordai, Ph.D

Budapest
2026

Table of Contents

List of Abbreviations

BE	Base Editor
Cas9	CRISPR associated protein 9
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
dCas9	Dead Cas9
DSB	Double-Strand Break
engRNA	Essential nicking guide RNA
epegRNA	Engineered pegRNA
HDR	Homology-Directed Repair
hPSC	Human Pluripotent Stem Cell
M-MLV	Moloney-Murine Leukemia Virus
MMR	Mismatch Repair
nCas9	Nickase Cas9
NHEJ	Non-Homologous End-Joining
NLS	Nuclear Localisation Signal
PAM	Protospacer Adjacent Motive
PBS	Primer Binding Site
PE	Prime Editor
PEAR	Prime Editing Activity Reporter
pegRNA	Prime Editing Guide RNA
ProPE	Prime Editing with Prolonged Editing Window
RT	Reverse Transcriptase
RTT	Reverse Transcriptase Template
sgRNA	Single Guide RNA
SNP	Single Nucleotide Polymorphism
SPELL	Streamlined Prime Editing with fixed-Length PBS Leverage
SpRY Cas9	PAM-flexible SpCas9 variant
tpgRNA	Template Providing RNA

1. Introduction

1.1 The CRISPR/Cas9 system

The discovery of CRISPR/Cas systems, often referred to as bacterial immune systems because they function as a defence mechanism against viruses and bacteriophages, has led to the development of a revolutionary new gene-editing technology (1). Today, because of their practicality and customizability, these systems and their numerous variants have become widely used tools in targeted genome editing of living organisms and in research (2,3). The most commonly used CRISPR/Cas system originates from *Streptococcus pyogenes* and consists of a complex of a Cas9 nuclease (often referred to as SpCas9) and a single guide RNA (sgRNA) (**Figure 1**). The 20-nucleotide-long spacer sequence of the sgRNA directs the complex to its target on a double-stranded DNA. An NGG protospacer-adjacent motif (PAM) sequence recognized by the SpCas9 protein, located downstream of the target sequence, is also essential, since in its absence, binding between the target and spacer sequences does not occur. Once the spacer sequence has hybridized to its target, the SpCas9 protein cleaves both strands of DNA three base pairs upstream of the PAM (2,3) (**Figure 1**). This double-strand break (DSB) is repaired by the cells' own repair mechanisms, and this process can be utilized for DNA editing. Which repair mechanism is activated depends on numerous factors, including the cell type, the cell's state, or the type of double-strand break. Generally speaking, these repair mechanisms either rejoin the strands, such as non-homologous end-joining (NHEJ), which is prevalent in mammalian cells and often involves the addition or removal of nucleotides, or they use a DNA template for repair, such as homology-directed repair (HDR), which operates primarily in dividing cells mainly in the S or G2 cell state (3,4).

Since in most mammalian cells DSBs are repaired with the error-prone end-joining process, that generates insertions or deletions (indels) at the targeted site, this is more suited to generate frameshift mutations (5,6), disrupt *cis*-regulatory elements (7), or when two sgRNAs are used at the same time, create large deletions (8). By providing a DNA template when cleaving at the genomic site, HDR can also be used for editing: to introduce point mutations, precise insertions, or deletions, as well as to integrate large, gene-sized DNA fragments (8,9) (**Figure 1**). However, despite numerous attempts to improve the efficiency of HDR, such as by optimizing DNA donors (10,11) or adding

HDR-enhancing chemicals (12), the majority of edits still contain unwanted indels or consist entirely of them (13). Cleavage of both strands of the DNA can also lead to chromosomal rearrangements or translocations (14), and although this technology holds great promise not only in research but also in medicine and agriculture, questions regarding its safe application in these fields still need to be resolved.

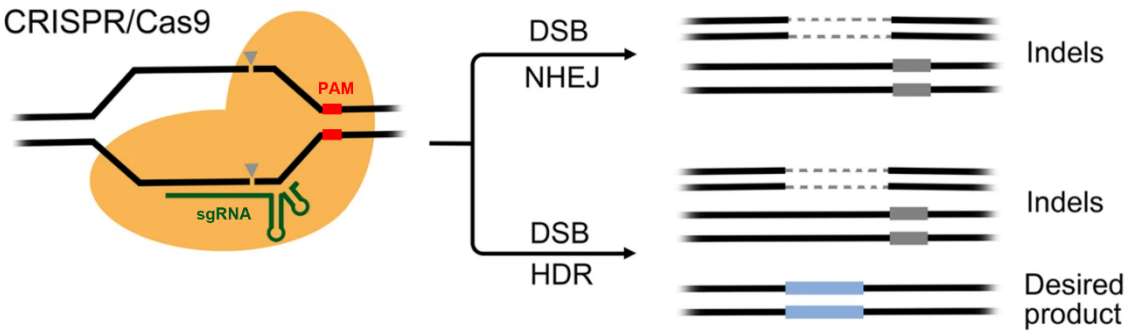

Figure 1. The SpCas9 nuclease and the two principal pathways of double-strand break (DSB)-mediated genome editing. The SpCas9 is shown in complex with its DNA target. The sgRNA is depicted in green, the SpCas9 protein is shown in yellow. The PAM sequence is indicated in red. Gray triangles denote the cleavage sites of the respective DNA strands. Non-homologous end joining (NHEJ) typically results in small insertions or deletions (indicated in gray), while homology-directed repair (HDR), in the presence of a repair template, enables the precise incorporation of a defined DNA sequence (indicated in blue). Figure adapted from (4), with minor edits.

SpCas9 can be highly active at target sequences, but it is also prone to cutting at sites that share sequence similarities with the original target, thereby causing unintended modifications (15). To reduce this so-called off-target effect, numerous variants of SpCas9 have been developed through rational design or directed evolution (16,17), and although these often do reduce off-target activity, the on-target efficiency of the proteins also diminishes, therefore, screening these variants for a given target is necessary to achieve optimal on-target but reduced off-target activity (18).

1.2 The new generation CRISPR/Cas9 based tools

A new generation of SpCas9-based tools, base editors (BE) and prime editors (PE), decrease the occurrence of unwanted modifications by using an inactive (dead) or nickase form of the protein (dSpCas9 and nSpCas9, respectively), which either possess no cleavage activity or cleave only one strand of the DNA. These methods are also better suited for making small, base-pair-level modifications to the genome.

1.2.1 The base editor (BE)

To create base editors, either a cytosine deaminase or a modified version of it, adenosine deaminase, has been fused to the dSpCas9 protein, enabling C to T, T to C, A to G, or G to A base conversions (19,20). Numerous BE variants are now available, which are more active or have editing windows of varying sizes, and C to G transversion base editors have also been developed (3,21,22). Although base editors are suitable for modifying a single base, their use often results in additional edits when a nearby C or A base is also edited; therefore, the most suitable variant for the given edit must be selected with careful consideration (3).

1.2.2 The prime editor (PE)

Prime editing is also capable of base pair substitution, although base editors can access different editing positions, as they perform editing within the target sequence, whereas prime editing allows for editing in the downstream region from the cleavage site of the nSpCas9 (3,23). The true practicality of prime editing, however, lies in its ability to perform any type of modification in the genome: not only substitutions, but insertions and deletions as well. This is made possible by a 3'-extension of the sgRNA, which results in the prime editing guide RNA (pegRNA), that serves as a template for an engineered version of the Moloney-murine leukemia virus (M-MLV) reverse transcriptase (RT) linked to the SpCas9 nickase protein. During prime editing, the PE protein, in complex with the pegRNA, recognizes and binds to the target sequence, after which the non-targeted strand is cleaved. The cleaved DNA strand then hybridizes to the 3' sequence of the pegRNA, known as the primer-binding site (PBS). Then, the RT elongates the non-targeted DNA strand based on the pegRNAs reverse transcriptase template (RTT) region, which is located between the PBS and the scaffold sequence and contains the desired modification to be introduced into the genome (**Figure 2a, b**). The newly synthesized 3' DNA flap initially competes with the original 5' flap of the cleaved DNA (flap equilibration), but the cells' DNA repair mechanisms are able to excise the 5' DNA flap, allowing the edited strand to be incorporated into the genome. The edit is finalized by replacing and repairing the targeted, unedited strand (3,23,24) (**Figure 2c-f**). This system was named PE2.

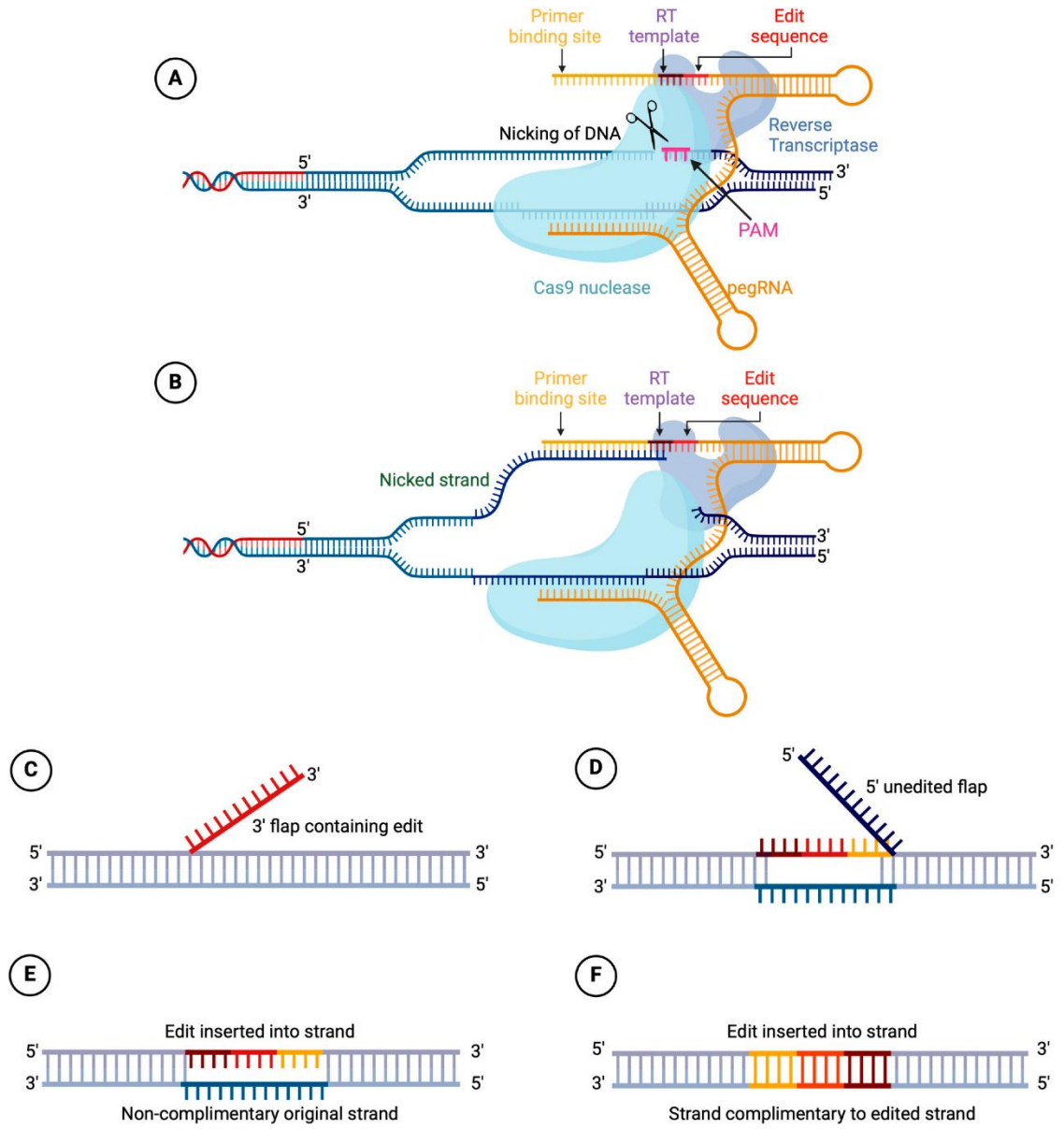

Figure 2. The components and mechanism of the prime editing system. a) A depiction of the main components of the prime editing system. **b**) The new DNA strand synthesis based on the RTT. **c, d**) Flap equilibration, the newly synthetized 3' DNA strand and the unedited 5' strand compete for binding to the original strand. **e, f**) The edited strand gets successfully incorporated into the genome. and Figure adapted from (24).

1.2.2.1 The limitations of PE and its advanced versions

The prime editing technology has been utilized to edit numerous endogenous targets, many of which are clinically relevant (e.g., HBB—the E6V mutation is responsible for sickle cell anemia (23); HEXA—a 4-base insertion causes Tay-Sachs disease (23); PRNP

– the G127V mutation confers resistance to prion diseases (23); CDKL5 – pathogenic variants are responsible for CDKL5 deficiency disorder (25); CTNNB1 – mutations can lead to hepatocellular carcinoma (26); CXCR4 – the P191A mutation prevents HIV infection (25), and IL2RB – certain mutant variants enable orthogonal IL-2 receptor reactivity in T-cell transfer therapy (25)) in order to investigate their broad applicability and specificity. It has also been demonstrated that PE is capable of gene modification in various cell types, such as immortalized and cancer-derived human cell lines (HEK293T, K562, U2OS, HeLa) (23) or in human pluripotent stem cells (hPSCs) and mouse embryonic stem cells (ES), the transfection of which is generally more challenging. Furthermore, PE has proven to be active in organoids, various model organisms (Drosophila, zebrafish, mice), and plants as well (27). Despite these successful editing results, however, low editing efficiency remains a challenge, leading to the development of several PE variants to overcome this limitation.

To improve the editing efficiency of the PE2 system, an additional sgRNA was introduced that nicks a secondary target site on the unedited strand. This prompts the cells' repair mechanisms to use the edited strand as a template to repair the opposing unedited strand, thereby increasing the incorporation of the modification into the genome. This system was named PE3, which successfully increased editing efficiency, although increased indel activity was also reported, and prime editing may still be relatively inefficient for certain genomic targets (23). By incorporating mutations into the PE protein that enhance its nicking function, codon optimization, and the addition of additional nuclear localization signals (NLS), PEmax was developed, which further increased the efficiency of prime editing (25).

Further improvements were achieved with the development of PE4, in which a dominant-negative MLH1 variant (MLH1dn) is co-expressed with PE2 (25). MLH1dn is known to inhibit mismatch repair (MMR), one of the repair pathways that influences prime editing, and it has already been demonstrated that inhibition of MMR-related genes can increase PE efficiency (28–30). As described in PE3, PE5 was created by adding a secondary nick to the unedited DNA strand (25). Furthermore, in addition to the intended edit, the synonymous mutations incorporated into the pegRNA RTT sequence may also reduce the effect of MMR, as this repair mechanism is less effective at recognizing heteroduplexes with three or more consecutive mismatches (25).

During the development of the PE6 system, several variants were created (PE6a–PE6g); some of these were smaller in size to facilitate delivery via viral vectors (PE6a and PE6b), while others contained modified SpCas9 and RT components to enhance the system's editing efficiency and specificity (PE6c–PE6f) (31).

To develop PE7, the La protein, a small, RNA-binding exonuclease-inhibiting factor capable of binding to the pol-U region of pegRNA transcribed by RNA polymerase III, was fused to PEmax. This system thus stabilizes the pegRNA by protecting its 3' end, and exhibits improved prime editing activity compared to previous PE versions (32,33).

Originally, Anzalone *et al.* tested prime editing on sequences located up to 33 nucleotides downstream of the cleavage site and reported detectable editing results (3,23). However, most studies on prime editing focus on edits within the target sequence (those located 1–3 spacer positions upstream of the cleavage site or within the PAM, likely because mismatching in the seed region of the target sequence disrupts SpCas9 activity or because the mutated PAM is no longer recognisable for the SpCas9) and optimize editing at these sites (25,34–36). Surprisingly few studies have evaluated prime editing at positions distant from the nick site (more than 10 nucleotides away), and in these, it was found that PE generally operates with poor efficiency at target-distal sites (37,38) (**Figure 3**). Increasing the efficiency of low-performing and target-distal edits broadens the applicability of prime editing-based techniques, including the development of therapies.

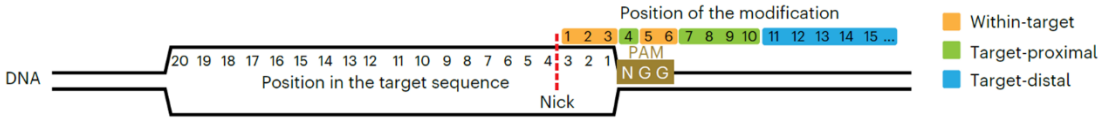

Figure 3. Specifying editing positions in prime editing. The nSpCas9 cleaves the DNA strand three nucleotides upstream of the PAM. Edits within the target region occur at these three nucleotides and at the GG region of the PAM sequence. Edits proximal to the target region are those located within 10 nucleotides of the cleavage site, while edits distal to the target region are located more than 10 nucleotides away. Figure adapted from (38).

PE is based on the conventional SpCas9 thus it recognises an NGG PAM sequence as well. To make certain edits easier to reach, when no NGG sequence can be found in close proximity, PAM flexible SpCas9 variants have been developed into prime editors. PE-SpG and PE-NG recognize a reduced NG PAM, while PE-SpRY (R is A or G, Y is T or

C) recognizes a relaxed NRN PAM. These versions, however, often show low editing efficiency (39).

1.2.2.2 Optimising the pegRNA

The properties of pegRNA are just as important as those of the PE protein, since the lengths of the PBS and RTT are critical for effective prime editing. The optimal length of these sequences depends on the target sequence: an effective PBS ranges from 8 to 14 nucleotides in length, and in general, pegRNAs with longer PBS (between 15 and 17 nucleotides) exhibit low editing efficiency, while the RTT is typically 10 to 20 nucleotides long, but multiple optimization steps are required to identify the most effective length. (23). Although numerous studies have attempted to reduce the challenges of pegRNA optimization, some focusing on melting temperature (40,41), others by incorporating mismatches into the spacer sequence (42), studying large pegRNA libraries (37,43–45) or treating transfected cells with cold shock (40), these attempts have met with only limited success, thus, the optimization of pegRNAs remains a substantial task.

The degradation of the 3'-end of the pegRNA by exonucleases, as it is not protected by the nSpCas9 can also lower the rate of successful prime editing. To address this issue, structured RNA motifs were incorporated into the 3'-end of pegRNA, thereby creating engineered pegRNAs (epegRNAs) to improve the efficiency of prime editing (34).

The structure of pegRNA, due to its intramolecular complementarity, can also influence the efficiency of prime editing. Since the spacer and PBS sequences bind to two separate strands of the same DNA segment, they are themselves complementary, and their binding may impair editing efficiency (40,42,46). The longer the PBS sequence, the more pronounced this effect could be, therefore, it is essential to find the optimal PBS length for a given target to achieve efficient prime editing.

Sontheimer and colleagues optimized prime editing by separating the pegRNA nicking and template providing functions and using conventional sgRNA, along with a circular RNA containing the PBS and RTT sequences, which reduces pegRNA misfolding, which can occur not only between the PBS and the spacer but also potentially between the RTT and scaffold sequences, and likely protects the PBS-RTT sequence from exonuclease degradation as well (47). The proPE method is also based on the separated pegRNA functions, but by using two separate guide RNAs. Cleaving of the non-target DNA strand is performed by the essential nicking guide RNA (engRNA), which is a

canonical sgRNA and utilizes a standard 20-nucleotide spacer sequence. The template for the newly synthesized DNA strand is provided by the template providing guide RNA (tpgRNA), which possesses the 3' extension of the pegRNA (the PBS and RTT sequences), but uses a short (11–15 nucleotide long) spacer sequence to avoid cleavage of the tpgRNA target site, located within close proximity to the engRNA target site (**Figure 4**). By using these two guide RNAs in appropriate ratios, several inhibitory factors, affecting the process of prime editing, can be eliminated simultaneously. It is possible to overcome limitations such as protein rebinding and repeated cleavage of the target sequence, which can disrupt an already successful edit or the digestion of the newly synthesized strand. This system also seems to be less sensitive for the inhibitory effect of spacer-PBS complementarity and long PBS sequences (38).

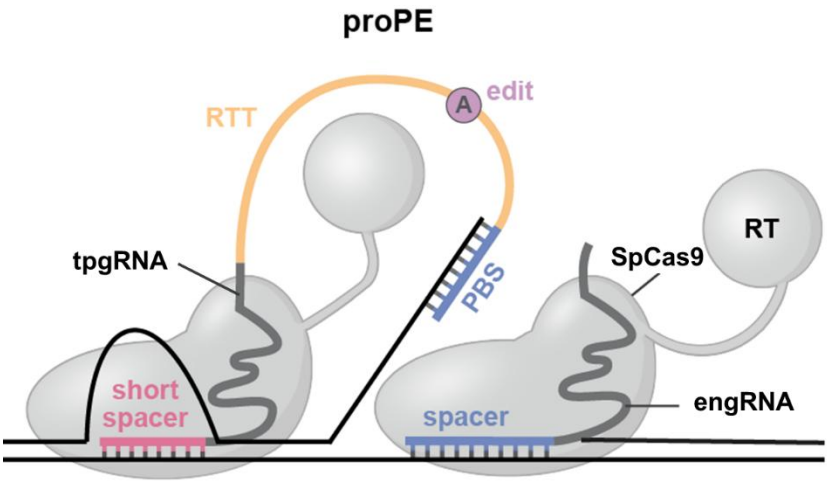

Figure 4. The main components of the proPE system. The complex of SpCas9 nickase and reverse transcriptase is shown in grey. The engRNA, a conventional sgRNA with a 20-nucleotide spacer sequence, is marked in blue. The tpgRNA contains the PBS and RTT sequences, which are marked in blue and yellow, respectively, and has a short (11–15 nucleotide long) spacer sequence, marked in pink. The figure shows the proPE system in cis orientation, where the PAM sequences are located on the same DNA strand. Adapted from (38), modified.

1.3 The PEAR system

PEAR (Prime Editing Activity Reporter) is a plasmid-based or HEK239T genome-integrated fluorescent reporter system that serves as a practical tool for assessing prime editing activity or isolating cells modified by prime editing. The system is based on a split *mScarlet* or *EGFP* gene separated by an intron sequence, with a mutated splicing site at

one end. By replacing the mutated AC base pair with GT via prime editing, the splicing site is restored, and the fluorescent signal can be detected by flow cytometry, or the corrected sequence in the genome-integrated system can be assessed using next-generation sequencing (NGS) (**Figure 5**).

PEAR is an easy-to-use and cost-effective reporter for testing various prime editing configurations (testing different PBS or RTT lengths, or second nick positions) or new PE-based systems. Using the plasmid-based PEAR system, in addition to editing the endogenous target (and thus using two pegRNAs, one for the PEAR plasmid and the other for the genomic target), it is possible to enrich edited cells by sorting *EGFP*- or *mScarlet*-positive cells (48).

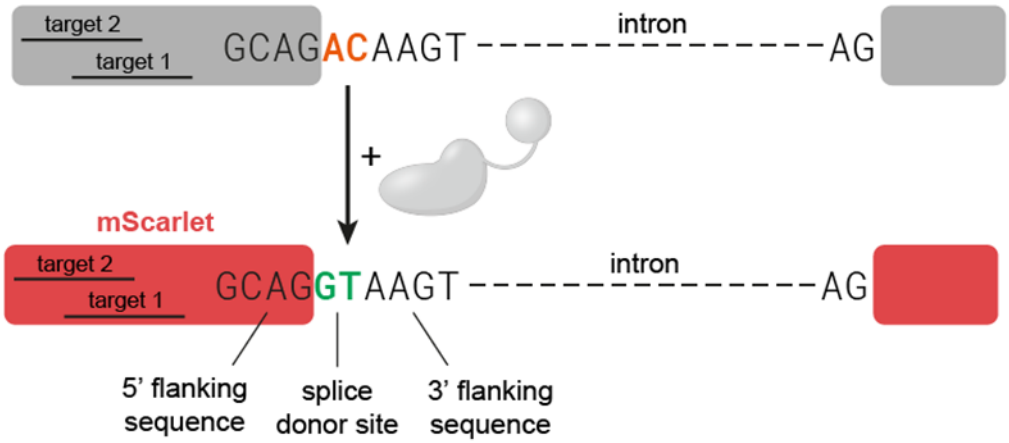

Figure 5. Schematic representation of the mScarlet PEAR system. The two mScarlet segments are separated by an intron sequence; the mutated splicing site is marked in red, and the corrected sequence is marked in green. The two target sequences used for the correction are shown on the left side of the mScarlet gene.

2. Objectives

1. We aim to investigate how reducing intramolecular complementarity of the pegRNA affects the efficiency and specificity of prime editing by:

- **1.1.** introducing mismatches into the spacer or PBS sequences,
- **1.2.** introducing 1–4 nucleotide deletions within the 20-nucleotide PBS sequence,
- **1.3.** combining PBS deletions with spacer mismatches, and
- **1.4**. assessing potential changes in off-target activity.

2. We also aim to evaluate the applicability of the proPE approach:

- **2.1.** in human pluripotent stem cells (hPSCs), and
- **2.2.** at clinically relevant target sites.

3. Methods

3.1 Materials

Restriction enzymes (BpiI (#ER1011) and BsmBI (#ER0451)), T4 ligase (#EL0012), Transcript Aid T7 High Yield Transcription Kit (#K0441), Dulbecco's modified Eagle's medium (DMEM) (#52100-047), Fetal Bovine Serum (#10500064), Turbofect Transfection Reagent (#R0532), Penicillin/Streptomycin (#15140-122), TrypLE Express Enzyme (#12604021), Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix (#A1413302) and Qubit dsDNA HS Assay kit (#Q32854) were purchased from Thermo Fischer Scientific. mTeSR1 media (#85850) and Accutase (#07920) were from STEMCELL Technologies. DNA oligonucleotides and the GenElute HP Plasmid Miniprep kit (# NA0160-1KT) used in plasmid purifications were acquired from Sigma-Aldrich. Q5 High-Fidelity DNA Polymerase (#M0491L), NEB Stable Competent *E. coli* (# C3040H) were from New England Biolabs Inc. NucleoSpin Gel and PCR Clean-up kit (#740609.250) was purchased from Macherey-Nagel. RNA Clean & Concentrator Kit (#R1018) was acquired from Zymo Research.

3.2 Plasmid construction

The PEAR fluorescent reporter target plasmid with *GFP* (pDAS12125_PEAR-GFP, Addgene #177178) and the pegRNA and tpgRNA cloning plasmids (pDAS12222-U6-pegRNA-BFP, Addgene #177181 and pDAS12069-U6-pegRNA-mCherry), were developed by Simon et al. (2022) (48). The *mScarlet* fluorescent reporter plasmid (pAT9752-BEAR-mScarlet, Addgene #162991)) and the 2nd nicking sgRNA and engRNA cloning plasmids (pAT9679-sgRNA-BFP (#162988) and pAT9658-sgRNA-mCherry (#162987)) were constructed by Tálas et al.(2021) (49).

The pegRNA and tpgRNA plasmids were cloned using a two-step cloning procedure: first, the linkers encoding the spacers were cloned into the pDAS12222 or pDAS12069 plasmids between the BpiI restriction sites, 3 units of BpiI enzyme, 2 units of T4 DNA ligase, 500 μM ATP, 1× Green buffer, 50 ng of vector, and 0.25 μM of each of the corresponding oligonucleotides. Next, in a second step, linkers encoding the PBS-RTT sequences were cloned between BsmbI sites under different buffer conditions (1 mM DTT

and 1× Tango buffer), but using the cloning method described above. To monitor transfection efficiency, the sgRNA-expressing plasmids contain a mCherry or TagBFP expression cassette. In some cases, a tevopreQ1 extension was also used for proPE experiments (34).

For prime editing experiments, either the pCMV-PE2 plasmid (created by Anzalone et al. (2019) (23) and acquired from Addgene #132775) or the same plasmid containing the PEmax modifications (R221K and N394K mutations in the SpCas9 nickase) were used as the PE and PEmax expressing plasmids. For non-prime editing experiments or for control samples the X330-Flag-dSpCas9 (Addgene #92113) plasmid was used as the dSpCas9-expressing plasmid, the pX330-Flag-wtSpCas9-H840A (Addgene #80453) as the nSpCas9-expressing plasmid and the pX330-Flag-wtSpCas9 (Addgene #92353) (50) as the WT-SpCas9-expressing plasmid.

3.3 Cell culturing and transfection

The HEK293T (CRL-3216) cells were obtained from ATCC. The HuES9 human embryonic stem cell line was a gift from Dr. Douglas Melton, Harvard University. (HuES9 NIH Approval number: NIHhESC-09-0022 and Health Care Research Council, Human Reproduction Committee in Hungary, Approval number: 6681/2012-EHR.) The cell lines were authenticated by their respective suppliers, and they tested negative for mycoplasma during our experiments.

The HEK293T cells were cultured in DMEM supplemented with 10% heat-inactivated FBS, 100 units/mL penicillin and 100 µg/mL streptomycin, at 37°C in a humidified atmosphere of 5% CO_2. HuES9 cells were maintained in mTeSR1 media on Geltrex coated plates at 37°C and 5% CO_2. HuES cells were passaged every 4-5 days using the ReLeSR cell dissociation reagent and were placed in mTeSR1 media.

In general, transfections of either of the cell types were performed in triplicates and transfected cells were analysed by flow cytometry three days post-transfection. Transfection efficiency was considered in both PEAR and genomic experiments. In PEAR experiments, only those mScarlet/GFP-positive cells were considered successfully edited that were also BFP/mCherry-positive (these were used to determine transfection efficiency and are found on pegRNA plasmids). For genomic experiments, samples were

always analysed by flow cytometry first, followed by genomic DNA purification. When evaluating sequencing data, samples were normalized for transfection efficiency.

3.3.1 Transfection of HEK293T cells

HEK293T cells were seeded on 48-well plates one day before transfection at a density of 3×10^4 cells/well. For experiments using the PEAR system, the molar ratio of 3:5:1:1 of the plasmid components PE, pegRNS, 2nd nicking RNA and PEAR target plasmid, respectively, was used, as a total of 598 ng DNA, or 350 ng DNA per well. For the BEAR-*GFP* and BEAR-*mScarlet* cell lines (described by Tálas et al. (2021) (49), and Simon et al. (2022) (48)) a total of 598 ng of DNA was used at a 4:5:1 (PE:pegRNA:2nd nicking RNA) molar ratio. In case of the genomic targets, the cells were transfected with a total of 350 ng of DNA per well at a 5:4:1 (PE2:pegRNA:2nd nicking RNA or PEmax:pegRNA:2nd nicking RNA for targets: *CDKL1, CTNNB1, CXCR4, IL2RB, HBG, COL7A1 T1, COL7A1 T2, CACNA1C T1* and *CACNA1C T2*)) molar ratio. For the wtSpCas9 experiments 232 ng DNA was used per well at a 1:1.5 SpCas9:sgRNA molar ratio. Before transfection, the DNA was mixed with 1 µL Turbofect reagent diluted in 50 µL serum-free DMEM, and the mixture was incubated for 20-30 min at RT before adding it to the cells.

For proPE experiments a total of 350 ng DNA/well was used (248 ng PE, 62 ng engRNA+tpgRNA and 40 ng 2nd nick sgRNA). The engRNAs and tpgRNAs were used in three different fractions: 1:1, 1:10, 1:40, in these conditions the engRNA amount were as follows: 1:1=31 ng, 1:10=3,1 ng, 1:40=0,8 ng.

3.3.2 Transfection of hPSC cells

HuES9 cells were transfected at ~70% confluence with a total of 1050 ng DNA/cuvette at the same molar ratio as described previously for the genomic targets, in a final volume of 20 µL in a 16-well nucleocuvette strip. Two hours before electroporation, the media was changed to fresh mTeSR1 containing a 10mM ROCK inhibitor (Y-27632) solution, then the cells were singularized using Accutase. 2×10^5 cells were mixed with 20 µL homemade electroporation buffer (described by Vriend et al. (2014) (52)) containing the plasmid DNA, then they were electroporated using a Lonza Amaxa 4D-Nucleofector with the CA-137 program. Transfected cells were plated on Geltrex coated 48-well plates in

500 µL ROCK inhibitor supplemented mTeSR1 and were incubated for 24 h, after which the media was changed to fresh.

3.4 Flow cytometry

Flow cytometric analysis was performed using an Attune NxT Acoustic Focusing Cytometer (Applied Biosystems, Life Technologies). As a general rule, signals from at least 20,000 live cells were recorded, with gating based on side and forward light scattering parameters. The BFP, GFP, mCherry, and mScarlet signals were excited using diode lasers with wavelengths of 405 nm (BFP), 488 nm (GFP), and 561 nm (mCherry and mScarlet), and emission was detected using 440/50 nm (BFP), 530/30 (GFP), 620/15 (mCherry), and 585/16 nm (mScarlet) filters. The data were analyzed using Attune Cytometric Software v.4.2.

3.5 Genomic DNA purification and genomic PCR

After flow cytometry genomic DNA was extracted using the Puregene DNA Purification protocol (Gentra Systems Inc). The amplicons required for next-generation sequencing were generated from genomic DNA samples using a two-round PCR to attach Illumina adapters. (The first-round PCR primers used to amplify the genomic sequences are listed in the appendix.). PCR was done in a PCRmax Alpha AC2 Thermal Cycler using Q5 high-fidelity polymerase with Q5 buffer, and 150 ng of genomic DNA in a total volume of 25 µL. The thermal cycling profile of the PCR was: 98 °C, 30 s; 35 × (denaturation: 98 °C, 20 s; annealing: Tm, 20 s; elongation: 72 °C, 20 s; 72 °C, 5 min). The Illumina adapters were added in a second PCR reaction using Q5 high-fidelity polymerase, Q5 buffer, and 1 µl of the first-step PCR product, for a total volume of 25 µl. The thermal cycling profile of the PCR was: 98 °C, 30 s; 35 × (98 °C, 20 s; 67 °C, 30 s; 72 °C, 20 s); 72 °C, 2 min. After agarose gel electrophoresis, the amplicons were purified using the NucleoSpin Gel and PCR Clean-up Kit, and the samples were quantified and pooled using the Qubit dsDNA HS Assay Kit.

3.6 Next generation sequencing

Sequencing on an Illumina NextSeq instrument was performed by Delta Bio 2000 Ltd. Reads were aligned to the reference sequence using BBMap. The aligned reads were considered as the total reads for each sample.

Indels at the pegRNA and 2nd nicking sgRNA target sites were computationally identified from the aligned reads. Indels were searched for at ± 2 bp around the nick/cut sites. For each sample, indel frequency was determined as (number of reads with an indel)/(number of total reads). The frequency of single substitution mutations without indels generated by prime editing was determined as the percentage of (sequencing reads with the intended modification without indels)/(number of total reads). Conversely, the frequency of intended insertions or deletions generated by prime editing was determined as the percentage of (all sequencing reads with the intended modification)/(number of total reads). For these samples, the indel background was calculated from reads containing different types of indels than the intended edit. In both cases, reads with the intended modifications were identified by searching for a sequence stretch containing the desired edit flanked by 5-5 matching nucleotides. When calculating the edit or indel percentage for each sample, the respective background edit or indel percentage derived from untransfected cells was subtracted.

The following programs were used to analyse the NGS data: BBMap 38.08, Samtools 1.8, BioPython 1.71 and PySam 0.13.

For each CYP gene, gene-specific primers were used and gene-specific reads were identified based on sequence differences between the two genes. Reads derived from non-gene-specific primer-annealing and from mixed PCR products due to template switching were excluded by exploiting two gene-specific motifs located at different positions of the amplicon.

3.7 Analysing literature data for the comparison of effective PBS lengths for pegRNAs and epegRNAs

We analysed Dataset 3, 4, 7 and 8 from the Table S2 in the publication of Yu and Kim et al. (37) to compare the average PE efficiencies of pegRNAs grouped according to their

PBS length between Library-Small and Library-epegRNA, when used with PE2max and PE4max systems. We included editing data in the analyses only from editing via the conventional 'NGG' PAM. In case of some pegRNAs, we identified more and different editing data for a single pegRNA in the data files, therefore we excluded these pegRNAs from our analysis.

3.8 Statistics

To compare two groups, statistical significance was assessed using a two-tailed unpaired or paired t-test. To compare more than two groups, the homogeneity of variances was tested using the Brown–Forsythe test, and the normality of the data was tested using the D'Agostino–Pearson omnibus (K2) test. For normally distributed data sets, statistical significance was assessed using one-way ANOVA and Tukey's post hoc test (when comparing individual groups with each other) or Dunnett's test (when comparing individual groups with a control group). In cases where the data did not meet the assumptions of normal distribution but satisfied the conditions for Box-Cox transformation, the transformed data were analysed as described above. Otherwise, the Kruskal-Wallis test was applied with the Dunn test. Statistical tests were performed using GraphPad Prism 9.1.2 software.

4. Results

4.1 Mismatches in the pegRNA can improve editing efficiency

The efficiency of prime editing depends heavily on the length of the 3'-extension of the pegRNA, particularly the PBS sequence (23,36,37,40). Since prime editing efficiency varies in a target-dependent manner, it is essential to optimize the PBS length for each individual case. As the effective length falls within a fairly wide range (typically between 8 and 15 nucleotides), finding the appropriate PBS can be a considerable burden (23). In addition, Ponnienselvan and colleagues suggested that there might be differences in the preferred PBS length depending on whether a 3' end-protection is added to the pegRNA (40). However, when Sarah Laura Krausz analysed the data from Yu and Kim *et al*. using bioinformatics and compared preferences for prime editing PBS lengths (37), when PE2max or PE4max was paired with conventional pegRNAs or epegRNAs, we found no substantial difference in the range of the most effective PBS lengths (it was between 11 and 14 nucleotides for conventional and end protected pegRNAs as well) (**Figure 6**).

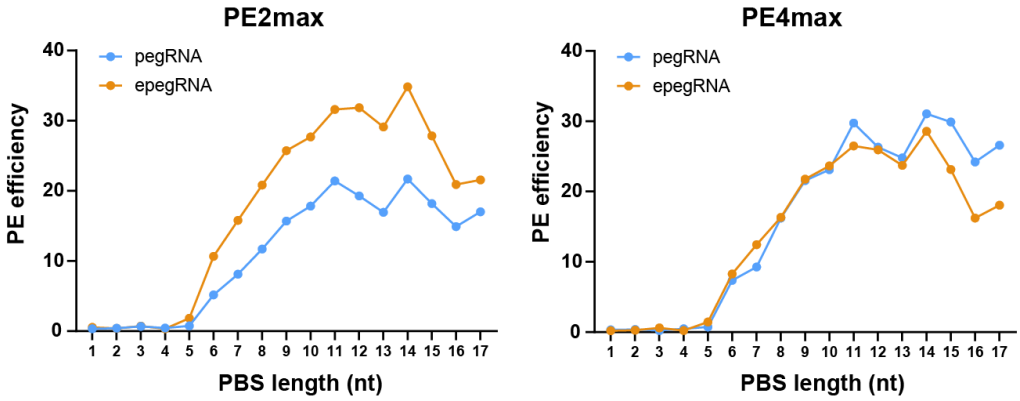

Figure 6. Analysis of data from the study by Yu and Kim *et al*. The average PE efficiency of pegRNAs and epegRNAs with different PBS lengths—which were used with either the PE2max or PE4max systems—was calculated based on data from the Library-Small and Library-epegRNA datasets. Analysis conducted by Sarah Laura Krausz.

Several prediction tools have been developed to address this optimization problem, such as PRIDICT, PRIDICT 2.0, and DeepPrime (37,43–45); and while these can reduce the number of PBS lengths tested, experimental testing of multiple pegRNAs is still required to identify the most effective ones. In general, however, it is true that pegRNAs

with longer PBS lengths (between 15 and 17 nucleotides) often have relatively low editing efficiency. Previous studies have also shown that pegRNAs may adopt a misfolded structure, presumably due to the compatibility between the spacer and PBS sequences, which can reduce editing efficiency (46). This hypothesis is supported by the fact that the splitting of pegRNAs, as well as the prolonged cooling time applied during the re-folding of in vitro transcribed pegRNAs for ribonucleoprotein experiments, was able to increase editing efficiency in some cases (38,46,47).

4.1.1 Mismatches in the pegRNA spacer

Fei *et al*. attempted to reduce the potential indels generated during prime editing by introducing mismatches into the pegRNA spacer sequence, and they achieved some success not only in reducing the number of indels, but also increasing editing efficiency, provided that the mismatches were placed in the appropriate locations (42). However, the effects of PBS sequence length and melting temperature were not considered in their work, which can determine the locations and effects of mismatches. To account for these factors and to investigate the effect of spacer and PBS modifications on spacer-PBS intramolecular complementarity and on prime editing efficiency, we first introduced mismatches into the spacer and then into the PBS sequence, while applying different PBS lengths (**Figure 7a**).

We performed our analysis using the PEAR system, both in a plasmid-based form and in a version integrated into the HEK293T genome. First, we analysed the mismatch tolerance of the SpCas9 protein. It is known that its tolerance varies with targets and spacer sequences, in general, however, it tolerates mismatches occurring at positions distal to the PAM site quite well, although there are exceptions where such variations completely abolish cleavage (15,53,54). In the genome-integrated PEAR system, using two *mScarlet* targets (shifted by 7 nucleotides) and one *EGFP* target, we applied the SpCas9 nuclease with sgRNAs that contained a single mismatch starting from the 10th spacer position. The results obtained by NGS showed slightly reduced but still efficient cleavage for the two *mScarlet* targets, while the *EGFP* target was more sensitive to mismatches (**Figure 7b**). This latter outcome may be associated with the use of a 21G-sgRNA, which contains an additional 5' G nucleotide at the 21st spacer position to ensure efficient expression from the U6 promoter used in plasmids expressing sgRNAs and

pegRNAs. Fu and colleagues have previously shown that extending the spacer sequence in the 5' direction by two Gs increases nuclease specificity; thus, in our case, the EGFP target may exhibit heightened sensitivity to incorporated base pair mismatches due to the effect of the 21-base-pair-long spacer (15).

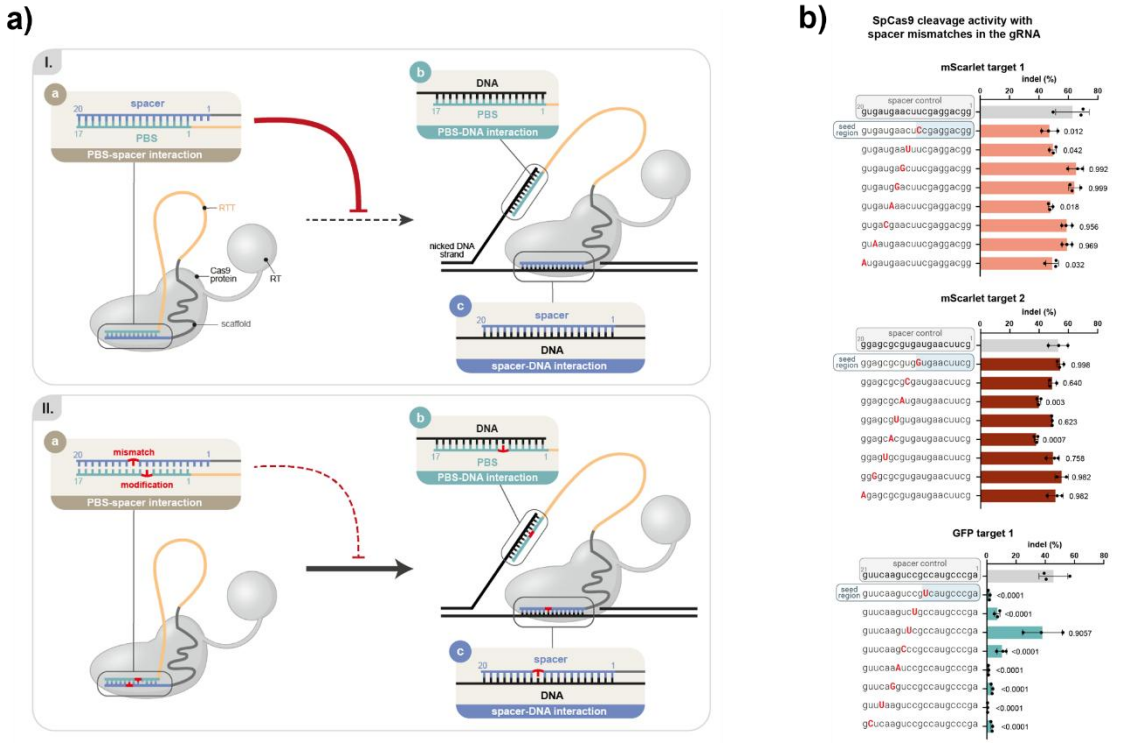

Figure 7. Mismatch sensitivity of SpCas9. a) A schematic diagram illustrating the complementarity between the pegRNA spacer and PBS, as well as methods for its reduction. **b**) SpCas9's tolerance for mismatches at the three PEAR targets integrated into the HEK293T genome. On the vertical axis, the spacer sequences are displayed in 5'-3' orientation; mismatches are indicated by capital letters and highlighted in red. The indel values were determined using next-generation sequencing. The bars show the mean of n=3 samples and the mean ± SD. The P-values are shown above. The differences between the control group and the samples with mismatched spacers using one-way ANOVA and the Kruskal–Wallis test the bars and indicate differences from the non-mismatch control sgRNA.

Based on previous reports in the literature (53), which are supported by the results of the present experiments, that SpCas9 generally tolerates mismatches between the target and the sgRNA spacer sequences well in cases where these are distant from the PAM site, we investigated whether these differences affect the complementarity between the pegRNA spacer and PBS sequences, and if so, how they affect prime editing activity. We designed pegRNAs that target the two *mScarlet* target sequences and whose spacer

contain one, two, or three nucleotide differences, mainly at PAM-distal positions. To evaluate the activity of PE3 in the plasmid-based PEAR system, we used varying PBS lengths (10, 13, 17) for both targets, and to explore the extent of this complementary reducing effect, we used mismatches at positions 8 and 10 of the spacer, as well as in the spacer segment beyond the PBS-complementary region (**Figure 8**). More than half of the mismatches occurring beyond position 10, affecting the PBS-complementary region, significantly increased prime editing efficiency, by up to 7.2-fold, while only one of all the different mismatched pegRNAs reduced editing efficiency (**Figure 8**). Based on these results, even a single mismatch can increase the efficiency of prime editing, provided that the mismatch is located within the PBS-complementary region, as it compensates for (or possibly even overcomes) the negative effects resulting from the disruption of spacer-target pair formation. Mismatches beyond the complementary region had no enhancing effect; in fact, few of them even reduced efficiency, just as variations near the PAM, which had a predominantly detrimental effect (in 7 out of the 12 cases examined), which suggests that, in these cases, the weakening of the process by mismatches was stronger than the positive effect resulting from the reduction in PBS-spacer interactions (**Figure 8**).

Our assumption is that SpCas9 separates the two DNA strands at target sites longer than 20 nucleotides. Thus, elongating the PBS sequence to 20 nt may lengthen the PBS-non-target DNA complex, which could potentially enhance reverse transcription initiation for some targets, while not increasing the complementarity between the spacer and the PBS. We tested these extended PBS sequences on the two *mScarlet* targets as well and observed increased editing efficiency at one of them compared to PBS17; however, when we introduced mismatches into the pegRNAs with 20-nt PBSs, we did not observe an increase in editing efficiency compared to the enhancement of mismatched pegRNAs containing PBS17 (**Figure 8**).

When we introduced multiple mismatches into the spacer region simultaneously, and these mutations fell within the PBS-complementary region, most pegRNAs tested were able to enhance editing efficiency, only a couple of them caused a significant decrease. Thus, although multiple mismatches may also have a positive effect on the efficiency of

prime editing, it appears that the detrimental effect is more frequent than in the case of a single mismatch (**Figure 8**).

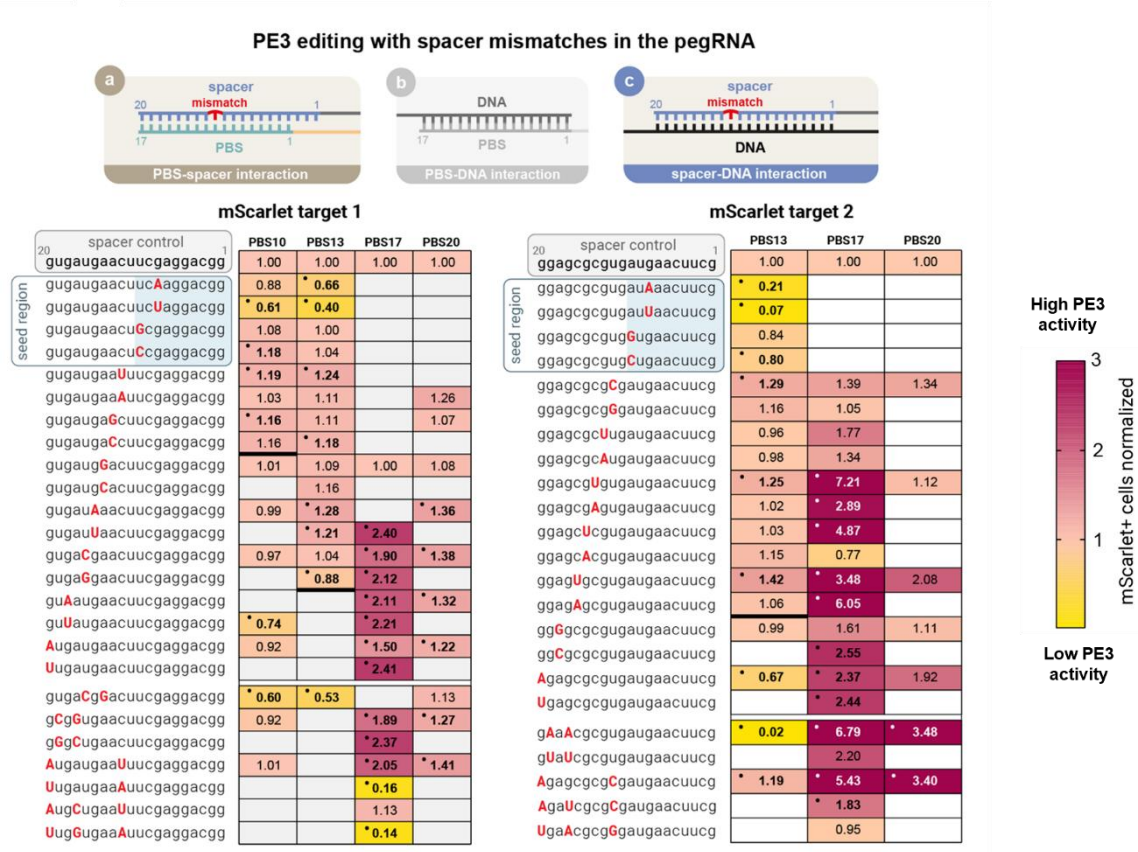

Figure 8. The effect of spacer mismatches on PE3 efficiency for plasmid-based PEAR targets in HEK293T cells. The PE3 efficiency values shown in the heatmaps were normalized to the corresponding, no-mismatch pegRNA control values. A significant deviation from the control is indicated by a black dot (or a white dot for darker-coloured cells) in the upper left corner. The thick horizontal black lines indicate the end of complementarity between the spacer and PBS sequences. The three small schematic diagrams above the heatmaps show the interacting sequences that may be affected by the modifications (those not affected are grey). The spacer sequences are displayed along the vertical axis in a 5′–3′ orientation, with mismatched nucleotides indicated in uppercase letters and highlighted in red. The differences between the control and samples containing a mismatched spacer were assessed for statistical significance using one-way ANOVA, the Kruskal–Wallis test or unpaired two-tailed t-tests.

To investigate whether disrupting intramolecular complementarity has a similar effect in a genomic context, we tested some of the previously used pegRNAs that contained spacer mismatches on *mScarlet* target 1 integrated into the HEK293T genome. In line with previous results, the mismatched pegRNAs tested increased prime editing efficiency by up to twofold, even in the case of shorter PBSs of 10 and 13 nucleotides in length (**Figure 9**).

For the *EGFP* target on which SpCas9 showed little tolerance for mismatches, none of the mismatching pegRNAs effected prime editing positively.

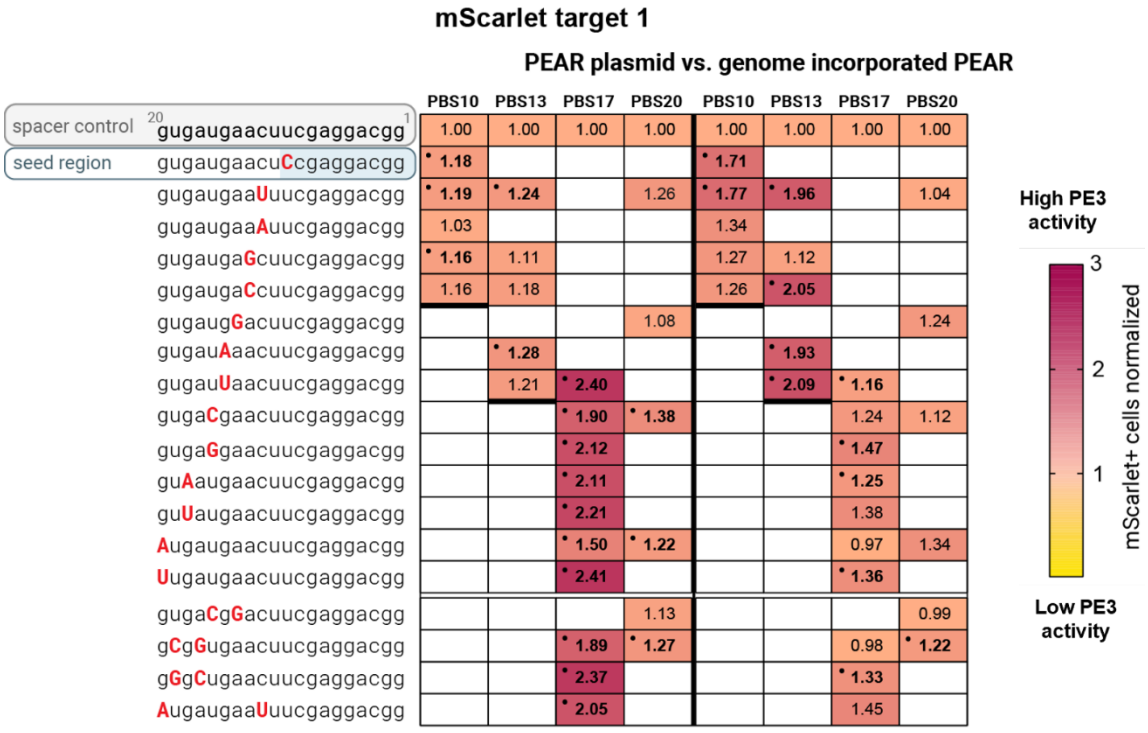

Figure 9. The effect of spacer mismatches on PE3 efficiency for genome integrated PEAR targets in HEK293T cells. The PE3 efficiency values shown in the heatmaps were normalized to the corresponding, no-mismatch pegRNA control values. The plasmid-based PEAR results presented here are from Figure 5. A significant deviation from the control is indicated by a black dot (or a white dot for darker-coloured cells) in the upper left corner. The thick horizontal black lines indicate the end of complementarity between the spacer and PBS sequences. The spacer sequences are displayed along the vertical axis in a 5′–3′ orientation, with mismatched nucleotides indicated in uppercase letters and highlighted in red. The differences between the control and samples containing a mismatched spacer were assessed for statistical significance using one-way ANOVA, the Kruskal–Wallis test or unpaired two-tailed t-tests.

4.1.2 Mismatches in the pegRNA PBS

Incorporating mismatches into the PBS sequence rather than in the spacer, may offer the advantage of lowering PBS complementarity, even in cases where SpCas9 shows limited tolerance to mismatches in the spacer. In this case as well, the effect of mismatches on prime editing is likely position-dependent, as they may weaken the formation and stability of the PBS-non-target strand heteroduplex, thereby hindering the initiation of reverse transcription and, consequently, the efficiency of prime editing. Furthermore, since reverse transcriptase engages with the phosphate groups and sugar backbone of the 3’-end of the DNA primer and the 5’-end of the RNA in the heteroduplex in a non-

sequence-specific manner, mismatches at these positions may reduce RT binding to the heteroduplex and thus reduce the efficiency of editing. Based on above and the cryo-electron microscopy structure of PE (55), we hypothesize that mutations starting from position 5 of the PBS can effectively reduce intramolecular complementarity while not interfering with the process of RT binding and the initiation of reverse transcription.

We re-examined the three plasmid-based PEAR targets using pegRNAs with PBS of varying lengths (10, 13, and 20 nucleotides) that contained one or multiple mismatches in the PBS. For the two *mScarlet* targets, the mismatches in the shorter PBSs increased efficiency in only one of the 23 cases examined, but in most cases significantly reduced it. However, for longer PBSs, we observed an increase in editing that reached as high as 6.7-fold, and in only 4 of the cases studied was a 1.5-fold increase not achieved, which is likely due to the fact that by using PBS20, the RNA–DNA heteroduplex is more stable even in the presence of mismatches (**Figure 10**).

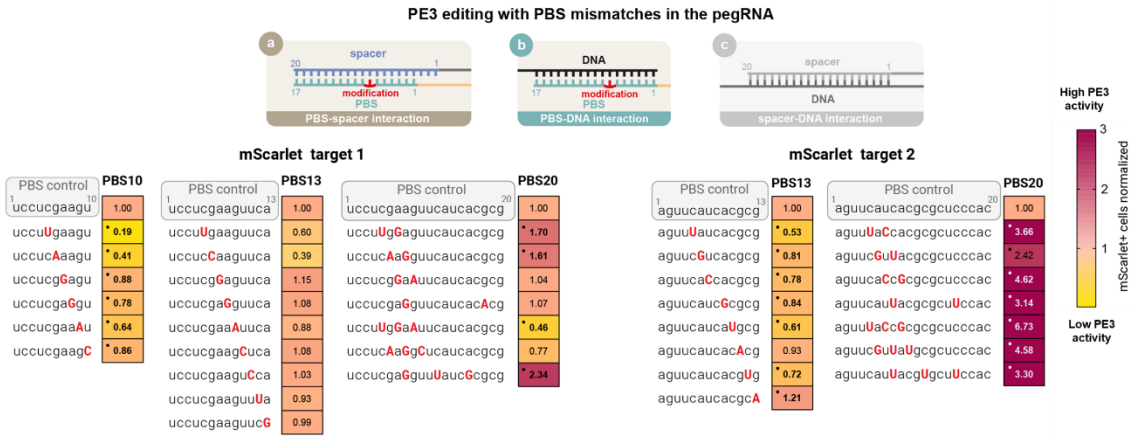

Figure 10. The integration of mismatched pairs into the PBS sequences of the two mScarlet targets. In the case of PBS20, multiple mismatches were incorporated concurrently. The PBS sequences are displayed along the vertical axis in the 5′–3′ orientation, with mismatched nucleotides marked in uppercase and highlighted in red. The PE efficiency values shown on the heatmaps were normalized to appropriate pegRNA controls that do not contain mismatches. Statistically significant deviations from the control are indicated by a black dot in the upper left corner (or a white dot in darker-colored cells). We tested the significance of the differences between the control group and the samples containing inadequate PBS using one-way ANOVA or the Kruskal–Wallis test.

Interestingly, in the case of the *EGFP* target, mismatches placed within the long PBS still reduced editing efficiency.

Based on the results so far, the intramolecular complementarity of the PBS and spacer sequences appears to negatively affect prime editing efficiency, consistent with findings

of previous studies employing different experimental methods (40,46). We have also shown that even a single mismatch can be sufficient to mitigate this suppressive effect, provided it occurs in the complementary region.

4.2 Deletions in the PBS sequence and the combination of spacer and PBS modifications

4.2.1 Deletions in the pegRNA PBS

We then investigated whether deletions introduced into the long (20-nucleotide) PBS sequence exhibit the same effect as mismatches. We introduced deletions of 1 to 4 nucleotides, starting at position 6 of the PBS, into the pegRNAs targeting the two *mScarlet* sites in the plasmid-based PEAR system (**Figure 9**). As controls, we also tested pegRNAs containing a shorter PBS region corresponding to the intact sequence located upstream of the implanted deletions. According to this rule, if the deletions started at position 10, we used a pegRNA with a 9-nt PBS as a control, as well as pegRNAs with a long PBS that did not contain a deletion. As observed with mismatches, deletions of different sizes and of various positions were also capable of increasing prime editing efficiency relative to the pegRNA with a long PBS, particularly at the *mScarlet* target 2 site. Here, pegRNAs containing a deletion in the PBS region were even able to surpass the efficiency of the pegRNA optimized for this target site. Since deletions starting at position 13 resulted in the highest efficiency for both targets, we selected these variants for subsequent experiments (**Figure 11**).

In most cases, an intact sequence remains in the PBS following the deletion of the nucleotides, which we hypothesized contributes to the stability of the heteroduplex between the RNA and the non-target DNA strand. To investigate whether this remaining 3'-end sequence plays any role, we selected three pegRNAs containing a deletion and scrambled this remaining region of the PBS. This yielded conflicting results: compared to the non-scrambled, but deletion containing variants and the optimal pegRNAs, the rearranged sequence either disrupted, enhanced, or had no effect on prime editing activity. This suggests that the 3′ region of the PBS influences prime editing efficiency in ways other than simply through binding between the PBS and the non-targeted DNA strand,

likely by affecting pegRNA stability or protein-RNA interactions, although to varying degrees among different pegRNAs.

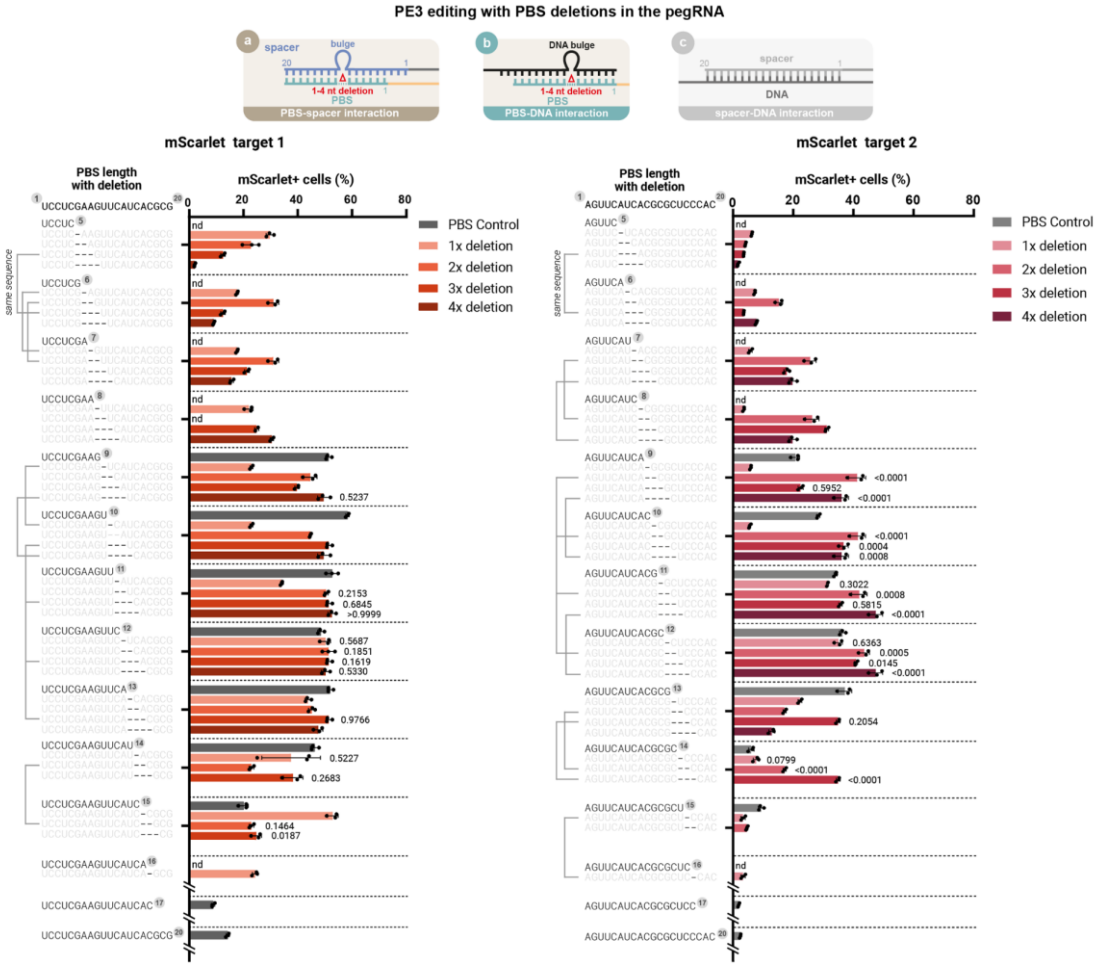

Figure 11. The effect of 1–4 nt deletions within the long (20 nt) PBS sequence for the two mScarlet targets. Experiments were conducted using PE3 in the plasmid-based PEAR system in HEK293T cells. The three small schematic diagrams positioned above the heat maps and bar charts depict the interacting sequences potentially affected by the modifications, while unaffected regions are shown in grey. PBS sequences are presented along the vertical axis in a 5′–3′ orientation. Sequences shown in black denote the PBS segments that remained intact upstream of the deletion site(s), with the number following each sequence indicating its length; these also correspond to the PBS sequences of the pegRNAs used as controls. Thin black lines on the left side of the panels connect cases in which distinct deletions yielded identical sequences. Bars represent the mean and mean ± SD of n = 3 samples. P values are indicated above the columns and reflect deviations from the no-deletion control. Differences between pegRNAs with the control PBS length and samples containing PBS20 with deletion(s) were assessed for statistical significance using either one-way ANOVA or the Kruskal–Wallis test.

4.2.2 Combining spacer and PBS modifications

We next examined whether the effects of spacer mismatches and PBS deletions are cumulative at the two *mScarlet* targets in the plasmid-based PEAR system. We selected one- and two-base deletions starting at position 13 (located within a 20-nt PBS) that performed well on both targets and paired them with mismatches that are not directly opposite these deletions and are located at PAM-distal positions (**Figure 12**). For *mScarlet* target 2, all combinations significantly enhanced prime editing efficiency compared to pegRNAs containing only deletions. At the other target, none of the combinations reduced efficiency, although no improvement was observed either. We obtained similar results when comparing the efficiency of pegRNAs with combined modifications to those with optimal PBS lengths (**Figure 12**). Thus, mismatches placed in the spacer and small deletions within the long, 20-nt PBS may represent an successful approach to reduce intramolecular complementarity, which we further investigated in a genomic context.

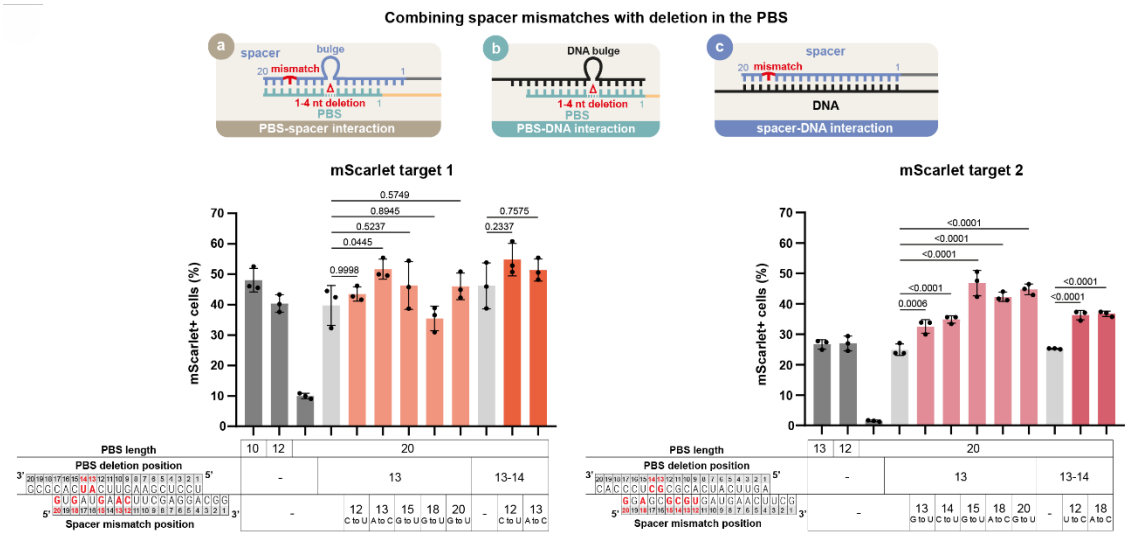

Figure 12. The effect of combining 1–2 nt deletions in PBS20 with spacer mismatches on PE3 activity in the plasmid-based PEAR system and in HEK293T cells. a) Altered positions within the PBS and spacer are highlighted in red. P values are indicated above the columns and represent deviations from the control containing only the deletion. The three small schematic diagrams at the top depict the interacting sequences that may be affected by the modifications.

We selected 10 genomic targets previously described in the literature (23) and compared the spacer- and PBS-modified pegRNAs with the ones used in the original

study in HEK293T cells. Each newly designed pegRNA contained a one-nucleotide deletion within the 20-nt PBS at position 13, along with an additional one-nucleotide mismatch (transversion or transversion) at position 13, 15, or 18 of the spacer. Of the 10 edits, 2 outperformed the optimal pegRNAs reported in the literature (**Figure 13)**. This, together with our previous results, suggests that the combination of spacer and PBS modifications is likely to further increase editing efficiency compared to conventional pegRNAs, although exhaustive optimization is required to identify the most effective combinations.

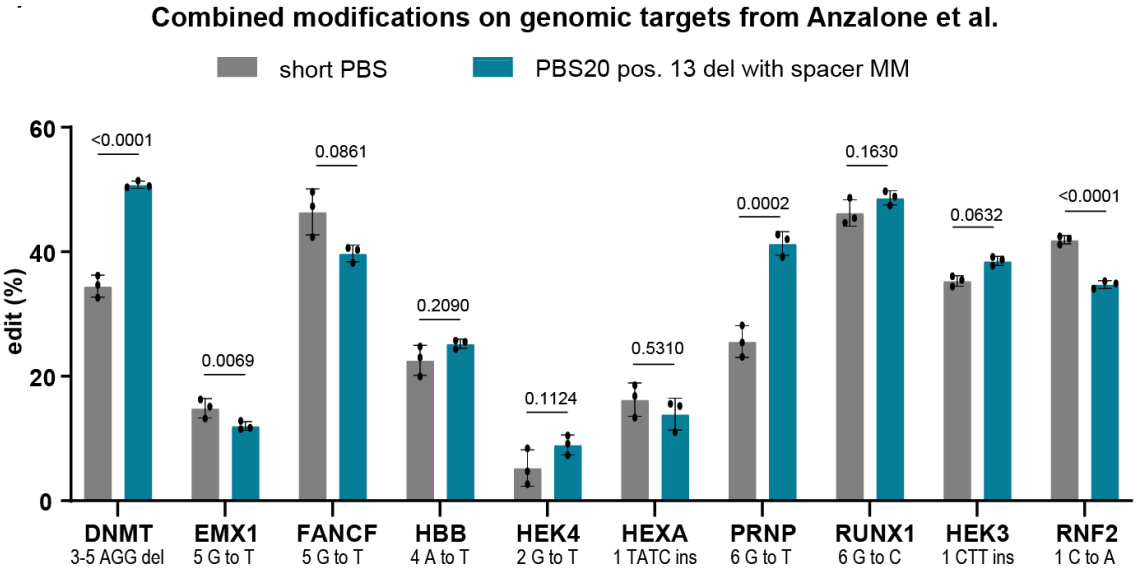

Figure 13. The effect of combined modifications of the spacer and PBS on endogenous targets. Effect of combining spacer mismatches (transitions or transversions) with a single-nucleotide deletion at position 13 in PBS20 across 10 endogenous targets in HEK293T cells. Shown here are PE3 editing outcomes for pegRNAs derived from Anzalone *et al.* with shorter PBS lengths (*DNMT*: PBS13, *EMX1*: PBS15, *FANCF*: PBS14, *HBB*: PBS8, *HEXA*: PBS12, *PRNP*: PBS12, *RUNX1*: PBS17, *HEK3*: PBS13, *RNF2*: PBS15), as well as for the best-performing combinations. Bars represent the mean and mean ± SD of n = 3 samples. Values were determined using next-generation sequencing. Differences between samples were assessed for statistical significance using one-way ANOVA or the Kruskal–Wallis test.

4.3 The impact of a single nucleotide deletion in the PBS on editing efficiency and off-target activity

When we compared pegRNAs containing combined modifications with pegRNAs containing only a deletion, in which the deletion also occurred only at position 13 of the 20-nt PBS, the results showed that additional spacer mismatches are not necessarily

required, as only pegRNAs at the *PRNP* target site exhibited increased editing efficiency compared to the control containing only a deletion. In most cases, pegRNAs containing only the deletion have already achieved the editing efficiency of the versions described in the literature, suggesting that this approach could reduce the burden associated with optimizing combined modifications, and even spare the PBS-length optimization step altogether. This solution has been named SPELL (Streamlined Prime Editing with fixed-Length PBS Leverage) (**Figure 14a**).

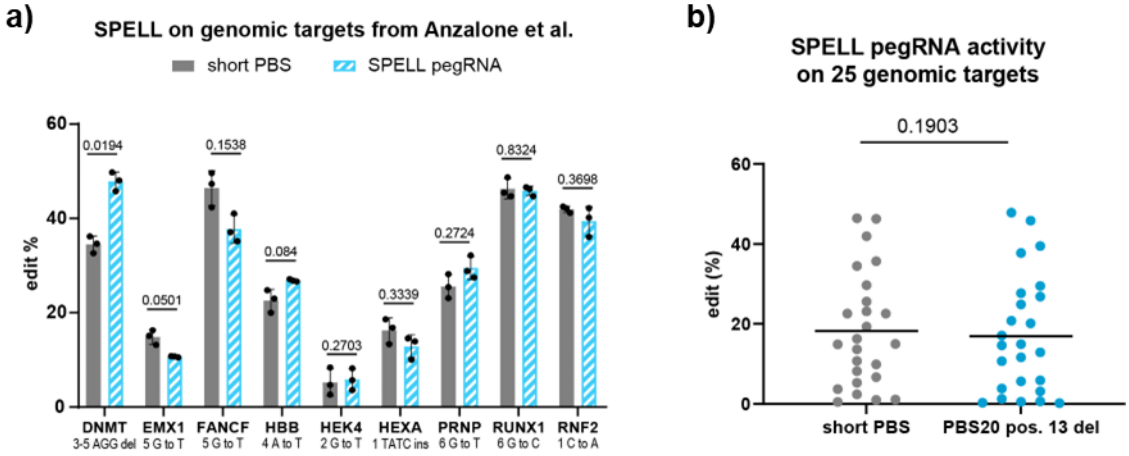

Figure 14. The activity of SPELL pegRNA on endogenous targets. a) Results are from the same experiment as Figure 10., here only the short PBS pegRNAs and deletion-only pegRNAs shown. Bars represent the mean and mean ± SD of n = 3 samples. **b**) Summarized short PBS and SPELL pegRNA PE3 editing results on genomic targets. **a, b**) Values were determined using next-generation sequencing. Differences between samples were assessed for statistical significance using one-way ANOVA or the Kruskal–Wallis test.

To further investigate this approach, we examined seven previously tested pegRNAs (38) as well as nine additional targets, five of which had been previously characterized (26,56), to see if we could achieve the prime editing efficiency of the original pegRNAs using SPELL pegRNAs. In both groups, more than half of the SPELL pegRNAs exhibited editing efficiency comparable to that of the original pegRNAs with shorter PBSs. In total, we tested 25 targets in the HEK293T genome and observed no significant difference between the overall activity of the previously characterized pegRNAs and the SPELL pegRNAs (**Figure 14b**). In addition, we evaluated two targets in a hPSC cell line (in the *LRRK2* and *GBA1* genes) (57), observing a 1.5-fold increase in editing efficiency for the

GBA1 target with a PBS containing a single deletion, while no significant change was detected for the *LRRK2* gene (**Figure 15a**).

Furthermore, in the analysis of two targets with known off-target sites (*HEK4* and *HBB*) (23,58), where not only off-target indel formation but also off-target prime editing activity was observed, during PE2 editing the use of SPELL pegRNAs did not result in increased off-target activity compared to the previously tested pegRNAs (**Figure 15b**).

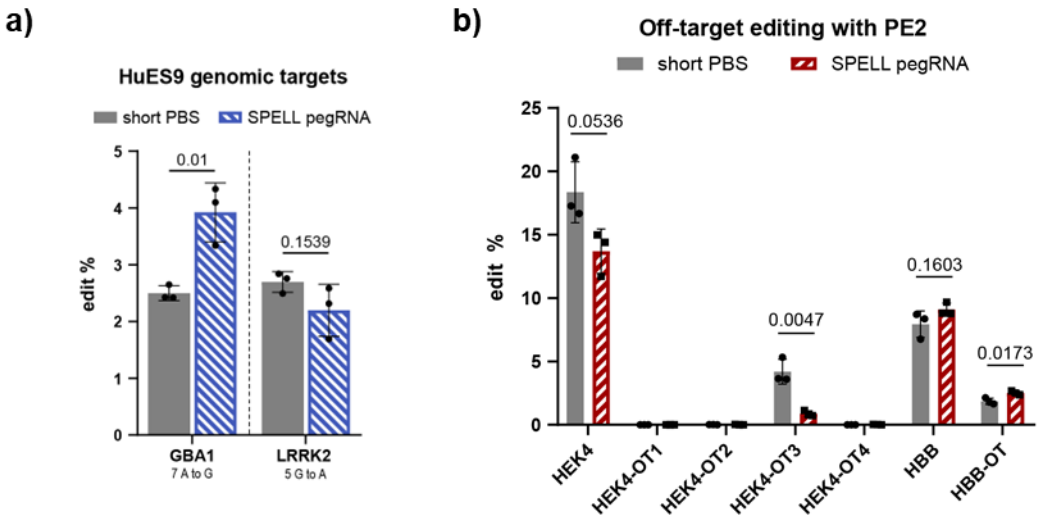

Figure 15. Activity of SPELL pegRNAs in a hPSC cell line and off-target effects. a) PE3 editing at two genomic targets in a hPSC line using pegRNAs with short PBS (PBS13) or deletion-only SPELL pegRNAs. **b**) Off-target editing assessed with PE2 in HEK293T cells at previously characterized off-target sites, HEK4 (22) and HBB (55), using pegRNAs with short PBSs (*HEK*4: PBS8; *HBB*: PBS13) and SPELL pegRNAs. **a, b**) Bars represent the mean and mean ± SD of n = 3 samples. Values were determined using next-generation sequencing.

4.4 SPELL pegRNAs do not act by restoring nicking activity

At the beginning of our work, Éva Bakos conducted an experiment in which we compared the cleavage activity of the SpCas9 nuclease using conventional sgRNAs and pegRNAs containing a long (15–17 nucleotide) PBS sequence. The results showed that even at targets where prime editing has high activity with these pegRNAs, albeit with varying efficiency, cleavage activity is reduced compared to sgRNA, leading to the conclusion that the PBS-spacer complementarity in the pegRNAs may affect the nicking function (**Figure 16a**).

However, when we compared the cleavage activity of SpCas9 using pegRNAs containing PBSs of different lengths, one of which was the SPELL pegRNA containing a deletion, we found that although these pegRNAs yielded prime editing results of varying efficiency at multiple targets, their cleavage activity did not differ significantly (**Figure 16b, c**). This surprising result suggests that this self-complementary effect does not necessarily influence nicking activity, but rather affects another step of prime editing, for example efficient RT priming. We have previously observed a similar phenomenon when the RTT of the pegRNA interacted with the stem-loop 2 (SL2) unit of the scaffold, and although this reduced PE efficiency, cleavage activity was maintained (59). Both findings suggest that a more comprehensive investigation of the prime editing process is needed to track the precise effects of pegRNA modifications.

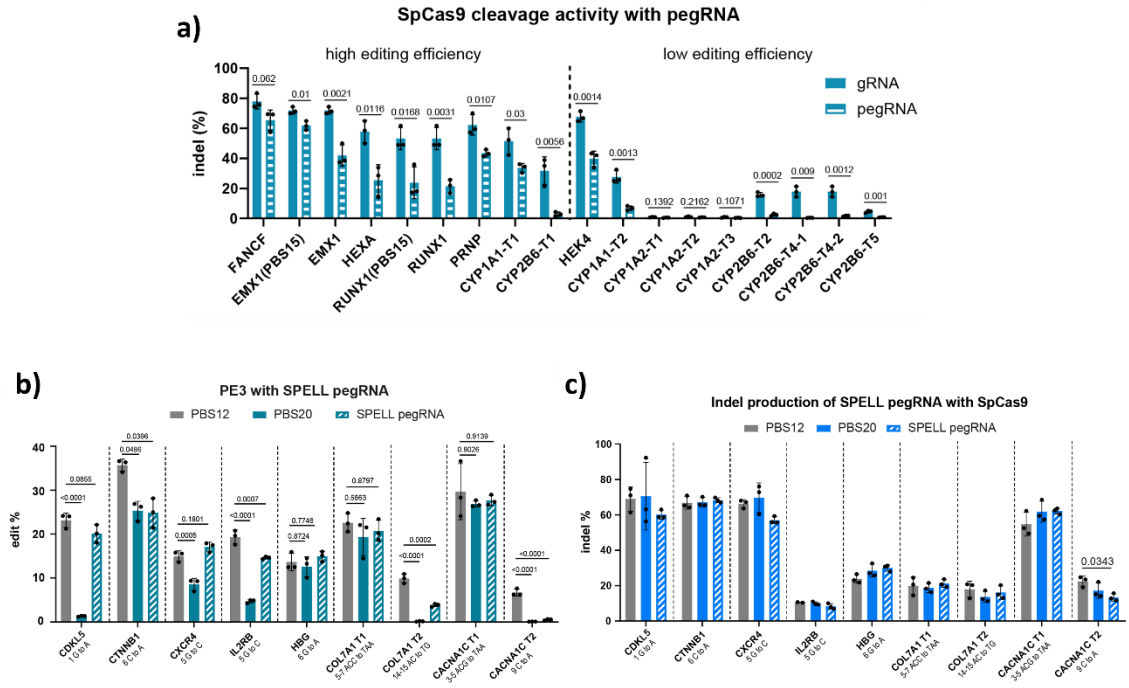

Figure 16. SpCas9 cleavage efficiency of sgRNAs and pegRNAs. a) Comparison of SpCas9 cleavage activities on endogenous targets in HEK293T cells for conventional sgRNAs and pegRNAs with 17 nt long PBS (for the RUNX1-1, CYP1A1, CYP1A2, and CYP2B6 targets epegRNAs were used). Indel values were determined by using next generation sequencing. experiment in HEK293T cells were conducted by Éva Bakos. **b, c**) On every target a short, 12 nucleotides, a long, 20 nucleotide and SPELL pegRNA was used. The P-values are shown above the columns and indicate the differences relative to the PBS12 control. The bars show the mean of the n=3 sample and the mean ± SD. The values were determined using next-generation sequencing. We tested the significance of the differences between the samples using unpaired two tailed t-tests one-way ANOVA or Kruskal-Wallis test

4.5 ProPE enhances editing efficiency on low performing and target-distal edits

The proPE method, which divides the functions of pegRNA between two guide RNAs, enabled efficient prime editing even without extensive optimization, as this system appears to work well even with 17-nucleotide-long PBSs, where the PBS-spacer complementarity is likely the strongest and which are thus often ineffective. It also appears to be less sensitive to the editing position. While conventional PE showed declining efficiency the further the position of the edit was located from the nick position proPE demonstrated increased efficiency even at editing positions distant from the nick site (edits located >10 positions downstream, **Figure 3**), where prime editing typically exhibits low editing efficiency (below 5% in HEK293T cells), thereby successfully broadening the editing window of prime editing (38).

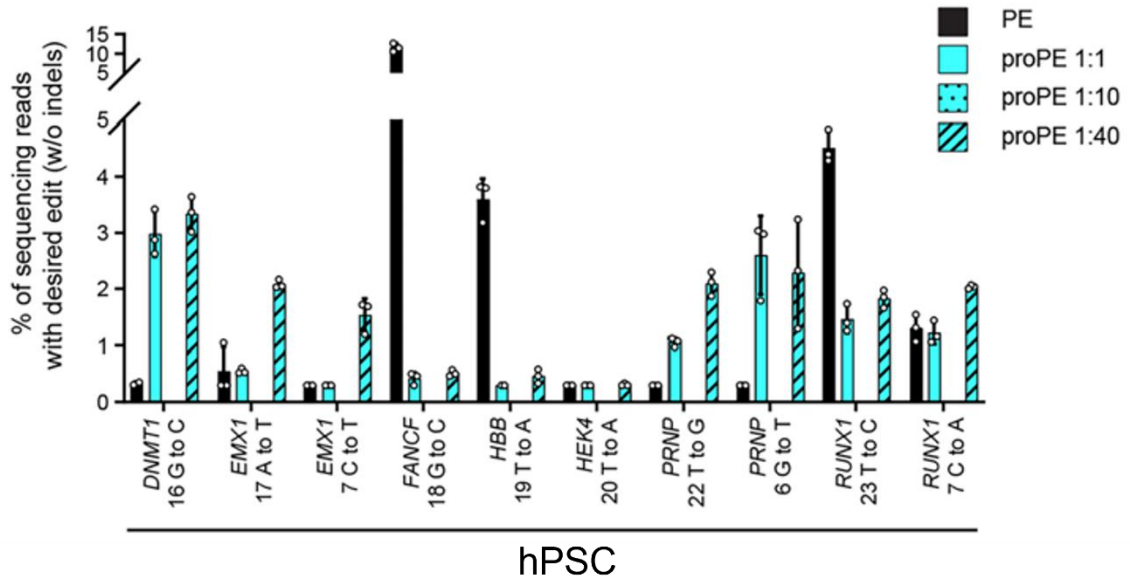

Figure 17. Characterization of proPE efficiency on endogenous targets in a hPSC cell line
We tested proPE (using PE5max protein and tpgRNA end protection) on 10 endogenous targets in an hPSC cell line at three different engRNA ratios. Results were obtained via NGS, and data are presented as mean values of n = 3 samples or as mean ± SD. Pe5max with epegRNAs was used as a control.

To determine whether the same effects could be observed in other cell lines and to investigate the broader applicability of the system, 10 endogenous targets were tested in a hPSC cell line; and for the majority of these, editing had to be performed at a position distant from the target. Prime editing is quite efficient when applied in HEK293T cells,

likely due to weaker mismatch repair, however, hPSC cells are generally more difficult to transfect, and prime editing is typically less active. With proPE however, improvements in editing efficiency and specificity (indel/edit ratio) were observed in more than half of the tested targets at at least one of the three tested engRNA ratios (the optimal ratio was found to be 1:40) (**Figure 17**).

This feature is particularly useful, as more than half of known human pathogenic SNPs are located farther from the target site than can be targeted with conventional SpCas9-based PE (43). To demonstrate the power of proPE, we compared its activity to that of PE by editing multiple pathogenic SNPs located in the RYR2 and SCN5A genes, which are associated with cardiovascular diseases, as well as in the KRT12 gene, which causes hereditary corneal dystrophy. All edits were located more than 10 base pairs downstream of the cleavage site, and in each case, PE was compared with two different tpgRNAs under proPE conditions. With the exception of the KRT12 target, proPE exhibited increased activity on all other targets with at least one of the two tested tpgRNAs, and in most cases, editing specificity was also higher with proPE, demonstrating its potential (**Figure 18**).

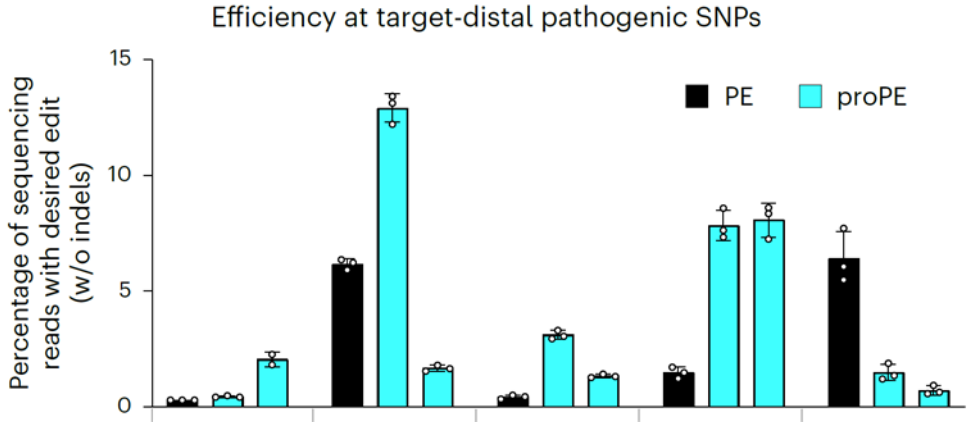

Figure 18. The efficiency of proPE on human pathogenic targets. The efficacy of ProPE (using PE5max protein and tpgRNAs for end protection) was tested on five pathogenic SNPs, each with two tpgRNAs, using three different engRNA ratios (1:1, 1:10, 1:40). For each target, only the best-performing ratios are shown. Results were obtained by NGS, and data are presented as mean values from n = 3 samples and as mean ± SD. Pe5max with epegRNAs was used as a control.

5. Discussion

Although the discovery and development of CRISPR/Cas-based systems have revolutionized the field of gene editing, prime editing stands out among them due to its versatility. However, inconsistencies in its efficiency continue to pose a challenge to its wider application.

The appropriate structure of the prime editing guide RNA, as well as the carefully optimized lengths of the PBS and RTT sequences, are crucial for efficient editing. Based on our experiments, the presence of the 3'-end of pegRNA weakens its cleavage activity compared to conventional gRNAs when using the SpCas9 nuclease, suggesting that an inhibitory effect occurs upon extension of the guide RNA. The complementarity of the pegRNA appears to be a logical explanation, as the spacer and PBS sequences share a significant complementary region (depending on the length of the PBS). We found that the efficiency of prime editing can be improved by using mismatching spacer or PBS sequences, particularly when using long (17- or 20-nucleotide) PBS sequences, which typically result in reduced editing efficiency and possess the longest complementary region to the spacer. Although this is not necessarily the primary limiting factor in all cases, PBS-spacer complementarity can clearly influence the prime editing process. These results support the findings of Fei and colleagues, who incorporated mismatches into the pegRNA spacer to reduce potential re-cleavage of the edited target sequences and thereby mitigate the formation of indels during prime editing. Their results also showed that these mismatches can improve editing efficiency when they occur in the complementary regions of the PBS and the spacer, suggesting a broader impact of mismatches (42). However, it is clear that the most effective mismatches and their positions are difficult to predict, and finding the best approach is likely to depend on the target sequence and require significant effort. Deletions in the long (20-nucleotide) PBS sequence appeared to be more reliable; a single-nucleotide deletion at position 13 generally had a positive effect on editing efficiency, although the exact mechanism of this effect has not yet been identified.

Previous studies have confirmed that the correct structure and proper expression of pegRNA are crucial for efficient prime editing. For conventional, in vitro transcribed pegRNAs, editing efficiency was increased when an optimized folding protocol was used to increase the proportion of properly structured pegRNAs during their use in RNP

complexes (46). Li and colleagues demonstrated that stabilizing the pegRNA structure by adding an extra G/C pair to one of the scaffold loops, or by replacing a few non-G/C pairs with G/C pairs within the same loop, resulted in higher editing efficiency for indel edits. These types of pegRNAs were named apegRNAs (60). Regarding the complementarity of the PBS spacer, an alternative solution had previously been proposed. During the development of ePE, Liu and colleagues further extended the pegRNA sequence with a Csy4 recognition site. They hypothesized that this 20-nucleotide-long extra sequence, which could form a hairpin-like structure, would prevent the ends of the pegRNA from folding the molecule into a circular shape. Using this method, they observed a significant increase in the efficiency of prime editing (61). Similarly, epegRNAs also utilize structured RNA motifs, although here it is for the purpose of protecting against cellular exonucleases, which can cleave the 3'-terminus of the epegRNA. In their article, Nelson and colleagues discuss that a 3'-pseudoknot sequence (such as evopreQ1 used for epegRNAs) does protect these RNAs from degradation and increases editing efficiency by reducing the number of truncated, ineffective pegRNAs (34). Similar results were reported by Zhang and colleagues, who extended the pegRNA with Xrn1-resistant RNA motifs (62). Like ePE pegRNAs, these may also help stabilize the pegRNA structure.

Although SPELL pegRNAs do not employ structured 3’ extension and do not modify the backbone structure, they may also promote the correct conformation of the pegRNA by reducing the complementarity of the PBS spacer. It would be interesting to compare or even combine the SPELL pegRNA method with the aforementioned pegRNA variants, in part to see if there is an additive effect from the different approaches, or if the SPELL method can also assist in the PBS optimization step in these cases. It would also be worthwhile to investigate the mechanisms underlying SPELL and whether this method works in other model organisms, so that it could be further developed and applied more widely.

ProPE also could provide a solution for pegRNA self-complementarity, although it aims to address multiple inhibitory factors affecting prime editing efficiency simultaneously by separating the cleavage and RT-priming functions of pegRNA, One of the method’s greatest achievements is that it enables the editing of sequences distant from the target, thereby opening the possibility of correcting numerous human pathogenic SNPs. It may be argued that the use of two target sites limits its application, as it requires

two PAM sequences located relatively close to each other; however, characterization of proPE revealed that the system was functional even with two PAMs spaced approximately 45–70 base pairs apart (depending on whether the PAMs were on the same or opposite strands), making the system quite flexible. Certainly, PAM-flexible PE variants also offer the possibility of targeting closer pathogenic SNPs; however, their efficiency is often quite low, and when compared to proPE SpRY-based PE and proPE showed declined efficiencies, making them less desirable options (38).

Given that prime editing is a constantly evolving technique, I believe that the methods and results presented in this study contribute to improving the usability of prime editing, expand its scope of application, and deepen our understanding of the underlying mechanisms of the process.

6. Conclusions

In this work, I aimed to investigate the effect of reduced complementarity within the pegRNA on the efficiency of prime editing, as well as the efficiency of the proPE method for edits located distant from the target (>10 positions downstream from the nick) and on human pathogenic SNPs.

- When we introduced mismatches into the pegRNA spacer while using PBS sequences of different lengths (10, 13, 17, and 20 nucleotides), experiments conducted in the plasmid-based PEAR system revealed that mismatches located in the second half of the spacer but still within the spacer-PBS complementary region can improve prime editing efficiency, although the effects were inconsistent. The greatest improvement was observed for 17-nucleotide-long PBSs. The positioning of mismatches within the PBS sequence was mostly detrimental when using shorter (10- or 13-nucleotide) sequences, whereas for 20-nucleotide-long PBSs, we observed improvements, albeit to varying degrees.
- Furthermore, we investigated modifications to a 20-nucleotide-long PBS sequence in the plasmid-based PEAR system, this time using 1–4-nt deletions starting at position 6. We observed varying effects depending on the length and location of the deletions; however, in some cases, the pegRNAs containing the deletion were able to match or even exceed the efficiency of the pegRNA with the PBS length optimal for the given targets.
- When we examined mismatches in the spacer combined with 1- or 2-nt deletions in the PBS in the PEAR system and later on multiple endogenous targets, we found that although in certain cases the mismatches could help enhance the effect of PBS deletions, a single deletion at position 13 of the 20-nt-long PBS achieved optimal pegRNA efficiency for the majority of targets. We named this method SPELL pegRNA and conducted further experiments with it on endogenous targets (25 in total) and an hPSC cell line; we found that in more than half of the cases, SPELL pegRNAs were just as effective as pegRNAs with short PBS sequences.
- When examining the off-target activity of SPELL pegRNAs, we also observed no notable differences compared to previously characterized pegRNAs. Overall, SPELL

pegRNAs are capable of achieving effective prime editing even without extensive optimization.

- Finally, we evaluated proPE for edits located distal to the target and found that, in these cases, proPE enables more efficient editing compared to conventional PE, thereby allowing the editing of hard-to-reach pathogenic SNPs.

7. Summary

Prime editing is an extremely practical but not particularly efficient genome editing method, as its efficiency can be limited by numerous bottlenecks arising during the complex process, and its effective application requires thorough optimization. One factor hindering efficiency is the complementarity of the pegRNA spacer and PBS sequences. Effective editing of every target molecule requires PBSs of varying lengths, which represents a considerable optimization burden; however, longer PBSs (15–17 nucleotides) generally exhibit lower efficiency for the majority of target molecules, likely due to the longer complementary segment associated with the spacer.

We found that by reducing this complementarity, either by introducing mismatches into the spacer or PBS sequences, or by introducing small deletions in a long (20 nt) PBS, or by combining spacer mismatches with PBS deletions, effective prime editing efficiency can be achieved, which sometimes even exceeds the optimal pegRNA efficiency at the given target. However, by introducing a single-nucleotide deletion at the 13th PBS position, we achieved the efficiency of pegRNAs with short PBSs previously characterized at the target site without increasing off-target activity, suggesting that this method could potentially reduce the burden of pegRNA optimization.

We also tested the split prime editor proPE on endogenous targets where the potential edit site was located distally relative to the nick site (>10 positions downstream). We successfully edited the majority of these endogenous and clinically relevant targets in two human cell lines, providing a useful method that could potentially make more than half of known, hard-to-reach human pathogenic SNPs editable.

8. References

1. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012 Aug 17;337(6096):816–21. doi:10.1126/science.1225829

2. Knott GJ, Doudna JA. CRISPR-Cas guides the future of genetic engineering. Science. 2018 Aug 31;361(6405):866–9. doi:10.1126/science.aat5011

3. Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nature Biotechnology. 2020 Jul 1;38(7):824–44. doi:10.1038/s41587-020-0561-9

4. Zhou L, Yao S. Recent advances in therapeutic CRISPR-Cas9 genome editing: mechanisms and applications. Mol Biomed. 2023 Apr 7;4(1):10. doi:10.1186/s43556-023-00115-5

5. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat Biotechnol. 2014 Dec;32(12):1262–7. doi:10.1038/nbt.3026

6. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science. 2014 Jan 3;343(6166):80–4. doi:10.1126/science.1246981

7. Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han R, Myacheva K, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat Biotechnol. 2016 Feb;34(2):192–8. doi:10.1038/nbt.3450

8. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013 Feb 15;339(6121):819–23. doi:10.1126/science.1231143

9. Jasin M, Rothstein R. Repair of Strand Breaks by Homologous Recombination. Cold Spring Harb Perspect Biol. 2013 Nov;5(11):a012740.

doi:10.1101/cshperspect.a012740 PubMed PMID: 24097900; PubMed Central PMCID: PMC3809576.

10. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol. 2016 Mar;34(3):339–44. doi:10.1038/nbt.3481

11. Paquet D, Kwart D, Chen A, Sproul A, Jacob S, Teo S, et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature. 2016 May;533(7601):125–9. doi:10.1038/nature17664

12. Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol. 2015 May;33(5):543–8. doi:10.1038/nbt.3198

13. Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell. 2017 Jan 12;168(1):20–36. doi:10.1016/j.cell.2016.10.044 PubMed PMID: 27866654.

14. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol. 2018 Aug;36(8):765–71. doi:10.1038/nbt.4192

15. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013 Sep;31(9):822–6. doi:10.1038/nbt.2623

16. Kim D, Luk K, Wolfe SA, Kim JS. Evaluating and Enhancing Target Specificity of Gene-Editing Nucleases and Deaminases. Annual Review of Biochemistry. 2019 Jun 20;88(Volume 88, 2019):191–220. doi:10.1146/annurev-biochem-013118-111730

17. Kovalev MA, Davletshin AI, Karpov DS. Engineering Cas9: next generation of genomic editors. Appl Microbiol Biotechnol. 2024 Feb 14;108(1):209. doi:10.1007/s00253-024-13056-y

18. Kulcsár PI, Tálas A, Ligeti Z, Tóth E, Rakvács Z, Bartos Z, et al. A cleavage rule for selection of increased-fidelity SpCas9 variants with high efficiency and no detectable off-targets. Nat Commun. 2023 Sep 16;14(1):5746. doi:10.1038/s41467-023-41393-5

19. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016 May;533(7603):420–4. doi:10.1038/nature17946

20. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature. 2017 Nov;551(7681):464–71. doi:10.1038/nature24644

21. Dadush A, Merdler-Rabinowicz R, Gorelik D, Feiglin A, Buchumenski I, Pal LR, et al. DNA and RNA base editors can correct the majority of pathogenic single nucleotide variants. npj Genom Med. 2024 Feb 26;9(1):16. doi:10.1038/s41525-024-00397-w

22. Wang D, Zhang Y, Zhang J, Zhao J. Advances in base editing: A focus on base transversions. Mutation Research - Reviews in Mutation Research. 2024 Jul 1;794:108515. doi:10.1016/j.mrrev.2024.108515

23. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019 Dec;576(7785):149–57. doi:10.1038/s41586-019-1711-4

24. BenDavid E, Ramezanian S, Lu Y, Rousseau J, Schroeder A, Lavertu M, et al. Emerging Perspectives on Prime Editor Delivery to the Brain. Pharmaceuticals. 2024 Jun 11;17(6). doi:10.3390/ph17060763

25. Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen PF, et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell. 2021 Oct 28;184(22):5635-5652.e29. doi:10.1016/j.cell.2021.09.018

26. Schene IF, Joore IP, Oka R, Mokry M, van Vugt AHM, van Boxtel R, et al. Prime editing for functional repair in patient-derived disease models. Nat Commun. 2020 Oct 23;11(1):5352. doi:10.1038/s41467-020-19136-7

27. Zhao Z, Shang P, Mohanraju P, Geijsen N. Prime editing: advances and therapeutic applications. Trends in Biotechnology. 2023 Aug 1;41(8):1000–12. doi:10.1016/j.tibtech.2023.03.004

28. Gvozdenovic A, Corn JE. Fueling next-generation genome editing with DNA repair. Current Opinion in Biomedical Engineering. 2023 Dec 1;28:100506. doi:10.1016/j.cobme.2023.100506

29. Ferreira da Silva J, Oliveira GP, Arasa-Verge EA, Kagiou C, Moretton A, Timelthaler G, et al. Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. Nat Commun. 2022 Feb 9;13(1):760. doi:10.1038/s41467-022-28442-1

30. Park JC, Kim YJ, Han JH, Kim D, Park MJ, Kim J, et al. MutSα and MutSβ as size-dependent cellular determinants for prime editing in human embryonic stem cells. Molecular Therapy Nucleic Acids. 2023 Jun 13;32:914–22. doi:10.1016/j.omtn.2023.05.015

31. Doman JL, Pandey S, Neugebauer ME, An M, Davis JR, Randolph PB, et al. Phage-assisted evolution and protein engineering yield compact, efficient prime editors. Cell. 2023 Aug 31;186(18):3983-4002.e26. doi:10.1016/j.cell.2023.07.039

32. Wolin SL, Cedervall T. The La Protein. Annual Review of Biochemistry. 2002 Jul 1;71(Volume 71, 2002):375–403. doi:10.1146/annurev.biochem.71.090501.150003

33. Yan J, Oyler-Castrillo P, Ravisankar P, Ward CC, Levesque S, Jing Y, et al. Improving prime editing with an endogenous small RNA-binding protein. Nature. 2024 Apr;628(8008):639–47. doi:10.1038/s41586-024-07259-6

34. Nelson JW, Randolph PB, Shen SP, Everette KA, Chen PJ, Anzalone AV, et al. Engineered pegRNAs improve prime editing efficiency. Nat Biotechnol. 2022 Mar;40(3):402–10. doi:10.1038/s41587-021-01039-7

35. Yan J, Oyler-Castrillo P, Ravisankar P, Ward CC, Levesque S, Jing Y, et al. Improving prime editing with an endogenous small RNA-binding protein. Nature. 2024 Apr;628(8008):639–47. doi:10.1038/s41586-024-07259-6

36. Petri K, Zhang W, Ma J, Schmidts A, Lee H, Horng JE, et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. Nat Biotechnol. 2022 Feb;40(2):189–93. doi:10.1038/s41587-021-00901-y

37. Yu G, Kim HK, Park J, Kwak H, Cheong Y, Kim D, et al. Prediction of efficiencies for diverse prime editing systems in multiple cell types. Cell. 2023 May 11;186(10):2256-2272.e23. doi:10.1016/j.cell.2023.03.034

38. Krausz SL, Simon DA, Bartos Z, Biczók Z, Varga É, Huszár K, et al. ProPE expands the prime editing window and enhances gene editing efficiency where prime editing is inefficient. Nat Catal. 2025 Oct;8(10):1100–16. doi:10.1038/s41929-025-01406-6

39. Jang H, Jo DH, Cho CS, Shin JH, Seo JH, Yu G, et al. Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. Nat Biomed Eng. 2022 Feb;6(2):181–94. doi:10.1038/s41551-021-00788-9

40. Ponnienselvan K, Liu P, Nyalile T, Oikemus S, Maitland SA, Lawson ND, et al. Reducing the inherent auto-inhibitory interaction within the pegRNA enhances prime editing efficiency. Nucleic Acids Research. 2023 Jul 21;51(13):6966–80. doi:10.1093/nar/gkad456

41. Lin Q, Jin S, Zong Y, Yu H, Zhu Z, Liu G, et al. High-efficiency prime editing with optimized, paired pegRNAs in plants. Nat Biotechnol. 2021 Aug;39(8):923–7. doi:10.1038/s41587-021-00868-w

42. Fei J, Zhao D, Pang C, Li J, Li S, Qiao W, et al. Mismatch prime editing gRNA increased efficiency and reduced indels. Nat Commun. 2025 Jan 2;16(1):139. doi:10.1038/s41467-024-55578-z

43. Mathis N, Allam A, Kissling L, Marquart KF, Schmidheini L, Solari C, et al. Predicting prime editing efficiency and product purity by deep learning. Nat Biotechnol. 2023 Aug;41(8):1151–9. doi:10.1038/s41587-022-01613-7

44. Mathis N, Allam A, Tálas A, Kissling L, Benvenuto E, Schmidheini L, et al. Machine learning prediction of prime editing efficiency across diverse chromatin contexts. Nat Biotechnol. 2024 Jun 21;1–8. doi:10.1038/s41587-024-02268-2

45. Kim HK, Yu G, Park J, Min S, Lee S, Yoon S, et al. Predicting the efficiency of prime editing guide RNAs in human cells. Nat Biotechnol. 2021 Feb;39(2):198–206. doi:10.1038/s41587-020-0677-y

46. Zhang W, Petri K, Ma J, Lee H, Tsai CL, Joung JK, et al. Enhancing CRISPR prime editing by reducing misfolded pegRNA interactions. Del Bene F, Stainier DY, editors. eLife. 2024 Jun 7;12:RP90948. doi:10.7554/eLife.90948

47. Liu B, Dong X, Cheng H, Zheng C, Chen Z, Rodríguez TC, et al. A split prime editor with untethered reverse transcriptase and circular RNA template. Nat Biotechnol. 2022 Sep;40(9):1388–93. doi:10.1038/s41587-022-01255-9

48. Simon DA, Tálas A, Kulcsár PI, Biczók Z, Krausz SL, Várady G, et al. PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells. Lapinaite A, Stainier DY, Hamilton JR, editors. eLife. 2022 Feb 23;11:e69504. doi:10.7554/eLife.69504

49. Tálas A, Simon DA, Kulcsár PI, Varga É, Krausz SL, Welker E. BEAR reveals that increased fidelity variants can successfully reduce the mismatch tolerance of adenine but not cytosine base editors. Nat Commun. 2021 Nov 3;12(1):6353. doi:10.1038/s41467-021-26461-y

50. Kulcsár PI, Tálas A, Huszár K, Ligeti Z, Tóth E, Weinhardt N, et al. Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. Genome Biology. 2017 Oct 6;18(1):190. doi:10.1186/s13059-017-1318-8

51. Chen Y, Tristan CA, Chen L, Jovanovic VM, Malley C, Chu PH, et al. A versatile polypharmacology platform promotes cytoprotection and viability of human pluripotent and differentiated cells. Nat Methods. 2021 May;18(5):528–41. doi:10.1038/s41592-021-01126-2

52. Vriend LEM, Jasin M, Krawczyk PM. Chapter Nine - Assaying Break and Nick-Induced Homologous Recombination in Mammalian Cells Using the DR-GFP Reporter and Cas9 Nucleases. In: Doudna JA, Sontheimer EJ, editors. Methods in Enzymology [Internet]. Academic Press; 2014 [cited 2024 Nov 12]. p. 175–91. (The Use of CRISPR/Cas9, ZFNs, and TALENs in Generating Site-Specific Genome Alterations). Available from: https://www.sciencedirect.com/science/article/pii/B978012801185000009X doi:10.1016/B978-0-12-801185-0.00009-X

53. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013 Sep;31(9):827–32. doi:10.1038/nbt.2647

54. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Molecular Therapy - Nucleic Acids. 2015 Jan 1;4:e264. doi:10.1038/mtna.2015.37

55. Shuto Y, Nakagawa R, Zhu S, Hoki M, Omura SN, Hirano H, et al. Structural basis for pegRNA-guided reverse transcription by a prime editor. Nature. 2024 Jul;631(8019):224–31. doi:10.1038/s41586-024-07497-8

56. Chen PJ, Liu DR. Prime editing for precise and highly versatile genome manipulation. Nat Rev Genet. 2023 Mar;24(3):161–77. doi:10.1038/s41576-022-00541-1

57. Li M, Zhong A, Wu Y, Sidharta M, Beaury M, Zhao X, et al. Transient inhibition of p53 enhances prime editing and cytosine base-editing efficiencies in human pluripotent stem cells. Nat Commun. 2022 Oct 27;13(1):6354. doi:10.1038/s41467-022-34045-7

58. Kim DY, Moon SB, Ko JH, Kim YS, Kim D. Unbiased investigation of specificities of prime editing systems in human cells. Nucleic Acids Res. 2020 Oct 9;48(18):10576–89. doi:10.1093/nar/gkaa764

59. Huszár K, Welker Z, Györgypál Z, Tóth E, Ligeti Z, Kulcsár PI, et al. Position-dependent sequence motif preferences of SpCas9 are largely determined by scaffold-complementary spacer motifs. Nucleic Acids Res. 2023 Jun 23;51(11):5847–63. doi:10.1093/nar/gkad323

60. Li X, Zhou L, Gao BQ, Li G, Wang X, Wang Y, et al. Highly efficient prime editing by introducing same-sense mutations in pegRNA or stabilizing its structure. Nat Commun. 2022 Mar 29;13(1):1669. doi:10.1038/s41467-022-29339-9

61. Liu Y, Yang G, Huang S, Li X, Wang X, Li G, et al. Enhancing prime editing by Csy4-mediated processing of pegRNA. Cell Res. 2021 Oct;31(10):1134–6. doi:10.1038/s41422-021-00520-x

62. Zhang G, Liu Y, Huang S, Qu S, Cheng D, Yao Y, et al. Enhancement of prime editing via xrRNA motif-joined pegRNA. Nat Commun. 2022 Apr 6;13(1):1856. doi:10.1038/s41467-022-29507-x

9. Bibliography of the candidate's publications

Publications related to the thesis

Biczók Z, Krausz SL, Simon DA, Tóth E, Varga É, Annus T, et al. Disrupting pegRNA intramolecular complementarity via PBS and spacer sequence alterations can enhance prime editing efficiency. Nucleic Acids Res. 2026 Apr 24;54(7):gkag292. doi:10.1093/nar/gkag292

Krausz SL, Simon DA, Bartos Z, Biczók Z, Varga É, Huszár K, et al. ProPE expands the prime editing window and enhances gene editing efficiency where prime editing is inefficient. Nat Catal. 2025 Oct;8(10):1100–16. doi:10.1038/s41929-025-01406-6

Publications not related to the thesis

Simon DA, Tálas A, Kulcsár PI, Biczók Z, Krausz SL, Várady G, et al. PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells. Lapinaite A, Stainier DY, Hamilton JR, editors. eLife. 2022 Feb 23;11:e69504. doi:10.7554/eLife.69504

10. Acknowledgements

I would like to thank my supervisor, Ervin Welker, for his continuous help and support during my doctoral studies and my work in the Welker group.

I would also like to thank Sarah Laura Krausz, whose idea sparked this project and thus made this work possible. I am grateful to her and Dorottya Anna Simon for their help in analysing the NGS data, as well as for introducing me to the necessary basics.

I would also like to thank all my co-authors for their wonderful work on the NAR publication, especially Dorottya Anna Simon, Eszter Tóth, and Éva Varga, without whose support the revision process could not have been completed.

I would like to thank my former and current colleagues in the Welker group for the supportive environment and the tremendous help. I am especially grateful to Ildikó Szűcsné Pulinka, Judit Szűcs, Vivien Karl, Veronika Csonka, and Viktória Faragó for their excellent technical assistance. I would also like to thank Viktória Faragó for her enormous help in compiling the figures for the two publications.

I am also grateful to György Váradi and Edit Szabó for their assistance with operating the flow cytometers and conducting the FACS measurements.

Finally, I would like to thank my family, especially my mother, father, and sister, for their endless support throughout my doctoral work. I am also very grateful to all my friends, with whom I was able to share all my difficulties, and who always responded with understanding and support.

Appendix

Prime sequences used for next generation sequencing

amplicon	Tm	sequence	
mScarlet	68	fwd	TCGTCGGCAGCgtcAGATGTGTATAAGAGACAgcttccccgagggctt caagt
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTGACTTAAAGGGGACCAACACA
GFP	70	fwd	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCCTCGTGACCACCCTGAC
		rev	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGACTTAAAGGGGACCAACACAT
DNMT1	66	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACAACAGCTTCATGTCAGCC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACGTTAATGTTTCCTGATGGTCC
EMX1	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGCTCAGCCTGAGTGTTGA
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCGTGGGTTTGTGGTTGC
FANCF	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTGCTGACGTAGGTAGTGC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACACGGATAAAGACGCTGGG
HBB	68	fwd	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGGTTGGCCAATCTACTCCC
		rev	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCTTCTCTGTCTCCACATGCC
HEK3	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGTGGGCTGCCTAGAAAGG
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCAGCCAAACTTGTCAACC
HEK4	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAACCCAGGTAGCCAGAGAC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTTTCAACCCGAACGGAG
HEXA	69	fwd	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATACAGGTGTGGCGAGAGG
		rev	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCCTCCTTTGGTTAGCA
PRNP	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCAGTGGAACAAGCCGAGT
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTGGTTGGGGTAACGGTG
RUNX1	68	fwd	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGAACTGGCAGGCACCGAGG
		rev	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGTGAGGCTGAAACAGTGACC

RNF2	62	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACGTCTC ATATGCCCCTTGG
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACGTAGG ATTTTGGTGGGACA
CYP1A1	70	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGGTCC CCAAAGGCCTGAAGAAT
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGGGT GAAGGTGTAGAGGTCG
CYP1A2	72	fwd	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGTCC CCAAAGGCCTGAAAAGT
		rev	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGAGGG TGGAGGTGTAGAGGTCAG
CYP2B6	72	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAGACC AGGACCATGGAACTCAGC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAAGC TTCCCCAAGTACCAAGGC
RYR2	66	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTCCCT TAAAAACGTCAAGCTC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGCGT CCCAAGAGGTCAAT
KRT12	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGGAGG TGGCTCTCTAGGTAT
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCAAA GCGCCTCCAAAAGAG
GBA1	65	fwd	TCGTCGGCAGCgtcAGATGTGTATAAGAGACAgCCTTGAGAT GCCTGGATCTTC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGACAA AGTTACGCACCCAAT
LRRK2	66	fwd	TCGTCGGCAGCgtcAGATGTGTATAAGAGACAgTAAGGGAC AAAGTGAGCACAG
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACATC TGAGGTCAGTGGTTATC
gpr58	66	fwd	TCGTCGGCAGCgtcAGATGTGTATAAGAGACAgCTGCTTCAA AGCACACCAGA
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGCTCG ATGAAGGCTCTTGTC
CDKL5	71	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTCAGA AGGCCCAGGGACAAAG
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGTA CTGGGCTCCGCAATTT
CTNNB1	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCATTCT GACTTTCAGTAAGGCAATGA
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGCCA TGGAACCAGACAGAAA
CXCR4	72	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTTGGCG TCTGGATCCCTGCCC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTGAG GGCCTTGCGCTTCTGG
IL2RB	71	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGACCAG GACAGGAAGGAGGAA
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGGCA GGTACAAAGTGGGAGG

HBG	71	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTTCATCCCTAGCCAGCCGC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACGGCTGACAAAAGAAGTCCTGG
COL7A1-T1	72	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGGTGAGACCAGTGTGCG
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGCACACTCCGCCAG
COL7A1-T2	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGTTCTGTCTCCTGCAGAAGTGCC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACACTGTAGGAAGGGGAACAAG
CACNA1C-T1	72	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGAGGTGTCATGGGGGATCTTT
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATGCCGCCACACACGACGAA
CACNA1C-T2	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAACTCACACCACCACCAACA
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACATCCTCCAAAGAGCTGCATCC
HEK4-OT1	68	fwd	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCATGGCTTCTGAGACTCA
		rev	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCTCCCTTGCACTCCCTGTCTTT
HEK4-OT2	67	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTGGCAATGGAGGCATTGG
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAAGAGGCTGCCCATGAGAG
HEK4-OT3	69	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCTGAGGCTCGAATCCTG
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGTGGCCTCCATATCCCTG
HEK4-OT4	68	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTCCACCAGAACTCAGCCC
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTCGGTTCCTCCACAACAC
HBB-OT	67	fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCCACGTTCACTTTGCCCCA
		rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTATCTTAAACCAACCTGCTCACTGG

Spacer, PBS and RTT sequences used for PE experiments

Spacer mismatches

Target name	**Intended edit**	**RTT length**	**PBS length**	**RTT-PBS sequence (5'-3')**	**2nd nicking spacer (5'-3')**
mScarlet target 1	33-34 AC to GT	31	10	acuuaccugcgucacggucacggcgccgccguccucgaagu	ggaccaacacauauuuaaag
		31	13	acuuaccugcgucacggucacggcgccgccguccucgaaguuca	
		31	17	acuuaccugcgucacggucacggcgccgccguccucgaaguucaucac	
		31	20	acuuaccugcgucacggucacggcgccgccguccucgaaguucaucacgcg	
mScarlet target 2	26-27 AC to GT	38	13	acuuaccugcgucacggucacggcgccgccguccucgaaguucaucacgcg	
		38	17	acuuaccugcgucacggucacggcgccgccguccucgaaguucaucacgcgcucc	
		38	20	acuuaccugcgucacggucacggcgccgccguccucgaaguucaucacgcgcucc cac	

Target name	**Mismatch position**	**Original base**	**Mismatching base**	**Mismatching base pair**	**Sequence (5'-3')**
mScarlet target 1					gugaugaacuucgaggacgg
	8	G	A	C-A	gugaugaacuucAaggacgg
	8	G	U	C-U	gugaugaacuucUaggacgg
	10	U	G	A-G	gugaugaacuGcgaggacgg
	10	U	C	A-C	gugaugaacuCcgaggacgg
	12	C	U	G-U	gugaugaaUuucgaggacgg

	12	C	A	G-A	gugaugaaAuucgaggacgg
	13	A	G	U-G	gugaugaGcuucgaggacgg
	13	A	C	U-C	gugaugaCcuucgaggacgg
	14	A	G	U-G	gugaugGacuucgaggacgg
	14	A	C	U-C	gugaugCacuucgaggacgg
	15	G	A	C-A	gugauAaacuucgaggacgg
	15	G	U	C-U	gugauUaacuucgaggacgg
	16	U	C	A-C	gugaCgaacuucgaggacgg
	16	U	G	A-G	gugaGgaacuucgaggacgg
	18	G	A	C-A	guAaugaacuucgaggacgg
	18	G	U	C-U	guUaugaacuucgaggacgg
	20	G	A	C-A	Augaugaacuucgaggacgg
	20	G	U	C-A	Uugaugaacuucgaggacgg
	14+16	A+U	G+C	U-G + A-C	gugaCgGacuucgaggacgg
	17+19	A+U	G+C	U-G + A-C	gCgGugaacuucgaggacgg
	17+19	A+U	C+G	U-C+A-G	gGgCugaacuucgaggacgg
	12+20	C+G	U+A	G-U + C-A	AugaugaaUuucgaggacgg
	12+20	C+G	A+U	G-A+C-U	UugaugaaAuucgaggacgg
	12+17+20	C+A+G	U+C+A	G-U+U-C+C-A	AugCugaaUuucgaggacgg
	12+17+20	C+A+G	A+G+U	G-A+U-G+C-U	UugGugaaAuucgaggacgg
mScarlet target 2					ggagcgcgugaugaacuucg
	8	G	A	C-A	ggagcgcgugauAaacuucg
	8	G	U	C-U	ggagcgcgugauUaacuucg
	10	A	G	U-G	ggagcgcgugGugaacuucg
	10	A	C	U-C	ggagcgcgugCugaacuucg

	12	U	C	A-C	ggagcgcgCgaugaacuucg
	12	U	G	A-G	ggagcgcgGgaugaacuucg
	13	G	U	C-U	ggagcgcUugaugaacuucg
	13	G	A	C-A	ggagcgcAugaugaacuucg
	14	C	U	G-U	ggagcgUgugaugaacuucg
	14	C	A	G-A	ggagcgAgugaugaacuucg
	15	G	U	C-U	ggagcUcgugaugaacuucg
	15	G	A	C-A	ggagcAcgugaugaacuucg
	16	C	U	G-U	ggagUgcgugaugaacuucg
	16	C	A	G-A	ggagAgcgugaugaacuucg
	18	A	G	U-G	ggGgcgcgugaugaacuucg
	18	A	C	U-C	ggCgcgcgugaugaacuucg
	20	G	A	C-A	Agagcgcgugaugaacuucg
	20	G	U	C-U	Ugagcgcgugaugaacuucg
	17+19	G+G	A+A	C-A+C-A	gAaAcgcgugaugaacuucg
	17+19	G+G	U+U	C-U+C-U	gUaUcgcgugaugaacuucg
	12+20	U+G	C+A	A-C+C-A	AgagcgcgCgaugaacuucg
	12+17+20	U+G+G	C+U+A	A-C+C-U+C-A	AgaUcgcgCgaugaacuucg
	12+17+20	U+G+G	G+A+U	A-G+C-A+C-U	UgaAcgcgGgaugaacuucg

PBS mismatches

Target name	PBS length	Mismatch position	Original base	Mismatching base	Mismatching base pair	PBS	RTT	Spacer	2nd nick spacer
mScarlet target 1	10					uccucgaagu	acuuaccugcgucacggucacggcgccgccg	gugaugaacuucgaggacgg	ggaccaacacacauauuuaaag
		5	C	U	G-U	uccuUgaagu			
		6	G	A	C-A	uccucAaagu			
		7	A	G	U-G	uccucgGagu			
		8	A	G	U-G	uccucgaGgu			
		9	G	A	C-A	uccucgaaAu			
		10	U	C	A-C	uccucgaagC			
	13					uccucgaaguuca			
		5	C	U	G-U	uccuUgaaguuca			
		6	G	A	C-A	uccucCaaguuca			
		7	A	G	U-G	uccucgGaguuca			
		8	A	G	U-G	uccucgaGguuca			
		9	G	A	C-A	uccucgaaAuuca			
		10	U	C	A-C	uccucgaagCuca			
		11	U	C	A-C	uccucgaaguCca			
		12	C	U	A-C	uccucgaaguuUa			
		13	A	G	G-U	uccucgaaguucG			
	20					uccucgaaguucaucacgcg			
		5+7	C+A	U+G	G-U+U-G	uccuUgGaguucaucacgcg			
		6+8	G+A	A+G	C-A+U-G	uccucAaGguucaucacgcg			
		7+9	A+G	G+A	U-G+C-A	uccucgGaAuucaucacgcg			

		8+18	A+G	G+A	U-G+C-A	uccucgaGguucaucacAcg			
		5+7+9	C+A+G	U+G+A	G-U+U-G+C-A	uccuUgGaAuucaucacgcg			
		6+8+10	G+A+U	A+G+C	C-A+U-G+A-C	uccucAaGgCucaucacgcg			
		8+12+16	A+C+A	G+U+G	U-G+G-U+U-G	uccucgaGguuUaucGcgcg			
mScarlet target 2	13					aguucaucacgcg	acuuaccugcgucacggucacggcgccgccguccucga	ggaggcgguggaaguaacuucg	
		5	C	U	G-U	aguuUaucacgcg			
		6	A	G	U-G	aguucGucacgcg			
		7	U	C	A-C	aguucaCcacgcg			
		9	A	G	U-G	aguucaucGcgcg			
		10	C	U	G-U	aguucaucaUgcg			
		11	G	A	C-A	aguucaucacAcg			
		12	C	U	G-U	aguucaucacgUg			
		13	G	A	C-A	aguucaucacgcA			
	20					aguucaucacgcgcucccac			
		5+7	C+U	U+C	G-U+A-C	aguuUaCcacgcgcucccac			
		6+8	A+C	G+U	U-G+G-U	aguucGuUacgcgcucccac			
		7+9	U+A	C+G	A-C+U-G	aguucaCcGcgcgcucccac			
		8+16	C+C	U+U	G-U+G-U	aguucauUacgcgcuUccac			
		5+7+9	C+U+A	U+C+G	G-U+A-C+U-G	aguuUaCcGcgcgcucccac			

		6+8+10	A+C+C	G+U+U	U-G+G-U+G-U	aguucGuUaUgcgcucccac			
		8+12+16	C+C+C	U+U+U	G-U+G-U+G-U	aguucauUacgUgcuUccac			

PBS deletions

Target name	**PBS length**	**Deletion position**	**PBS sequence (5'-3')**	**RTT length**	**RTT sequence (5'-3')**	**spacer sequence (5'-3')**	**2nd nick spacer (5'-3')**
mScarlet target 1	7		uccucga	31	acuuaccugcgucacggucacggcgccgccg	gugaugaacuucgaggacgg	ggaccaacacauauuuaaag
	8		uccucgaa				
	9		uccucgaag				
	10		uccucgaagu				
	11		uccucgaaguu				
	12		uccucgaaguuc				
	13		uccucgaaguuc				
	14		uccucgaaguucau				
	15		uccucgaaguucauc				
	17		uccucgaaguucaucac				
		6	uccuc-aaguucaucac				
		7	uccucg-aguucaucac				
		8	uccucga-guucaucac				
		9	uccucgaa-uucaucac				
		10	uccucgaag-ucaucac				

		11	uccucgaagu-caucac				
		12	uccucgaaguu-aucac				
		13	uccucgaaguuc-ucac				
		14	uccucgaaguuca-cac				
		15	uccucgaaguucau-ac				
		16	uccucgaaguucuac-c				
		17	uccucgaaguucauca-				
		6+7+8	uccuc---guucaucac				
		7+8+9	uccucg---uucaucac				
		8+9+10	uccucga---ucaucac				
		9+10+11	uccucgaa---caucac				
		10+11+12	uccucgaag---aucac				
		11+12+13	uccucgaagu---ucac				
		12+13+14	uccucgaaguu---cac				
		13+14+15	uccucgaaguuc---ac				
		14+15+16	uccucgaaguuca---c				
		15+16+17	uccucgaaguucau---				
	20		uccucgaaguucaucacgcg				
		6	uccuc-aaguucaucacgcg				
		7	uccucg-aguucaucacgcg				
		8	uccucga-guucaucacgcg				
		9	uccucgaa-uucaucacgcg				
		10	uccucgaag-ucaucacgcg				
		11	uccucgaagu-caucacgcg				
		12	uccucgaaguu-aucacgcg				
		13	uccucgaaguuc-ucacgcg				
		14	uccucgaaguuca-cacgcg				

		15	uccucgaaguucau-acgcg				
		16	uccucgaaguucuac-cgcg				
		17	uccucgaaguucauca-gcg				
		6+7	uccuc--aguucaucacgcg				
		7+8	uccucg--guucaucacgcg				
		9+10	uccucgaa--ucaucacgcg				
		10+11	uccucgaag--caucacgcg				
		11+12	uccucgaagu--aucacgcg				
		12+13	uccucgaaguu--ucacgcg				
		13+14	uccucgaaguuc--cacgcg				
		14+15	uccucgaaguuca--acgcg				
		15+16	uccucgaaguucau--cgcg				
		16+17	uccucgaaguucauc--gcg				
		6+7+8	uccuc---guucaucacgcg				
		7+8+9	uccucg---uucaucacgcg				
		8+9+10	uccucga---ucaucacgcg				
		9+10+11	uccucgaa---caucacgcg				
		10+11+12	uccucgaag---aucacgcg				
		11+12+13	uccucgaagu---ucacgcg				
		12+13+14	uccucgaaguu---cacgcg				
		13+14+15	uccucgaaguuc---acgcg				
		14+15+16	uccucgaaguuca---cgcg				
		15+16+17	uccucgaaguucau---gcg				
		16+17+18	uccucgaaguucauc---cg				
		6+7+8+9	uccuc----uucaucacgcg				
		7+8+9+10	uccucg----ucaucacgcg				
		8+9+10+11	uccucga----caucacgcg				

		9+10+11+12	uccucgaa----aucacgcg				
		10+11+12+13	uccucgaag----ucacgcg				
		11+12+13+14	uccucgaagu----cacgcg				
		12+13+14+15	uccucgaaguu----acgcg				
		13+14+15+16	uccucgaaguuc----cgcg				
		14+15+16+17	uccucgaaguuca----gcg				
mScarlet target 2	7		aguucau	38	acuuaccugcgucacggucacggccgccgccguccucga	ggagcgcgugaugaacuucg	
	8		aguucauc				
	9		aguucauca				
	10		aguucaucac				
	11		aguucaucacg				
	12		aguucaucacgc				
	13		aguucaucacgcg				
	14		aguucaucacgcgc				
	15		aguucaucacgcgcu				
	17		aguucaucacgcgcucc				
		6	aguuc-ucacgcgcucc				
		7	aguuca-cacgcgcucc				
		8	aguucau-acgcgcucc				
		9	aguucauc-cgcgcucc				
		10	aguucauca-gcgcucc				
		11	aguucaucac-cgcucc				
		12	aguucaucacg-gcucc				
		13	aguucaucacgc-cucc				
		14	aguucaucacgcg-ucc				
		15	aguucaucacgcgc-cc				
		16	aguucaucacgcgcu-c				

		17	aguucaucacgcgcuc-				
		6+7+8	aguuc---acgcgcucc				
		7+8+9	aguuca---cgcgcucc				
		8+9+10	aguucau---gcgcucc				
		9+10+11	aguucauc---cgcucc				
		10+11+12	aguucauca---gcucc				
		11+12+13	aguucaucac---cucc				
		12+13+14	aguucaucacg---ucc				
		13+14+15	aguucaucacgc---cc				
		14+15+16	aguucaucacgcg---c				
		15+16+17	aguucaucacgcgc---				
	20	-	aguucaucacgcgcucccac				
		6	aguuc-ucacgcgcucccac				
		7	aguuca-cacgcgcucccac				
		8	aguucau-acgcgcucccac				
		9	aguucauc-cgcgcucccac				
		10	aguucauca-gcgcucccac				
		11	aguucaucac-cgcucccac				
		12	aguucaucacg-gcucccac				
		13	aguucaucacgc-cucccac				
		14	aguucaucacgcg-ucccac				
		15	aguucaucacgcgc-cccac				
		16	aguucaucacgcgcu-ccac				
		17	aguucaucacgcgcuc-cac				
		6+7	aguuc--cacgcgcucccac				
		7+8	aguuca--acgcgcucccac				
		8+9	aguucau--cgcgcucccac				

		9+10	aguucauc--gcgcucccac				
		10+11	aguucauca--cgcucccac				
		11+12	aguucaucac--gcucccac				
		12+13	aguucaucacg--cucccac				
		13+14	aguucaucacgc--ucccac				
		14+15	aguucaucacgcg--cccac				
		15+16	aguucaucacgcgc--ccac				
		16+17	aguucaucacgcgcu--cac				
		6+7+8	aguuc---acgcgcucccac				
		7+8+9	aguuca---cgcgcucccac				
		8+9+10	aguucau---gcgcucccac				
		9+10+11	aguucauc---cgcucccac				
		10+11+12	aguucauca---gcucccac				
		11+12+13	aguucaucac---cucccac				
		12+13+14	aguucaucacg---ucccac				
		13+14+15	aguucaucacgc---cccac				
		14+15+16	aguucaucacgcg---ccac				
		15+16+17	aguucaucacgcgc---cac				
		6+7+8+9	aguuc----cgcgcucccac				
		7+8+9+10	aguuca----gcgcucccac				
		8+9+10+11	aguucau----cgcucccac				
		9+10+11+12	aguucauc----gcucccac				
		10+11+12+13	aguucauca----cucccac				
		11+12+13+14	aguucaucac----ucccac				
		12+13+14+15	aguucaucacg----cccac				
		13+14+15+16	aguucaucacgc----ccac				
		14+15+16+17	aguucaucacgcg----cac				

Combined modfications

Target name	PBS length	RTT length	Deletion position	Mismatch position	Original base	Mismatching base	Mismatching base pair	Spacer sequence (5'-3')	PBS sequence (5'-3')	RTT sequence (5'-3')	2nd nick sequence (5'-3')
mScarlet target 1	10	31						gugaugaacuucgaggacgg	uccucgaagu	acuaccugcgucacggucacggcgccgccg	ggaccaacacauauuuaaag
	12							gugaugaacuucgaggacgg	uccucgaaguuc		
	20							gugaugaacuucgaggacgg	uccucgaaguucaucacgcg		
			13					gugaugaacuucgaggacgg	uccucgaaguuc-ucacgcg		
			13	12	C	U	G-U	gugaugaaUuucgaggacgg	uccucgaaguuc-ucacgcg		
			13	13	A	C	U-C	gugaugaCcuucgaggacgg	uccucgaaguuc-ucacgcg		
			13	15	G	U	C-U	gugauUaacuucgaggacgg	uccucgaaguuc-ucacgcg		
			13	18	G	U	C-U	guUaugaacuucgaggacgg	uccucgaaguuc-ucacgcg		
			13	20	G	U	C-A	Uugaugaacuucgaggacgg	uccucgaaguuc-ucacgcg		
			13+14					gugaugaacuucgaggacgg	uccucgaaguuc--cacgcg		

			13+14	12	C	U	G-U	gugaugaaUuucg aggacgg	uccucgaaguuc-- cacgcg		
			13+14	13	A	C	U-C	gugaugaCcuucg aggacgg	uccucgaaguuc-- cacgcg		
mScarlet target 2	12	38						ggagcgcgugauga acuucg	aguucaucacgc	acuuaccugcgucacggucacggcgccgccguccucga	
	13							ggagcgcgugauga acuucg	aguucaucacgcg		
	20							ggagcgcgugauga acuucg	aguucaucacgcgc ucccac		
			13					ggagcgcgugauga acuucg	aguucaucacgc- cucccac		
			13	13	G	U	C-U	ggagcgcUugaug aacuucg	aguucaucacgc- cucccac		
			13	14	C	U	G-U	ggagcgUgugaug aacuucg	aguucaucacgc- cucccac		
			13	15	G	U	C-U	ggagcUcgugaug aacuucg	aguucaucacgc- cucccac		
			13	18	A	C	U-C	ggCgcgcgugaug aacuucg	aguucaucacgc- cucccac		
			13	20	G	U	C-U	Ugagcgcgugaug aacuucg	aguucaucacgc- cucccac		
			13+14					ggagcgcgugauga acuucg	aguucaucacgc-- ucccac		
			13+14	12	U	C	A-C	ggagcgcgCgaug aacuucg	aguucaucacgc-- ucccac		
			13+14	18	A	C	U-C	ggCgcgcgugaug aacuucg	aguucaucacgc-- ucccac		

Target name	Intended edit	PBS length	RTT length	Deletion position	Mismatch position	Original base	Mismatching base	Mismatching base pair	Spacer sequence (5'-3')	PBS sequence (5'-3')	RTT sequence (5'-3')	2nd nick sequence (5'-3')
DNMT	3-5 AGG del	13	12						gauuccuggugccagaaaca	uucuggcaccagg	ucccgucaccgu	gcccuucagcuaaaauaaagg
		20		13	13	G	A	A-C	gauuccuAgugccagaaaca	uucuggcaccag-aaucccc		
EMX1	5 G to T	15	13						gaguccgagcagaagaagaa	uucuucugccucggac	augggagcacuuc	gacaucgaugucucccccau
		20		13	13	A	G	U-G	gaguccgGgcagaagaagaa	uucuucugccucg-acucagg		
FANCF	5 G to T	8	17						ggaaucccuucugcagcacc	gcugcaga	ggaaaagcgaucaaggu	gggguccaggugcugacgu
		20		13	18	A	G	U-G	ggGaucccuucugcagcacc	gcugcagaaggg-uuccaug		
HBB	4 A to T	13	14						gcauggugcaccugacuccug	gagucaggugcac	agacuucuccacag	gccuugauaccaaccugccca
		20		13	13	C	U	G-U	gcauggugUaccugacuccug	gagucaggugca-cauggug		
HEK4	2 G to T	8	10						ggcacugcggcuggaggugg	ccuccagc	uuaaccccaa	gucccuuccuuccacccagcc
		20		13	15	U	G	A-G	ggcacGgcggcuggaggugg	ccuccagccgca-ugccacc		

HEXA	1 TATC	12	18						guaccugaaccguauauccua	gauauacgguuc	agucagggccauaggaua	gcuuucaccuucaaaugcca
		20		13	13	A	G	U-G	guaccugaGccguauauccua	gauauacgguuc-gguacca		
PRNP	6 G to T	12	12						gcaguggugggggccuugg	aggccccccacc	auguagacgcca	gcauguuuucacgauaguaa
		20		13	15	G	U	C-U	gcaguUgugggggccuugg	aggccccccacc-cugcccc		
RUNX1	6 G to C	15	15						gcauuuucaggaggaagcga	cuuccuccugaaaau	ugucugaaggcaucg	augaagcacuggggguacga
		20		13	13	C	U	G-U	gcauuuuUaggaggaagcga	cuuccuccugaaa-ugcacc		
HEK3	1 CTT ins	13	13						ggcccagacugagcacguga	cgugcucagucug	ucugccaucaaag	gucaaccaguauccggugc
		20		13	15	A	G	U-G	ggcccGgacugagcacguga	cgugcucagucug-gcccca		
RNF2	1 C to A	15	14						gucaucuuaguca uuaccug	guaaugacuaagaug	aacgaacaccucau	gucaaccauuaagcaaaacau
		20		13	13	U	C	A-C	gucaucuCaguca uuaccug	guaaugacuaag-ugacugc		

SPELL pegRNA

Target name	**Intended edit**	**PBS length**	**Deletion position**	**PBS sequence (5'-3')**	**RTT length**	**RTT sequence (5'-3')**	**spacer sequence (5'-3')**	**2nd nick sequence (5'-3')**
CYP1A1-T1	7 A to G	13		CAGUGCCAGGUGC	16	GCUGGCUCACCCUUGA	ACCCGCACCUGGCACUGUCA	GGGCCUGCCGGAUGGUGUCC
		20	13	CAGUGCCAGGUG-GGGUUCU				
CYP1A2-T4	8 A to G	13		CACCAUCCGGCAG	38	AUUGGCUCCAUGCCCGUGCUGGUGCUGAGCCGCCUGGA	GGGCCUGCCGGAUGGUGUCC	ACCCGCACCUGGCACUGUCA
		20	13	CACCAUCCGGCA-GCCCUGG				
CYP2B6-T2	9 A to G	13		AGGGCCCCGCCCU	19	CCCAUGGCCGCCUCCCACC	GCAGAGGGCGGGGCCCUGGU	GCCUACUCAAAUCCUUUCUG
		20	13	AGGGCCCCGCCC-CUGCCCC				
CYP2B6-T3	10 A to G	13		GGGCCCCGCCCUC	30	CACCCUAACUCCCAUGACCGCCUCCCACCA	GGCAGAGGGCGGGGCCCUGG	GCCUACUCAAAUCCUUUCUG
		20	13	GGGCCCCGCCCU-UGCCCCU				
CYP2B6-T4	11 A to G	13		CCCCGCCCUCUGC	23	CCCAUGGCCGCCUCCCACCAGGG	AGGGGCAGAGGGCGGGGCCC	
		20	13	CCCCGCCCUCUG-CCCUUUU				
RYR2-T2	12 A to G	14		GUGGAUAGUGUCAU	20	GUCAUGAUCGUGUUCCCCAU	AAGAUGACACUAUCCACAUG	GUUUAUUACAUAUCUCUCCG

		20	13	GUGGAUAGU GUC-UCUUCCU				
KRT12-T2	13 A to G	13		UGGAUAAGG UGCG	25	CUUAAUGAUACAUUAGCUU CCUACC	AGCUCGCACC UUAUCCAGGU	AUCCACUGAU UGAAGACCUC
		20	13	UGGAUAAGG UGC-AGCUCUA				
GBA1	7 A toG	13		GUACCAUGU GGUC	18	CUUACCCUAGAGCCUCCU	AGCCGACCAC AUGGUACAGG	ACCCUUACCUA CACUCUCUG
		20	13	GUACCAUGU GGU-GGCUGGA				
LRRK2	5 G to A	14		UCAGCAAUC UUUGC	13	AGCAAUGCUGUAG	AUUGCAAAGA UUGCUGACUA	GACAGACCUG AUCACCUACC
		20	13	UCAGCAAUC UUU-CAAUGAU				
CDKL5	1 G to A	12		UCCUCUAGG AGU	20	AUAUUGACACAAUUCCCCA A	GAGGGACUCC UAGAGGACUG	GCAGAACCGCC ACUCAUUCA
		20	13	UCCUCUAGG AGUCCUCCU A				
CTNNB1	6 C to A	12		UCCAGGUAA GAC	23	UGGCACCAGAAUGGAUUAC AGAG	GCAACAGUCU UACCUGGACU C	GCCACUCAUAC AGGACUUGGG
		20	13	UCCAGGUAA GACGUUGCU G				
CXCR4	5 G to C	12		UGACUUGUG GGU	18	UGACCGCUUCUACGCCAA		

		20	13	UGACUUGUGGGUGUUGUGU			GCAACCACCCACAAGUCAUUG	GCAUCUUUGCCAACGUCAGUG
IL2RB	5 G to C	12		CUUUGAAAGACA	17	UCCCAAGCCUCCGACUA	GCCAGGUGUCUUUCAAAGUAG	GCUCCCUCCAAGUUGUCCACG
		20	13	CUUUGAAAGACACUGGAGU				
HBG	6 G to A	12		GCCUUGACAAGG	24	UGGCCAGCCUUGCCUUGAUCAAUA	GUUUGCCUUGUCAAGGCUAU	GCCUGGCUAAACUCCACCCAU
		20	13	GCCUUGACAAGGAAACUUG				
COL7A1-T1	5-7 ACC to tAA	12		AGCAAUUCUGCC	30	AGGCUGUGAGCGGGACAGCUCGGUAAAGUG	AGGCUGGCAGAAUUGCUCAC	GUGUGCGGCUGCGGGGUCUC
		20	13	AGCAAUUCUGCCGCCUCUG				
COL7A1-T2	14-15 AC to tG	12		AUAGUGGGCGUA	30	CCUGAGGCAUCUGUCUGACAGACGCCAGGU	GCCCACUACGCCCACUAUACC	GCAAGCUCCCAGCGGGUGACA
		20	13	AUAGUGGGCGUAUGGGAAG				
CACNA1C-T1	3-5 ACG to tAA	12		CUGGCAGGCCCA	30	ACAGGGCCAGCAGGGCCUUGUUUGCUUAGU	GUCUCCUGGGCCUGCCAGACA	GGCCCAGGCUGUACAUCUUC
		20	13	CUGGCAGGCCCAGAGAGGG				
	9 C to A	12		AGCCAACAGGAG	30	CAGGUUGCUCAAGGAGUUCCAUUACCUGAG	GCCUCCCUCCUGUUGGCUCUC	

CACNA1C-T2		20	13	AGCCAACAGGAGGAGGUCA				GUGAUGAAGAGGAAGAGGAGA

Off-target analysis

Target name	**Target sequence (5'-3')**	**Intended edit**	**PBS length**	**Deletion position**	**PBS sequence (5'-3')**	**RTT length**	**RTT sequence (5'-3')**	**spacer sequence (5'-3')**	**2nd nick sequence (5'-3')**
HEK4	CCACCTCCAGCCGCAGTGCC	2 G to T	8		ccuccagc	10	uuaaccccaa	ggcacugcggcuggaggugg	gucccuuccuuccaccagcc
			20	13	ccuccagccgcaugccacc				
HEK4-OT1	CCTCCTCCGGCCGCAGTGCA		8		ccuccagc				
			20	13	ccuccagccgcaugccacc				
HEK4-OT2	CCCCCTCCAGCCGCAGAGCC		8		ccuccagc				
			20	13	ccuccagccgcaugccacc				
HEK4-OT3	CCACCTCCAGCCGTCGTGCC		8		ccuccagc				
			20	13	ccuccagccgcaugccacc				
HEK4-OT4	GGCATCACGGCTGGAGGTGG		8		ccuccagc				
			20	13	ccuccagccgcaugccacc				
HBB	CAGGAGTCAGgTGCACCATG	4 A to T	13		gagucaggugcac	14	agacuucuccacag	gcauggugcaccugacuccug	gccuugauaccaaccugccca

			20	13	gagucaggugca -cauggug				
HBB -OT	CAGGAGTCAGATG CACCATG		13		gagucaggugca c				
			20	13	gagucaggugca -cauggug				