SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3382.

SARAH LAURA KRAUSZ

Patobiokémia

című program

Programvezető: Dr. Csala Miklós, egyetemi tanár

Témavezető: Dr. Welker Ervin, tudományos tanácsadó

ProPE: A New Gene Editing Approach for Increasing the Editing Efficiency and Specificity of Prime Editing

PhD thesis

Sarah Laura Krausz

Molecular Medicine Doctoral School
Semmelweis University

Supervisor: Ervin Welker, D.Sc
Official reviewers: Törőcsik Beáta, Ph.D
Kovács Tibor, Ph.D

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

Members of the Complex Examination Committee: Zsolt Rónai, D.Sc
József Kardos, Ph.D

Budapest
2025

Table of Contents

List of Abbreviations

AAV	Adeno-Associated Virus
CIGAR	Compact Idiosyncratic Gapped Alignment Report
ClinVar	Clinical Variant Database
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CYP genes	Cytochrome P450 genes
engRNA	Essential Nicking Guide RNA
etpgRNA/epegRNA	Engineered tpgRNA / pegRNA
GFP	Green Fluorescent Protein
HEK293 cell line	Human Embryonic Kidney cell line
HuES cell line	Human Embryonic Stem cell line
indel	Insertion / Deletion
K562 cell line	Human Chronic Myelogenous Leukemia cell line
PAM	Protospacer Adjacent Motif
PBS	Primer Binding Site
PE	Prime Editing
MCP	MS2 Coat Protein, binding MS2 aptamer
MLH1d	Dominant-Negative Mutant of MLH1
N22p	22-amino-acid from the N protein, binding BoxB aptamer
NGS	Next-Generation Sequencing
PEAR	Prime Editing Activity Reporter
pegRNA	Prime Editing Guide RNA
proPE	Prime Editing with Prolonged Editing Window
R1	Raw Data File 1, from i5 Direction
R2	Raw Data File 2, from i7 Direction
RT	Reverse Transcriptase
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
RTT	Reverse Transcription Template
SaCas9	Staphylococcus aureus Cas9 (alternative CRISPR nuclease)
sgRNA	Single guide RNA
SNP	Single Nucleotide Polymorphism

SnPE	Split pegRNA prime editor
SpCas9	Streptococcus pyogenes Cas9
sPE	Split Prime Editor
SpRY	A PAM-flexible SpCas9 variant
tdMCP	Tandem Dimeric MS2 Coat Protein, binding MS2 aptamer
tpgRNA	Template-Providing Guide RNA
U2OS cell line	The Uppsala 2 Osteosarcoma cell line

1. Introduction

The discovery of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) revolutionized gene editing by enabling the gene editing machinery to be easily reprogrammed through changing the so-called spacer sequence within its RNA(1). *Streptococcus pyogenes* Cas9 (SpCas9) is the most widely used CRISPR-associated nuclease, enabling site-specific genome modifications via the induction of DNA double-strand breaks(2,3). These breaks are subsequently repaired by the cell's DNA repair pathways, typically resulting in small insertions or deletions (indels), or precise sequence changes when a donor DNA template is provided(4). However, the use of nucleases can lead to chromosomal rearrangements and on-target aberrations(5), which are detrimental to therapeutic applications. Moreover, the donor template used for homology-directed repair can also integrate into the genome via alternative pathways. To mitigate these risks, several approaches have been developed that utilize a nickase variant of Cas9, which introduces single-strand rather than double-strand breaks(6–8). Prime editing represents the most recent and is considered one of the most promising advancement among these nickase-based editing systems(8). It enables the precise installation of substitutions, small deletions, and insertions without inducing double-strand breaks or requiring a donor DNA template, thereby minimizing the occurrence of unintended modifications.

Prime editing is a complex genome engineering process that employs a ribonucleoprotein complex known as the prime editor. The prime editor comprises a fusion protein (consisting of a nickase version of SpCas9 and a reverse transcriptase) and a 3′-extended single guide RNA, referred to as the prime editing guide RNA (pegRNA). The 3' extension of the pegRNA contains the primer binding site (PBS) and the reverse transcriptase template (RTT), which includes the desired edit. The mechanism of prime editing can be described in six sequential steps. First, the prime editor binds to the target site which is complementary to its spacer sequence and introduces a single-strand break (nick) on the non-target DNA strand (Fig. 1/1.). The released 3' DNA end hybridizes to the complementary PBS sequence of the pegRNA and functions as a primer for the reverse transcriptase (Fig. 1/2.). The reverse transcriptase then extends the non-target DNA strand along the RTT template, thereby incorporating the desired edit into the DNA (Fig. 1/3.). This newly synthesized strand forms a 3′ DNA flap that equilibrates with the original, unedited 5′ flap (Fig. 1/4.). The 5′ flap is preferably cleaved by an endogenous

flap endonuclease, then the edited strand is ligated (Fig. 1/5.). Finally, the introduced edit may become fixed by mismatch repair or during replication (Fig. 1/6.).

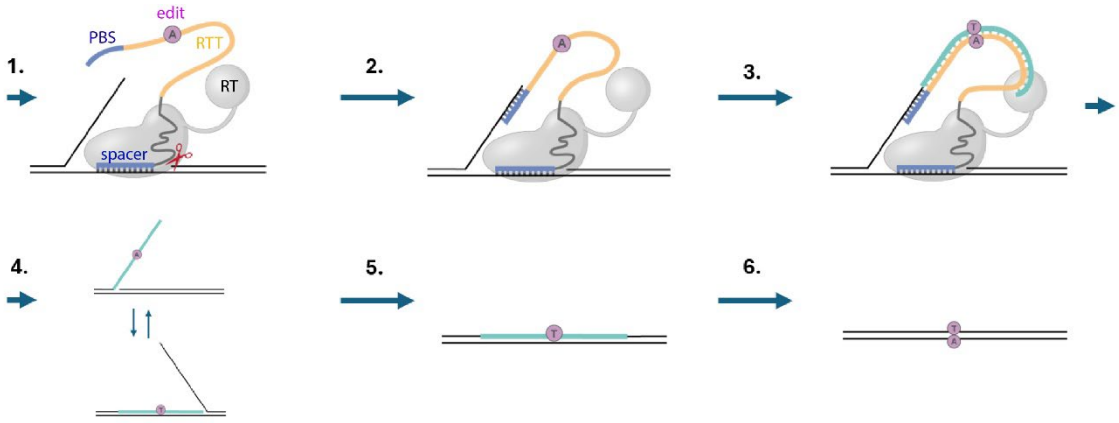

Figure 1.: Schematic representation of the prime editing process. The prime editor binds to and nicks its target site. The PBS hybridizes to the released 3' DNA strand, which serves as a primer for the reverse transcriptase. The reverse transcriptase then extends the DNA end along the RTT template. The newly synthesized strand forms a 3′ DNA flap that equilibrates with the original 5′ flap. The 5′ flap is preferably cleaved by endogenous endonucleases, after which the edited strand is ligated. Finally, the edit may become fixed by the mismatch repair or during replication. PBS, RTT, RT refer to the primer binding site, reverse transcription template, reverse transcriptase, respectively.

Unfortunately, the efficiency of PE to introduce certain edits remains low, likely due to a combination of inhibitory factors. The complexity of the PE process makes it difficult to identify the determinants that are responsible for this low activity. A variety of recent studies have attempted to identify these limiting factors and increase the efficiency of PE by addressing key bottlenecks. Many of these efforts have achieved considerable success; some of these approaches are introduced below.

The 1st step of PE (Fig. 1/1.) may be inhibited by the PBS-spacer intramolecular interaction as the two sequences are complementary. Reduced activity associated with long PBSs is well-documented in the literature(9–13). While shortening the PBS can mitigate this inhibition, it may simultaneously weaken hybridization with the 3′ end of the target DNA, thereby reducing priming efficiency. As a result, each PBS typically requires extensive optimization. Several studies have attempted to alleviate this burden

by developing prediction tools(10,14–18), some of which based on the melting temperature (Tm) of the PBS(10,11). However, these approaches have achieved limited success, likely because additional factors contribute to determining the optimal PBS length and sequence. Other strategies aim to overcome this inhibition by introducing mismatches within the spacer sequence(19), refolding of the pegRNA prior to its application in RNP format(20), or subjecting transfected cells to cold shock(11).

The 2nd step of PE (Fig. 1/2.) may be hindered by the degradations of the 3' end of the pegRNA, as it is not protected by the bound Cas9 prime editor protein. The degradation of the 3' extensions of pegRNAs during prime editing has been demonstrated using northern blot and RT-qPCR analyses(21). The inhibitory nature of these degraded pegRNAs was highlighted by Nelson et al.(21) suggesting that degraded, editing-incompetent pegRNAs inhibit PE by occupying the target site, thereby preventing intact pegRNAs from binding and thereby from facilitating editing. An approach to mitigate this, is to incorporate a protective 3' end downstream of the PBS(21–25) Although these protective methods did not entirely prevent degradation(21,25), it could substantially increase prime editing efficiency(21–25).

The 4th step of PE (Fig. 1/4.) may be inhibited by 3’ truncated DNA flaps, which could be the result of incomplete reverse transcription or 3’ digestion. DNA flaps are known to undergo extensive digestion by exonucleases and endonucleases via DNA repair in cells(26–28). If the degradation of the newly transcribed DNA flap occurs during prime editing eliminating the intended edit, it unavoidably decreases PE efficiency. This diminishing effect has also been experimentally demonstrated by co-expressing a flap-nuclease in a recent study(29). While no study directly addressed this limitation, Koeppel and Weller et al.(29) found circumstantial evidence that more structured flaps, which are more resistant to degradation, can be inserted into the genome at higher efficiency.

The efficiency of incorporating the new DNA strand in step 5 (Fig. 1/5.) could be increased by designing an alternative pathway with two oppositely oriented pegRNAs, allowing direct hybridization of the new DNA strands(30,31). This approach proved to be particularly effective for creating large deletions and insertions(32–38).

The low editing efficiency frequently can be increased by nicking the unmodified DNA strand, which biases the mismatch repair and can improve the fixation of the edit in

the final step of the process (Fig. 1/6.) (8). However, this also increases the occurrence of the indel background at the nicked sites(8).

Further work revealed that in the final step (Fig. 1/6.) mismatch repair is responsible for the reduced efficiency of PE for many targets and edits. Attempts to alleviate this problem have been made by inhibiting the DNA repair protein MLH1(39,40) or by introducing an additional edit so that then the resulting double-edit is no longer a substrate for mismatch repair(41).

Even after the successful installation of the edit, the measured efficiency may be reduced as SpCas9 nuclease is known to repeatedly cleave its target if the target sequence remains unmodified during consecutive cleavage-repair cycles(42–44). A large body of evidence shows that PE is more efficient when edits alter the PAM sequence(9,13,45), which is required for Cas9 recognition. This bottleneck can be addressed by co-editing the PAM site if it does not interfere with the function of the sequence; this strategy has been shown to enhance editing efficiency at other target positions(13,45).

In addition to low editing efficiency and the frequent occurrence of indels in certain cases, two major challenges limit the broad therapeutic application of PE: the absence of a suitable PAM sequence at the targeted single-nucleotide polymorphism (SNP) site and the large size of the prime editor protein, which limits its compatibility with adeno-associated virus (AAV) vectors. To expand the targetable sequence space, PAM-modified PE variants have been developed(46–49). Furthermore, to address the AAV packaging limitation, smaller prime editors have been engineered by either separating the reverse transcriptase and the Cas9 nuclease components(50–52) or by modifying the reverse transcriptase domain itself(50,53–55).

2. Objectives

My first objective is to develop a python script for the bioinformatic analysis of Illumina NGS data obtained from prime editing experiments. This will allow the evaluation of prime editing efficiency and the characterization of indel backgrounds associated with editing. This tool helped in the development of an easy-to-use plasmid-based assay suitable for the enrichment of successfully edited cells and the systematic analysis of factors influencing editing.

My second objective is to develop a prime editing tool, proPE (Prime Editing with Prolonged Editing Window) which overcomes the spacer-PBS interaction limiting prime editing activity by providing the spacer and PBS on two distinct CRISPR/SpCas9 complexes. I also aim to determine the parameter ranges where it is functioning optimally.

My third objective is to explore its potential for therapeutic applications by identifying editing scenarios in which it enables more efficient and safer editing compared to conventional prime editing, and by evaluating its applicability in AAV-based delivery systems. Furthermore, we aim to explore its potential for reducing off-target effects, resulting from the system's design employing two distinct target sites.

My fourth objective is to investigate the mechanisms most likely responsible for the improved efficiency and specificity of proPE.

3. Methods

3.1 Plasmid construction

All linkers and PCR primers used in this study were purchased from Sigma-Aldrich. Oligonucleotides used for cloning the plasmid constructs in AAV production were kindly provided by Dr. Györgyi Ferenc, Laboratory of Nucleic Acid Synthesis, Institute of Plant Biology, HUN-REN BRC, Szeged, Hungary.

3.1.1 RNA-expressing plasmids

To monitor transfection efficiency, the RNA-expressing plasmids contain an mCherry expression cassette for GFP-PEAR experiments, and a TagBFP expression cassette for mScarlet-PEAR and NGS experiments. The construction of the RNA cloning plasmids is detailed below for each RNA type. Spacer coding linkers were inserted into the RNA cloning plasmids between BpiI sites using 3 units of the BpiI enzyme, 2 units of T4 DNA ligase, 500 μM ATP, 1× Green buffer, 50 ng vector, and 0.25 μM of each oligonucleotide. For RTT-PBS cloning, the unique linker was inserted into either the RNA cloning plasmid or the spacer-containing RNA cloning plasmids between Esp3I sites using 3 units of the Esp3I enzyme, 2 units of T4 DNA ligase, 500 μM ATP, 1× Tango buffer, 1 mM DTT, 50 ng vector, and 0.25 μM of each oligonucleotide.

Cloning plasmid for engRNA- and 2nd nicking RNA-expressing plasmids: Cloning plasmids for SpCas9 sgRNA were created by Simon and colleagues(56); pAT9658-sgRNA-mCherry and pAT9679-sgRNA-BFP. An SaCas9 engRNA cloning plasmid (pSLK20330-Sa-sgRNA-mCherry) was constructed by assembling linkers encoding the SaCas9 sgRNA scaffold into the BpiI and Esp3I digested pDAS12069-U6-pegRNA-mCherry using NEB HiFi Assembly.

Cloning plasmid for tpg/pegRNA-expressing plasmids: Tpg/pegRNA cloning plasmids without the tevopreQ1 extension were created by Simon and colleagues(56); pDAS12069-U6-pegRNA-mCherry and pDAS12222-U6-pegRNA-BFP. For most of the experiments tpg/pegRNAs with a tevopreQ1 extension(21) were used. This extension provides protection against the 3’ degradation of the tpg/pegRNA. For tpg/pegRNA cloning plasmids with a tevopreQ1 extension a linker containing an exchangeable cassette

followed by the tevopreQ1 extension was cloned between the Esp3I sites into pDAS12222-U6-pegRNA-BFP and pDAS12069-U6-pegRNA-mCherry plasmids, resulting in pSLK7824-U6-pegRNA-epeg-BFP and pSLK7822-U6-pegRNA-epeg-mCherry, respectively.

Cloning plasmid for petRNA expressing plasmids: A cloning vector for petRNA23 with the same background backbone as tpg/pegRNAs vector, pSLK20322-U6-petRNA-BFP, was constructed using NEB HiFi Assembly. PCR was employed to amplify the essential 5′ and 3′ components of the circular RNA (ribozyme, ligation arm, MS2, and flanking sequences) from the petRNA cloning plasmid (acquired from Addgene; #181802). The fragments were subsequently assembled into BpiI and Esp3I digested pDAS12222-U6-pegRNA-BFP.

Cloning plasmid for 5'-MS2 and 5'-BoxB RNA-expressing plasmids: To generate plasmids for subsequent cloning containing a 5′ MS2 or BoxB sequence along with the tevopreQ1 extension, the constructs pSLK20323-U6-MS2-epeg-BFP and pSLK20324-U6-BoxB-epeg-BFP were constructed using NEB HiFi Assembly. A linker containing either the MS2 or BoxB sequences, as well as the sequence of an exchangeable cassette followed by the tevopreQ1 extension4, were inserted into the BpiI and Eco32I digested pSLK7824-U6-pegRNA-epeg-BFP.

3.1.2 PEAR plasmids

The PEAR-mScarlet plasmid (#162991) used in this study and the pAT9624-BEAR-cloning plasmid (#162986) were created by Tálas and colleagues(57). PEAR-GFP plasmids with different eng-tpg target distances and with the SaCas9-engRNA target site were constructed from the pAT9624-BEAR-cloning plasmid (#162986), also created by Tálas and colleagues(57). The linkers coding the targets were cloned into pAT9624 plasmid between Esp3I sites.

3.1.3 Prime editor-expressing plasmids

The following plasmids were obtained from the non-profit plasmid distribution service Addgene: pCMV-PE2 (#132775) created by Anzalone et al.(9) (the original prime editor), pCMV-PEmax-P2A-hMLH1dn (#174828) created by Chen et al.(39) (offering enhanced prime editing by introducing mismatches in the prime editor and suppressing

mismatch repair by co-expressing hMLH1 domain) and pCMV-SaCas9-PE (#169851) created by Liu et al. (58) Several PE-expressing plasmids were constructed using NEB HiFi Assembly, as described below.

PE2max (offering enhanced prime editing by mismatches introduced in the prime editor): PCR fragment synthesized from pCMV-PEmax-P2A-hMLH1dn was used to insert the prime editor coding sequence into NotI and MssI linearized pCMV-PE2 (Addgene number #132775).

tdMCP-PE4max and N22-PE4max: Linkers were used to insert the tdMCP and N22 coding sequences into pCMV-PEmax-P2A-hMLH1dn vector linearized using SpCas9 in vitro.

nCas9max for sPE: It was created by removing the RT coding sequence from pCMV-PEmax-P2A-hMLH1dn. The plasmid was digested with EcoRI, overlapping segments were created by PCR from the same plasmid, and these fragments were assembled.

MCP-M-MLV-RTmax for sPE: RTmax was amplified via PCR from pCMV-PEmax-P2A-hMLH1dn, while the plasmid backbone and MCP were amplified from the MCP-M-MLV-RT vector (created by Liu et al. and acquired from Addgene, #181799). Then the PCR fragments were assembled.

PEmax (SPRY-Cas9): SPRY-Cas9 was amplified via PCR, while an upstream fragment was amplified from the PE2max plasmid to generate an overlapping fragment which were assembled into NotI and CpoI digested PE2max.

Inactive SpCas9 prime editor: Two regions of nSpCas9 coding sequence from the pCMV-PEmax-P2A-hMLH1dn vector were PCR amplified with primers overlapping and containing the D10A mutation in their flanking regions. The resulting fragments were assembled into NotI and BamHI digested pCMV-PEmax-P2A-hMLH1dn.

3.1.4 AAV plasmids

p601m-AAV-v3em-Nterm-PE2max (#198734) and p601m-AAV-v3em-Cterm-PE2max-ΔRNaseH-dualU6 (#198735), created by Davis et al.(59), were acquired from the non-profit plasmid distribution service Addgene.

p601m-AAV-v3em-Cterm-PE2max-ΔRNaseH-dualU6 construct was used to create plasmids containing the C terminal of PE, tpgRNA-expression cassette and a

second expression cassette either for the engRNA or for the 2nd nicking sgRNA. The constructs were assembled using NEB HiFi Assembly. For both constructs PCR was used to generate overlapping fragments for the assembly by amplifying the U6 promoter from AAV-v3em-Cterm-PE2max-ΔRNaseH-dualU6. PCR was used to amplify the tpgRNA coding sequence from 'tpg_PRNP(1,3-11)_6 G to T (G127V)_RTT23-PBS12_epeg'-expressing plasmid, the 2nd nicking sgRNA coding sequences from '2nd nick_PRNP(4)'-expressing plasmid and engRNA coding sequence from 'eng_PRNP(1)'-expressing plasmid. The two constructs were created by assembling the corresponding fragments with the NotI and HindIII digested p601m-AAV-v3em-Cterm-PE2max-ΔRNaseH-dualU6.

The sequences of all plasmid constructs were confirmed by Sanger sequencing (Microsynth AG).

3.2 Cell culturing, transfection and transduction

HEK293T (CRL-3216) cell line was from ATCC, was authenticated by its respective supplier and regularly tested negative for mycoplasma.

HEK293 cells were grown in DMEM supplemented with 10% heat-inactivated FBS with 100 units/ml penicillin and 100 µg/ml streptomycin. Cells were cultured at 37°C in a humidified atmosphere of 5% CO2 and were passaged every 3–4 days.

Transfections were performed in triplicates. Transfected cells were analyzed by flow cytometry three days post-transfection, either for PEAR experiments or to assess transfection efficiency, followed by genomic DNA purification.

3.2.1 Transfection of HEK293 cells

Table 1 shows the amount of plasmid transfected into HEK293 cells. The amount of the engRNA and tpgRNA coding plasmids varied for various proPE conditions; different fractions of the 31ng of 'standard' amount of engRNA coding plasmid were used (1/1: 31 ng, 1/5: 6.2 ng, 1/10: 3.1 ng, 1/12: 2.6 ng, 1/20: 1.5 ng, 1/40: 0.8 ng, 1/80: 0.4 ng), referred to as the 'engRNA fraction'. The amount of tpgRNA plasmid varied such that the sum of plasmids encoding engRNA and tpgRNA was equal to the amount of pegRNA in the corresponding control experiment. Different amounts of engRNA and/or tpgRNA coding plasmids were used in experiments shown in Figure 9,13,17. The

amount of prime editor was modified as detailed below for experiments shown in Figure 13 and Figure 17.

Table 1: Table shows the amount of the plasmids transfected into HEK293 cells. All experiments were performed with the indicated quantities highlighted by grey background. Quantities for conditions other than these (Figures 9) are indicated individually in the table.

	sum (ng)	prime editor/ SpCas9 (ng)	pegRNA / (engRNA + tpgRNA) (ng)	2nd nicking sgRNA (ng)	PEAR plasmid (ng)
PEAR - (without 2nd nicking sgRNA)	350	223	91	-	36
PEAR - (with 2nd nicking sgRNA)	350	223	55	36	36
genomic - (without 2nd nicking sgRNA)	350	248	102	-	-
genomic - (with 2nd nicking sgRNA)	350	248	62	40	-
Fig. 9 - (using only non-targeting 2nd nicking sgRNA)	331	211	52.5	34	34

Fig. 9: In the experiment with variable amounts of engRNA coding plasmid (26.25 ng, 8.75 ng, 4.77 ng, 2.50 ng, 1.28 ng, 0.65 ng, 0.33 ng, 0.16 ng) a fixed amount of tpgRNA coding plasmid (26.25 ng) was used. In the experiment with variable amounts of tpgRNA coding plasmid (51.85 ng, 45.11 ng, 38.37 ng, 31.63 ng, 24.89 ng, 18.15 ng,

11.41 ng, 4.67 ng) a fixed amount of engRNA coding plasmid (0.65 ng) was used. Plasmid encoding a non-targeting sgRNA was used to keep the amount of total DNA constant at 331 ng across conditions, when needed.

Fig. 13: Editing of sPE, SnPE and proPE was performed under conditions where the amounts of both the total RNA and protein coding plasmids, as well as the nicking activity, were kept equal. For sPE, a 3:1 ratio of nCas9 coding plasmid (ng) to RT coding plasmid (ng) was used, as this ratio seemed to demonstrate the highest activity according to Liu et al. (Extended Data Fig. 2f in their publication).

Fig. 17: In the PEAR experiment (without 2nd nicking sgRNA), the prime editor protein harboring a dead SpCas9 and the SaCas9 prime editor protein were used in a 1:3 ratio (dead SpCas9 prime editor: 56 ng, SaCas9 prime editor: 167 ng) with 1/1 engRNA fraction (31ng engRNA). Variable amounts of tpgRNA coding plasmids were used (60 ng, 45 ng, 30 ng, 27 ng, 23 ng, 20 ng, 17 ng, 13 ng, 10 ng, 7 ng, 3 ng, 0 ng), with the proportion targeting or non-targeting degraded tpg/pegRNA increased correspondingly as the tpgRNA amount was reduced.

HEK293 cells were seeded on 48-well plates 1 day before transfection at a density of 3-5×10^4 cells/well. For all experiments the total DNA was mixed with 0.9 µl turbofect reagent diluted in 50 µl serum-free DMEM and added to the cells after 20 min incubation at RT.

3.2.2 Transfection and transduction of HEK293T cells with AAV

AAV viruses were purchased from Creative Cell Kft. HEK293 cells were seeded on 48-well plates at a density of 3×10^4 cells/well one day before transduction. Cells were transduced with a total of 1.43x106 MOI of AAV, using a ratio of 1 : 0.84 : 0.16 for the N-terminal PE2max, C-terminal PE2max with the engRNA and tpgRNA, and C-terminal PE2max with the 2nd nicking sgRNA and tpgRNA, coding sequences, respectively. Genomic DNA was extracted three days post-transduction.

For AAV plasmid transfection 170.1 ng N-terminal PE2max, 146.4 ng C-terminal PE2max with the engRNA and tpgRNA and 28.5 ng C-terminal PE2max with the 2nd nicking sgRNA and tpgRNA, sequence coding AAV plasmids were used.

3.3 Flow cytometry

Flow cytometry analysis was carried out using an Attune NxT Acoustic Focusing Cytometer (Applied Biosystems by Life Technologies). As a rule, signals from a set target minimum of 10,000 viable single cells were acquired by gating based on the side and forward light-scatter parameters. BFP, GFP, mCherry, and mScarlet signals were detected using the 405 (for BFP), 488 (for GFP), and 561 nm (for mCherry and mScarlet) diode laser for excitation, and the 440/50 (BFP), 530/30 (GFP), 620/15 (mCherry), and 585/16 nm (mScarlet) filters for emission. Percentage of GFP or mScarlet positive cells was calculated as the proportion of GFP+mCherry or mScarlet+BFP double positive cells within the mCherry or BFP positive cell population, respectively. mCherry and BFP were used as indicators of the efficiency of transfection. Attune Cytometric Software v.4.2 was used for data analysis.

3.4 Genomic DNA purification and genomic PCR

After flow cytometry genomic DNA was extracted using the Puregene DNA Purification protocol (Gentra Systems Inc). Amplicons for next-generation sequencing were generated from the genomic DNA samples using two rounds of PCR to attach Illumina handles. PCR was done in a S1000 Thermal Cycler (Bio-Rad) or PCRmax Alpha AC2 Thermal Cycler using Q5 high-fidelity polymerase supplemented with Q5 buffer, and 150 ng of genomic DNA in a total volume of 50 µl. The thermal cycling profile of the PCR was: 98 °C 30 sec; 35 × (denaturation: 98 °C 20 sec; annealing temperature 30 sec; elongation: 72 °C); 72 °C 5 min. i5 and i7 Illumina adapters were added in a second PCR reaction using Q5 high-fidelity polymerase with supplied Q5 buffer and 1 µl of first step PCR product in total volume of 50 µl. The thermal cycling profile of the PCR was: 98 °C 30 sec; 35 × (98 °C 20 sec, 67 °C 30 sec, 72 °C 20 sec); 72 °C 2 min. Amplicons were purified by agarose gel electrophoresis. Samples were quantified with Qubit dsDNA HS Assay kit and pooled.

3.5 Next-generation sequencing, indel, and editing frequency analysis

Samples were sequenced on NextSeq (Illumina) with paired-end sequencing resulting in 2 ×150 bp reads, by Deltabio Ltd. Reads were aligned to the reference

sequence using BBMap. Primer dimers that were found among the aligned reads of the FANCF and PRNP amplicons, were removed from further analysis.

Indels at eng/pegRNA and 2nd nicking sgRNA target sites were computationally counted from the aligned reads. Indels without mismatches were searched at ± 2 bp around the cut sites. For each sample, indel frequency was determined as (number of reads with indel either at the eng/pegRNA target site or at the 2nd nicking sgRNA target site)/(number of total reads). Frequency of precise edits generated by prime editing was determined as the percentage of (sequencing reads with the desired modification without indels)/(number of total reads). For intended insertions or deletions in the window of ± 2 bp around the cut sites generated by editing, the frequency of precise edits was determined as the percentage of (all sequencing reads with the desired modification)/(number of total reads). For these samples, the indel background was calculated from reads containing indels without considering the desired indel edits. Reads with the intended modifications were identified by searching for a sequence stretch containing the desired edit, flanked by 5-5 matching nucleotides. We used modified sequence stretch if the corresponding WT sequence was found in less than 68% of the reads derived from the empty cells indicating high sequencing error in the region. In these cases, we modified the sequencing stretch by allowing any type of nucleotide at the position with high sequencing error and an additional nucleotide was added to the particular sequence, keeping 5 matching nucleotides on either side of the edit. The following software were used: BBMap 38.08, samtools 1.8, BioPython 1.71, and PySam 0.13 to analyse the NGS data. From the value of each independent sample in a triplicate the average edit or indel value of the empty cells was subtracted. When specificity values were calculated for the samples and the resulting values were less than 0.3%, then it was arbitrarily set to 0.3% to avoid unrealistic specificity values. The average of the three processed values of a triplicate was then calculated. Specificity values were calculated by dividing the edit values by the indel values for each sample of a triplicate and then taking the average of these ratios.

Several edits were introduced to the CYP gene. Studying modification at this 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.

3.6 Statistics

Homogeneity of variances was tested by Brown-Forsythe test and normality of residuals was tested by D'Agostino-Pearson omnibus (K2) test. For data sets with a normal distribution statistical significance was assessed by a two-tailed unpaired t-test for comparing two groups. For comparing more than two groups one-way ANOVA or RM one-way ANOVA (in the case of paired comparison) was used, followed by Šídák's multiple comparisons test (when comparing only the selected groups), Tukey's post hoc test (when comparing every group to each other group) or Dunnett's test (when comparing every group to a control group) was used. In cases where the data did not follow a normal distribution, significance was assessed using a two-tailed Mann-Whitney test for comparisons between two groups. For comparisons involving more than two groups, the Kruskal-Wallis test was used, or the Friedman test in the case of paired comparisons, both followed by Dunn's test. Statistical tests were performed using GraphPad Prism 9.2. Medians of each group are shown on each graph.

Fig. 16: For each normalized dataset, a straight-line model was fitted using non-linear regression with the least squares fitting method. The null hypothesis that the best-fit slope is the same for all datasets was tested with an extra sum-of-squares F-test. Error bars include error propagation.

4. Results

4.1 Python Scripting for PEAR Development

PEAR (Prime Editing Activity Reporter) is a plasmid-based fluorescent assay designed for the rapid evaluation of prime editing activity and the enrichment of edited cells. The PEAR plasmid contains a GFP sequence disrupted by a non-functional intron. In particular, the acceptor site of the intron is mutated, preventing splicing and resulting in non-functional GFP expression. This mutation can be corrected by prime editing, restoring functional GFP expression. The resulting percentage of the GFP-fluorescent cells reflects the prime editing efficiency (Fig. 2).

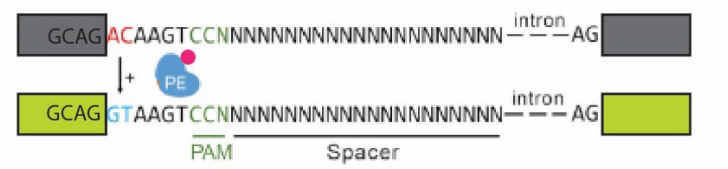

Figure 2.: Schematic figure of the PEAR system. The GFP sequence is disrupted by a non-functional intron, which can be corrected by prime editing and can result in the appearance of the green fluorescence signal.

PEAR plasmid facilitates enrichment of the cells where prime editing occurred on the genomic target. In enrichment experiments, two distinct pegRNAs are employed; one targeting the genomic site and another targeting the PEAR plasmid. Efficient editing of the PEAR plasmid leads to GFP fluorescence, and sorting GFP-positive cells enables the enrichment of cells with successfully edited genomic sites.

To analyze prime editing activity at the genomic sites using data obtained from Illumina next-generation sequencing (NGS), we developed a Python script. The analysis pipeline begins with the merging of paired-end sequencing reads. Typically, sequencing outputs two raw data files per index pair; the "R1" file containing sequences from one direction (i5) and the "R2" file from the opposite direction (i7). If the raw data are not pre-sorted by index pairs, a demultiplexing step is included to assign reads to their index pairs. Merging the raw data files creates sequences that combine the information from both the R1 and R2 reads, capturing the full sequence of the amplicon from the i5 direction. We created the appropriate reference files for aligning the merged reads and

then assessed the indel background around the nick sites from the CIGAR string, an alignment attribute, followed by the detection of a specific sequence string including the desired edit (Fig. 3).

Figure 3.: The pipeline of data analysis. The two raw data files (R1 and R2) were merged and aligned to its reference file. The aligned file was analyzed to obtain the editing percentage and the indel background of the editing.

The PEAR assay enables systematic examination of various parameters of a given prime editing system, which will be exploited in the following sections of this thesis.

4.2 Development of proPE

ProPE, short for prime editing with prolonged editing window, is a prime editing tool designed to enhance prime editing efficiency. A pegRNA has two different functional components: the reverse transcription template (RTT-PBS) required for reverse transcription and the spacer responsible for target recognition. The PBS region of the RTT-PBS is complementary to the spacer, creating an intramolecular interaction that may reduce prime editing activity (Fig. 4a).

To overcome this intramolecular inhibition, we separated the RTT-PBS and spacer into two different single guide RNAs (sgRNAs); the essential nicking gRNA (engRNA) and the template providing gRNA (tpgRNA). The engRNA nicks the target site recognized by its spacer sequence, which is identical to the spacer of the corresponding pegRNA, while the tpgRNA provides the primer and template for reverse transcription. The tpgRNA is directed to a nearby target in the vicinity of the nicked DNA strand by its short spacer sequence which facilitates binding without cleavage (Fig. 4b).

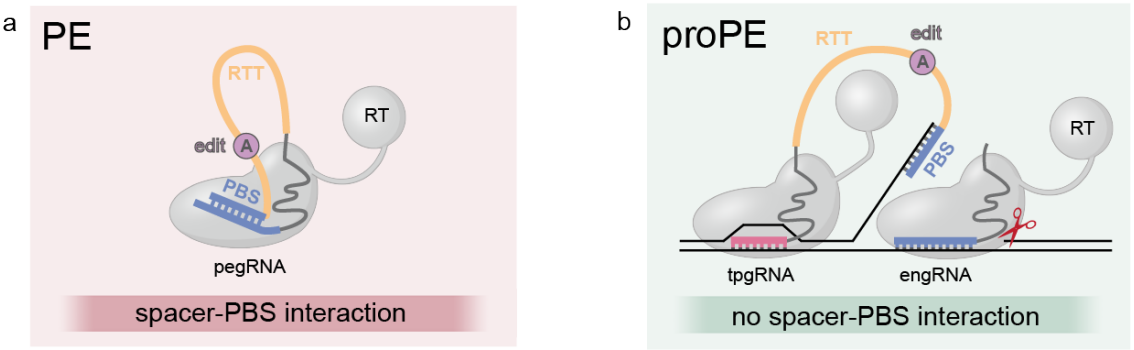

Figure 4.: Schematic figure illustrating the development of proPE; the inhibition of PBS–spacer intramolecular interaction observed in PE (**a**) can be overcome in proPE by separating the two sequences into distinct sgRNAs (**b**). engRNA, tpgRNA, pegRNA, PBS, RTT, RT refer to the essential nicking gRNA, template providing gRNA, prime editing gRNA, primer binding site, reverse transcription template, reverse transcriptase, respectively.

We introduced proof-of-concept experiments on the EMX1 genomic site and on a PEAR assay integrated into the genome. Specifically, we tested whether both the engRNA and a nearby targeting tpgRNA are required for editing and compared proPE to conventional PE. Our results indicate that both the engRNA and tpgRNA are essential for proPE editing, as no editing was observed in the absence of either RNA. Targeting the tpgRNA to a site near the engRNA target significantly enhances editing efficiency beyond that of conventional PE, although it is not strictly required (Fig. 5a,b).

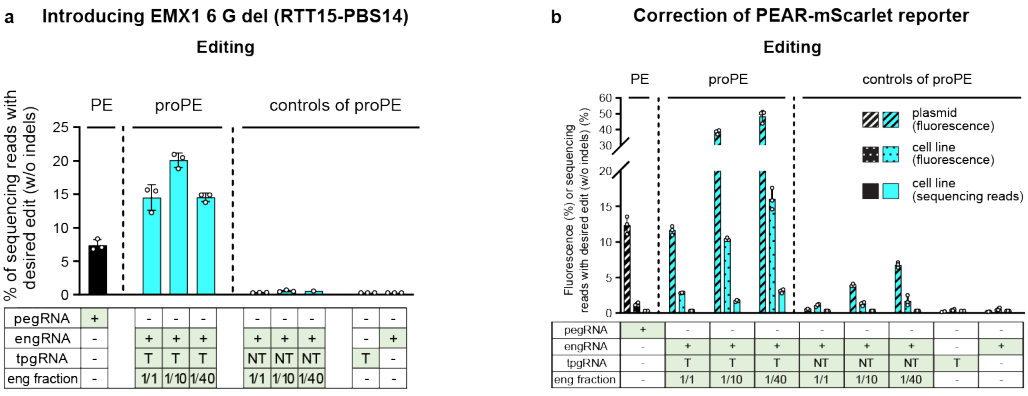

Figure 5.: Both engRNA and tpgRNA are essential for proPE editing. a, Amplicon sequencing results of introducing 6 G deletion at the EMX1 site by PE and by proPE with varying its components. b, The same experiment was performed using the PEAR plasmid

in HEK293 cells and a HEK293 cell line with genome-integrated PEAR copies. Fluorescence data were obtained from both cell lines, while NGS data was available only from the HEK293 cell line carrying genome-integrated copies. Different 'eng fractions' were used in the experiments which will be explained later. 'T' and 'NT' indicate the targeting and non-targeting tpgRNA, respectively. All data points were calculated as the mean of triplicates.

4.3 Characterization of proPE

Next, we identified some of its working parameters; specifically, we investigated how the distance between the engRNA and tpgRNA target sites and their orientation, the spacer length of the tpgRNA, and the amounts of engRNA and tpgRNA applied influence editing efficiency. These experiments were conducted using the PEAR assay, which allows systematic evaluation of an individual proPE parameter while keeping all others constant.

To examine the spatial configuration, we tested 26 different distances between the engRNA and tpgRNA target sites, in both trans and cis orientations (Fig. 6), using the same engRNAs and tpgRNAs for all the distances and orientations. We found that effective editing occurred when the two target sites were separated by 0 to 30 nucleotides in trans orientation, or 6 to 30 nucleotides in cis orientation, corresponding to ~70 bp and ~45 bp between the two NGG PAMs, respectively. Interestingly, a periodicity of around 10 nt is apparent in the editing efficiencies in both orientations, suggesting that the coiling of the DNA helix affects the optimal assembly of the proPE components and their efficiency (Fig. 6).

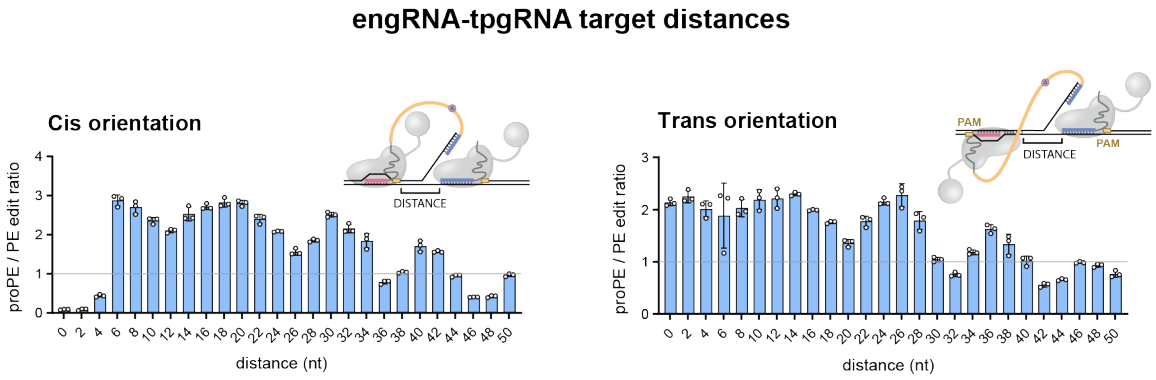

Figure 6.: PAM for tpgRNA is likely to be found at an appropriate distance from the engRNA target site. Fluorescence data, reflecting editing activity at various engRNA-tpgRNA distances, were obtained using the PEAR assay. ProPE/PE edit ratio is shown for both cis and trans orientations. Light grey line indicates the ratio, where the editing efficiency of PE and proPE is the same. All data points were calculated as the mean of triplicates.

These findings indicate that a PAM for the tpgRNAs is generally found at an appropriate distance from the engRNA target site. Indeed, our analysis of the ClinVar database revealed that in 98.6% of the engRNA targets, which are closest to a human pathogenic SNP, there is at least one tpgRNA target site within the optimal distance and orientation from the engRNA targets.

Nevertheless, we explored whether the targetable scope could be further expanded using SpRY, a Cas9 variant developed by Walton et al.(49), which recognizes a relaxed "NRN" PAM sequence. We tested the activity of PE-SpRY and proPE-SpRY on NGG targets and compared it to the WT variants. Consistent with the literature(60), SpRY-based prime editors showed low editing efficiency in both PE and proPE settings (Fig. 7).

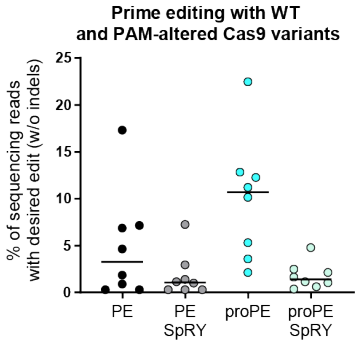

Figure 7.: SpRY-based prime editing systems exhibit low editing efficiency with both PE and proPE. Amplicon sequencing results show editing efficiencies for PE, the corresponding proPE, and their SpRY variants. All data points were calculated as the mean of triplicates.

Next, we investigated the effective length of the tpgRNA spacer ranging from 5 to 15. It is known that the spacer lengths shorter than 15 nt typically do not cleave their targets(61) (Fig. 8a), which is essential for tpgRNA function. We used four distinct proPE combinations to investigate the spacer length of the tpgRNA. Each combination is defined by a specific engRNA, a tpgRNA, and a second-nicking sgRNA. ProPE editing was still observed with tpgRNAs containing 5-nt-long spacers, however editing efficiency declined when spacers were shorter than 10 nt (Fig. 8b).

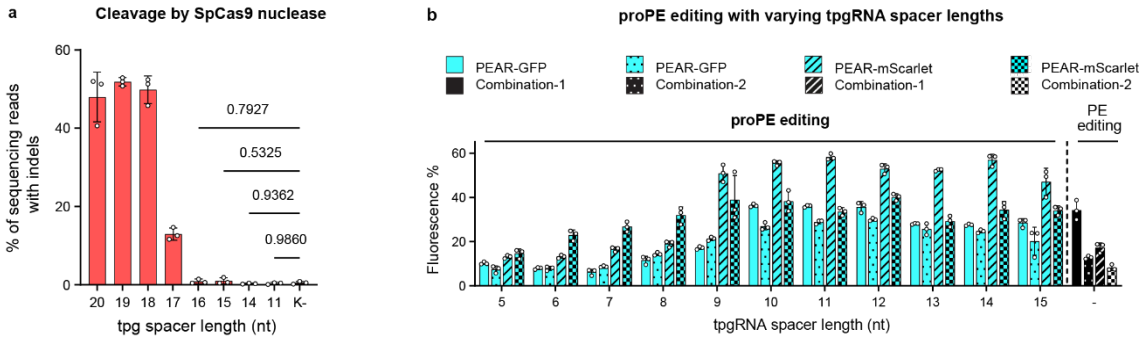

Figure 8.: Characterization of the optimal spacer length for tpgRNA. a, Amplicon sequencing data of cleavage by SpCas9 at the EMX1 site show that spacers shorter than 15 nt do not facilitate cleavage. K- indicates the control with inactive SpCas9. Statistical significance was assessed by a one-way ANOVA and p-values are indicated above the bars. **b,** Fluorescence data from PEAR assay using four different proPE combinations (two PEAR assay based on GFP and two on mScarlet) show that proPE editing is effective with tpgRNA spacer lengths between 10-15 nt. All data points were calculated as the mean of triplicates.

Finally, we examined the optimal amounts of engRNA and tpgRNA in proPE using four different proPE combinations. We decreased the amount of one specific RNA while keeping the other constant and added a non-targeting RNA to maintain the same total DNA amounts for transfection. We found that both too little and excessive amounts of the engRNA of the nicking complex reduced the editing efficiency. While with increasing amount of tpgRNA the editing efficiency reaches saturation (Fig. 9). Based on these results, all subsequent proPE experiments were performed using varying amounts of engRNA, and for the sake of clarity, only the best-performing condition is shown in most cases.

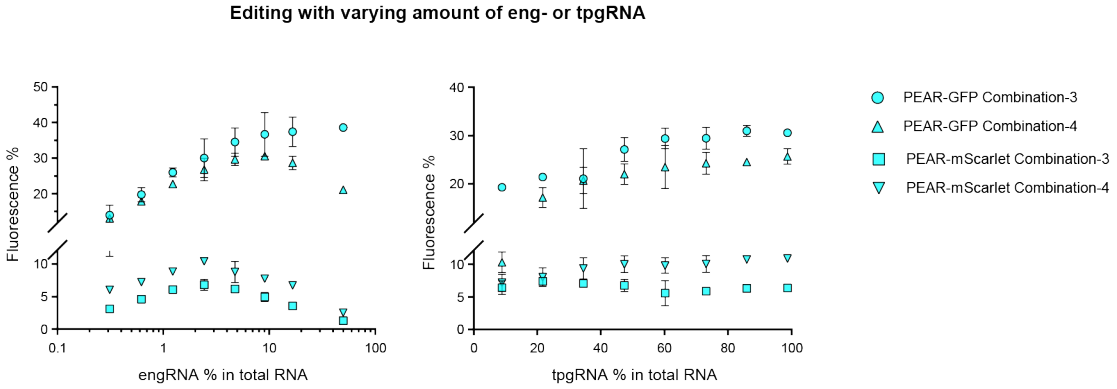

Figure 9.: Both insufficient and excessive amounts of engRNA reduce the editing efficiency, while with increasing the amount of tpgRNA leads to saturation of editing efficiency. Fluorescence data from PEAR assay using four different proPE combinations were used to identify the optimal ranges of engRNA and tpgRNA. All data points were calculated as the mean of triplicates.

4.4 ProPE enhances efficiency on low-performing edits

Since proPE generally improved editing efficiency compared to PE in earlier experiments (Fig. 5,6,7,8b), we next evaluated its performance on 13 frequently used edits from the literature(8,56,62,63). In contrast to the earlier results (Fig. 5,6,7,8b), an overall improvement with proPE was not observed on these 13 edits (Fig. 10a,b). To understand the discrepancy between the previous results (Fig. 5,6,7,8b) and those obtained from these 13 PE and proPE conditions (Fig. 10a,b), we examined the differences between them. We noticed that, while all 13 edits were located close to the target site, mostly within the target sequence, the desired edits in the previous figures (Fig. 5,6,7,8b) were positioned much farther from the target. This suggested that the distance of the edit from the target site might be a critical factor influencing efficiency.

To test this hypothesis, we categorized the edits into three groups based on their distance from the target site: the within-target edit group (positions 1, 2, 3, 5 and 6), the target-proximal group (positions 4, 7, 8, 9, 10), and the target-distal group (positions greater than 10) (Fig. 10c). We then performed an experiment using the same 8 targets (and tpgRNA targets for proPE) with the same PBSes in each group, but changed the editing position, and thus the sequence and the length of the RTT. We observed a marked decline in editing efficiency as the edit was positioned farther from the target site, with 4 out of 8 target-distal edits showing extremely low efficiency (Fig. 10d). Notably, proPE significantly improved the overall editing efficiency in the target-distal group, while also enhancing editing specificity (Fig. 10e,f).

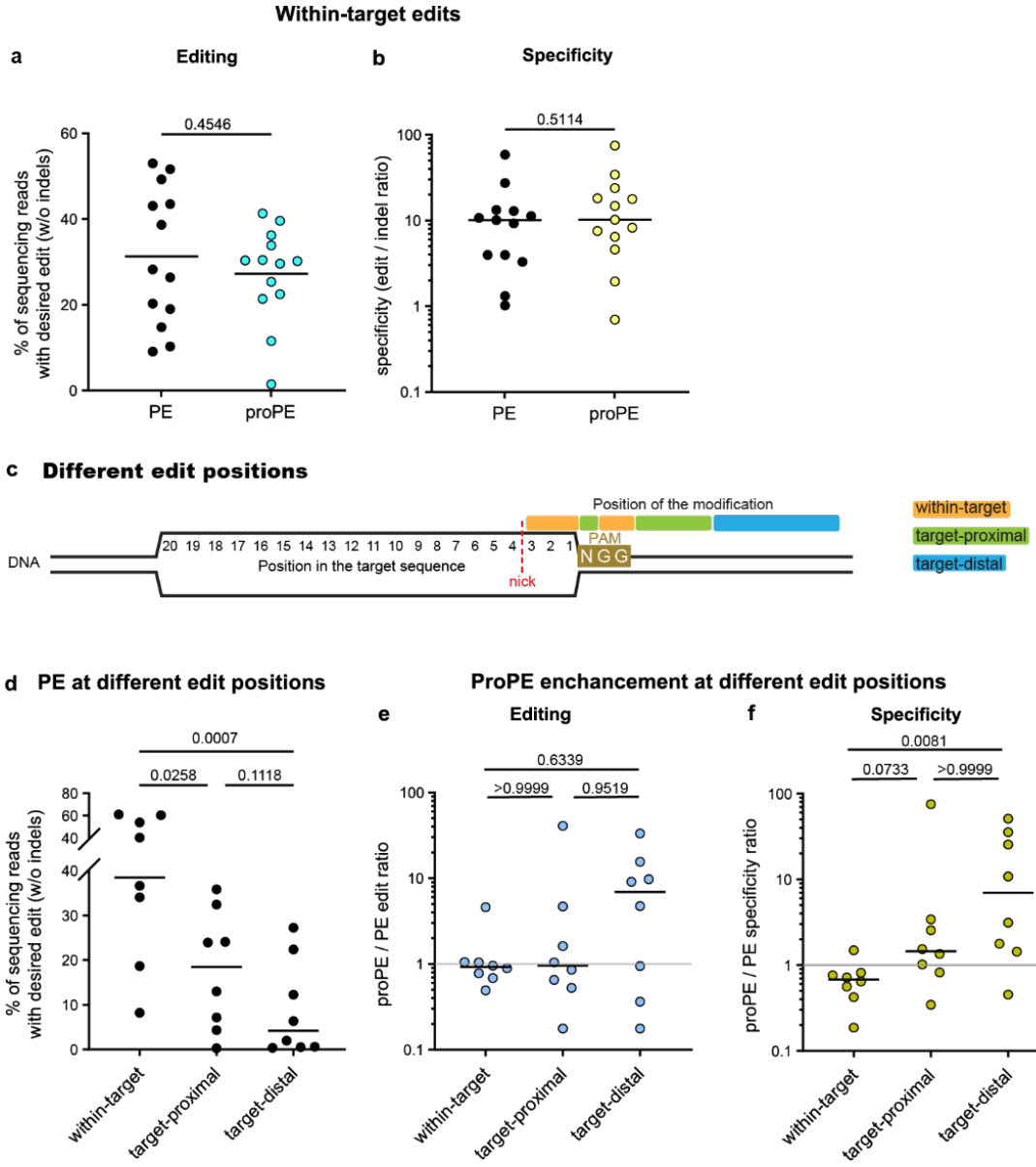

Figure 10.: The position of the edit influence PE editing and proPE improvement. **a,b** Amplicon sequencing data of PE and proPE editing on within-target edits. **c,** Schematic overview of the different edit groups defined by the distance from the nick site. **d,** Amplicon sequencing data of PE showing the decline in PE efficiency as the edit distance from the nick increases. **e,f,** Amplicon sequencing data demonstrating the ability of proPE to enhance both editing efficiency and specificity on target-distal edits. For data sets with normal distribution statistical significance was assessed by a RM one-way ANOVA (**d**) or by a two-tailed unpaired t-test (**a**). For data sets with a non-normal distribution statistical significance was assessed by a two-tailed Mann-Whitney test (**b**) and by a Friedman test (**e,f**). Light grey line indicates the ratio, where the editing

efficiency of PE and proPE is the same (**e,f**). All data were calculated as the mean of triplicates. **a,b,** N=13, **d-f**, N=8. All data points were calculated as the mean of triplicates.

We noticed that proPE tends to enhance the efficiency of introducing edits that conventional PE fails to introduce effectively. To support this observation, we evaluated proPE on 76 PE combinations, out of 130 total, in which conventional PE achieved less than 5% editing efficiency. In these cases, proPE substantially increased the overall efficiency, by 6.2-fold, up to 29%, while also improving the specificity (Fig. 11a,b).

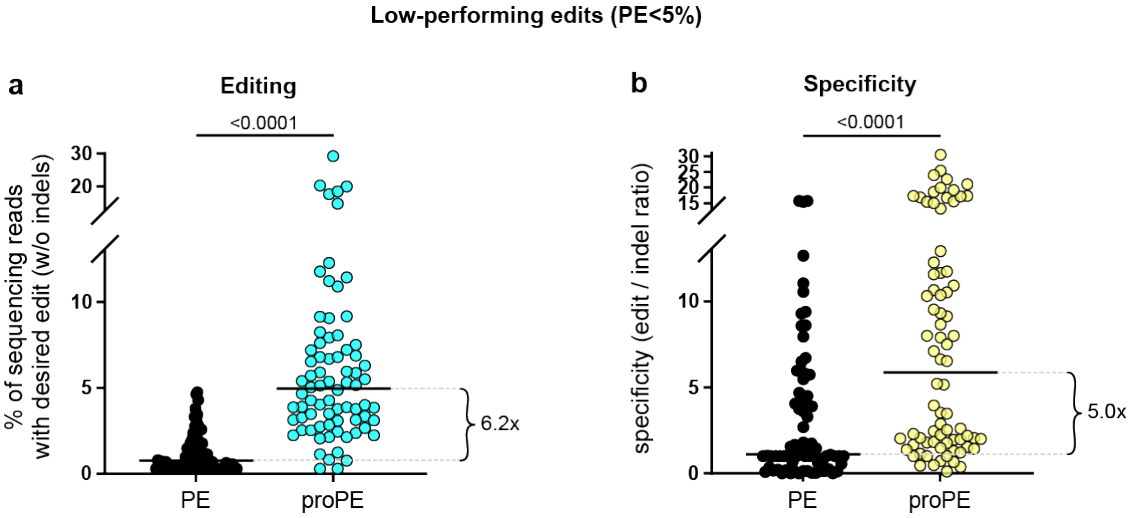

Figure 11.: ProPE enhances editing efficiency and specificity of introducing low-performing edits. a, Amplicon sequencing results of PE and proPE editing introducing several target-distal edits, showing that proPE increases editing efficiency specifically for edits where conventional PE fails to achieve high efficiency. **b,c,** Amplicon sequencing data demonstrates that proPE improves both the overall editing efficiency and specificity across 76 low-performing edits. Medians are shown, statistical significance was assessed by a two-tailed Mann-Whitney test, p-values and fold change between proPE and PE are indicated on the top and next to the dots, respectively, N=76. All data points were calculated as the mean of triplicates.

4.5 ProPE could reduce the off-target effect of editing

ProPE has the potential to reduce off-target effects compared to conventional PE, as it relies on two distinct target sites instead of one. To assess this, we examined off-

target editings while introducing a G-to-T substitution at position 2 of the HEK4 site. ProPE exhibited increased specificity, reducing off-target editing (Fig. 12a) as well as both on-target and off-target indels (Fig. 12b).

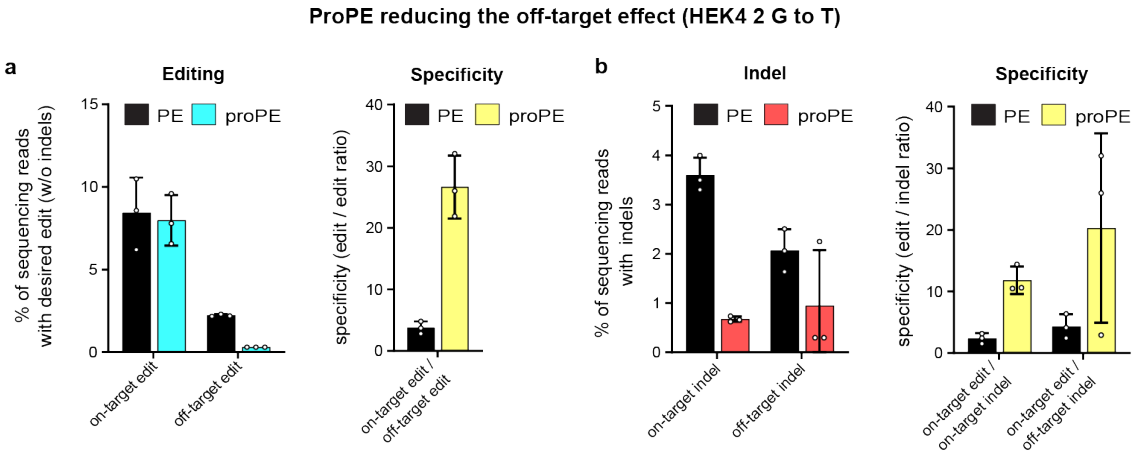

Figure 12.: ProPE could reduce off-target effect of prime editing. The installation of a G-to-T substitution at position 2 of the HEK4 site demonstrated comparable on-target editing activity between proPE and conventional PE. However, proPE exhibited reduced off-target editing (**a**), lower indel formation at off-target site and also at the on-target site (**b**), therefore improving specificity of the editing. All data points were calculated as the mean of triplicates.

4.6 Comparison of proPE with other split pegRNA systems

While we were developing proPE, two related methods, sPE-petRNA(51) and SnPE(64), were published. Both approaches separate the RTT-PBS sequence from the targeting spacer. In sPE-petRNA, the reverse transcriptase (RT) is uncoupled from the Cas9 nickase and fused to the MS2 coat protein (MCP), while the circular RNA carrying the RTT-PBS sequence contains an MS2 aptamer that is recognized by MCP. In contrast, SnPE utilizes a tethering approach in which a motif-binding protein, either tdMCP or N22p, is fused to the PE protein. This protein binds to an MS2 or BoxB aptamer, respectively, which is positioned at the 5' end of the linear RTT-PBS sequence. In summary, the two methods differ in whether the RTT-PBS is circularized(51) or kept linear(64) for optimized editing, and whether it is not recruited(51) or directly recruited to the nicking complex(64).

When comparing proPE to these other split systems on target-proximal and target-distal edits, proPE exhibited substantially higher editing efficiency than both sPE and SnPE (Fig. 13).

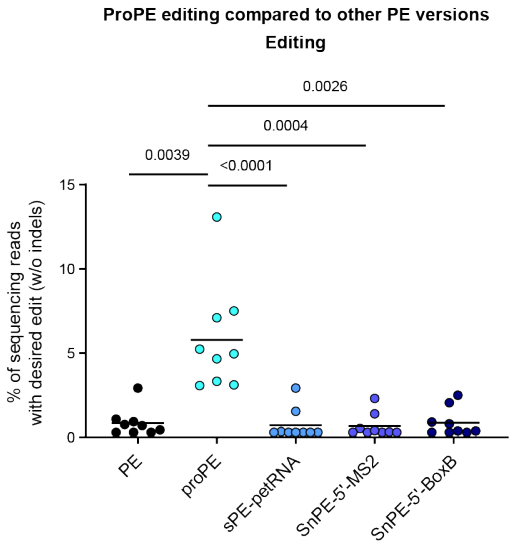

Figure 13.: ProPE exhibits higher editing efficiency than both sPE with petRNA and SnPE. Amplicon sequencing results comparing PE, proPE, sPE-petRNA, SnPE-5'-MS2, SnPE-5'-BoxB. Statistical significance was assessed by a Kruskal-Wallis test. All data points were calculated as the mean of triplicates.

4.7 ProPE is compatible with AAV delivery

To evaluate proPE's potential for therapeutic applications, we tested its compatibility with the AAV delivery system. Based on the study of Davis et al.(59), we designed three AAV vectors to overcome the size limitations of AAVs: the first encoded the N-terminal half of the prime editor, while the second and third carried the C-terminal half along with the tpgRNA, paired either with the engRNA or the second-nicking sgRNA for improved editing(9). Applying AAV-delivered proPE to HEK293 cells resulted in high editing efficiency (Fig. 14).

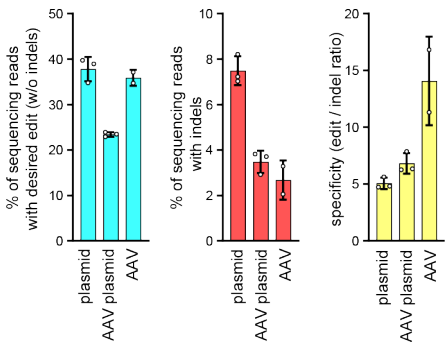

Figure 14.: ProPE is compatible with the AAV-delivery system. Amplicon sequencing results of conventional plasmid transfection, AAV plasmid transfection and AAV transduction. All data points were calculated as the mean of triplicates.

4.8 Overcoming the bottlenecks of prime editing

ProPE can increase editing efficiency for the majority of the corresponding pegRNAs in our experiments with the greatest increase achieved with low-performing-edits, including the target-distal edits. We identified four inhibitory bottlenecks in the prime editing process, which proPE is capable of mitigating.

4.8.1 Bottleneck-a caused by the spacer-PBS intramolecular interaction

PE is generally less efficient when longer primer binding sites (PBSs) is used(9–13), likely due to intramolecular interactions between the PBS and spacer regions of a pegRNA. We demonstrated that proPE successfully mitigates this limitation (Fig. 15) and propose that proPE achieves this by separating the PBS and spacer into distinct molecules, thereby eliminating these inhibitory intramolecular interactions.

To test whether proPE improves editing compared to PE where stronger intramolecular interactions are expected within the pegRNA, we tested the same six pegRNAs and tpgRNAs with both shorter and longer PBSs. We found that extending the PBS generally has a negative impact on PE, an effect not observed with proPE.

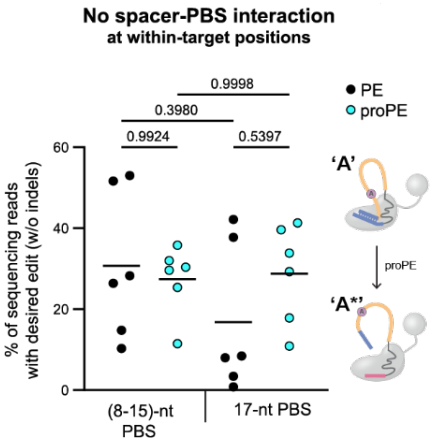

Figure 15.: ProPE overcomes the inhibition caused by the intramolecular interactions between the PBS and the spacer in PE. Amplicon sequencing results of editing with pegRNA and corresponding tpgRNAs with shorter and longer PBSs. Statistical significance was assessed by a one-way ANOVA as the data set followed normal distribution. All data points were calculated as the mean of triplicates.

4.8.2 Bottleneck-b caused by the 3’ degraded pegRNAs

The 3’ extension in pegRNAs is frequently degraded during prime editing, and the resulting 3'-degraded pegRNAs inhibit PE efficiency(21). We demonstrate that proPE successfully mitigates this limitation. We propose that a key factor enabling proPE to

overcome this limitation is the shorter dwell time of tpgRNAs on the target compared to pegRNAs.

The degradation of the 3' extensions of pegRNAs during prime editing has been demonstrated using northern blot and RT-qPCR analyses(21). Nelson et al.(21) demonstrated that such degraded pegRNAs can reduce editing efficiency by using degradation-mimicking pegRNA. We used the same approach to compare their effect on proPE and conventional PE (Fig. 16). Three types of tpg/pegRNAs were used: (1) *intact tpg/pegRNAs* with all necessary elements for editing, (2) *degraded tpg/pegRNAs* lacking the RTT and PBS to mimic natural degradation, and (3) *non-targeting degraded tpg/pegRNAs* as controls. To test inhibition, we gradually replaced *intact tpg/pegRNAs* with *degraded ones*. In separate control experiments, we used *non-targeting degraded tpg/pegRNA* (which do not compete for target binding) instead of the *degraded ones* to account for the reduced amount of intact tpg/pegRNA in transfection. In Figure 16, inhibitory effects that are normalized to the corresponding non-targeting degraded results are shown. The inhibitory effect of degraded pegRNAs on PE efficiency is apparent, while it is significantly less pronounced in the case of proPE.

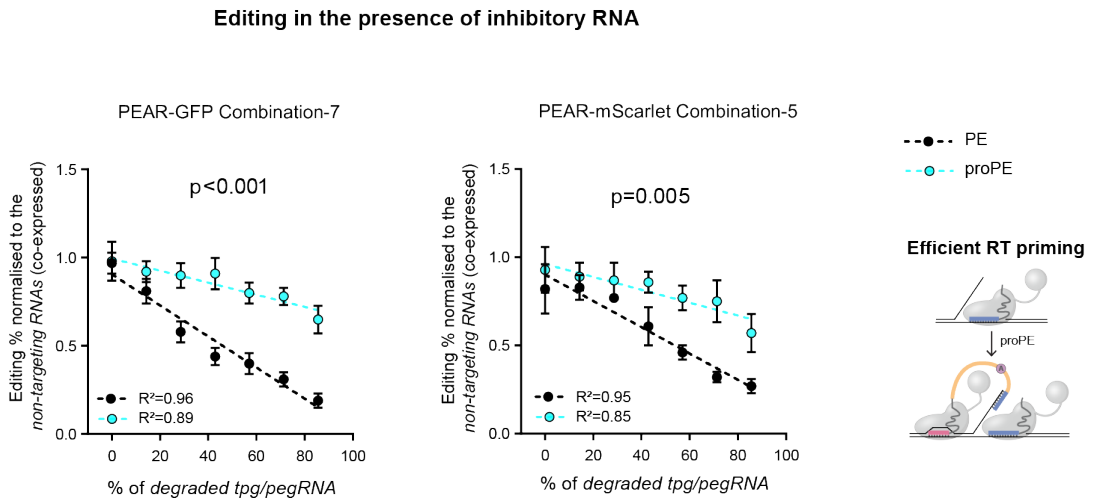

Figure 16.: ProPE mitigates the inhibition caused by 3' degraded pegRNAs in PE. PE or proPE editing was assessed in the presence of increasing amount of *degraded RNAs* targeting the same site. The editing values are normalized to the corresponding values from experiments where *non-targeting degraded RNAs* were co-expressed. The error bars include error propagation. A line is fitted to the data points, statistical significance was

assessed between the slope of the fitted lines by using the nonlinear module of GraphPad. (This figure is primarily the result of Dorottya Simon's contribution.)

We propose that proPE reduces the inhibitory effects seen with 3' degraded pegRNAs, due to the fast dissociation rate of its tpgRNAs complex. Wild-type SpCas9 is known to remain bound to its target DNA for extended periods, forming a stable post-cleavage complex(65). This prolonged binding is also observed(65) for the nickase version of SpCas9, that nicks the non-target strand. In contrast, inactive SpCas9 complexes with 20-nt-long spacers do not maintain this stable post-cleavage conformation and dissociates from the DNA much more quickly(65). Furthermore, it has been reported that the dissociation rate of inactive Cas9 complexes is even faster when shorter spacers are used(66), although the magnitude of this difference is smaller compared to the difference between the dissociation rates of active and inactive complexes(65,66). As a result, tpgRNAs (which are in inactive complexes with shorter spacers) are expected to dissociate more quickly than pegRNAs. Thus, an inhibitory, degraded tpgRNA is more likely to be replaced by an intact tpgRNA during the time window when the nicked DNA strand is available, which may explain their reduced inhibitory potential.

We aimed to design an experiment to test this hypothesis. Unfortunately, we could not directly compare the exact dwelling time differences between inactive tpgRNA complexes with short spacers and active complexes with 20-nt-long spacers, as using an active tpgRNA complex would introduce confounding factors. Specifically, the active complex would also nick the DNA at the tpgRNA target, potentially affecting mismatch repair and increasing background indel levels, making it difficult to make a reliable comparison. We were still able to test the difference between shorter (12 nt) and longer (20 nt) spacers in inactive complexes, although the difference in dwelling times was expected to be substantially smaller in this model than between PE and proPE.

The same experimental design was used in Figure 17 as in the Nelson article(21) as well as in Figure 16, except the tpgRNAs were complexed with a prime editor protein harboring inactive SpCas9, while the engRNAs were complexed with a nickase SaCas9 prime editor. Two identical proPE conditions were compared, with the only difference that the spacer length of the co-expressed degraded tpgRNAs was either 12 or 20

nucleotides. To test their inhibitory effect, editing was examined where the intact tpgRNAs were replaced by these degraded tpgRNAs with increasing amounts. To account for the effect of the reduced amount of intact tpgRNA, increasing amount of non-targeting degraded tpgRNAs, that do not compete for target binding, were added in control experiments. In Figure 17, the inhibitory effects (of degraded RNAs) are shown normalized to the corresponding non-targeting degraded results. The inhibitory effect of degraded tpgRNAs with 12-nt-long spacers is significantly lower than with 20-nt-long spacers.

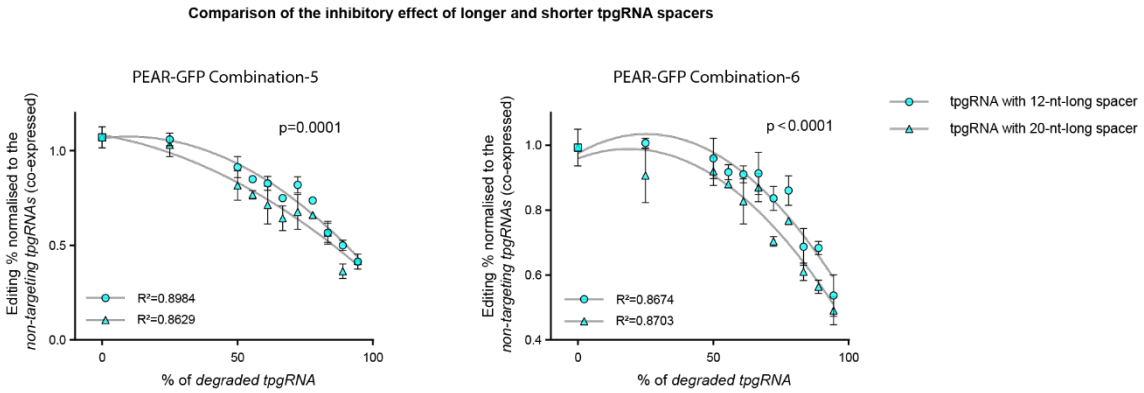

Figure 17.: The inhibitory effect of degraded tpgRNAs with 12-nt-long spacers is significantly lower than with 20-nt-long spacers. ProPE editing was assessed in the presence of increasing amount of *degraded tpgRNAs* targeting the same site. The editing values are normalized to the corresponding values from experiments where *non-targeting degraded RNAs* were co-expressed. ProPE editing in the presence of co-expressed *degraded tpgRNAs* was compared using different spacer lengths (12 nt and 20 nt). Data (mean ± s.d) were calculated from triplicates (n=3) of the targeting samples, each normalized to the average value of the triplicates of the corresponding non-targeting samples (with non-targeting degraded tpgRNA). Statistical significance was calculated between the parameters of the fitted curves. Centered second order polynomial (quadratic) model was used for the fitting in GraphPad. This model provides better fitting to these points than the linear model, of which parameters also resulted in significant differences between 12- and 20-nt-long spacers.

4.8.3 Bottleneck-c caused by the 3’ truncated new DNA strand

The 3’ truncation of the newly transcribed flap sequence reduces PE efficiency(29). We demonstrate that proPE successfully mitigates this limitation and propose that a key factor enabling proPE to overcome this limitation is the previously introduces shorter dwell time of tpgRNAs on the target compared to pegRNAs, allowing for faster turnover on the target site.

We investigated whether a degraded flap could be more efficiently re-elongated by proPE than by PE, by designing an experiment where double edit could only occur if a short flap is successfully prolonged. We maintained it by using a tpg/pegRNA which installs the first edit simultaneously with a tpg/pegRNA which harbors a 3’ extension with the 2nd edit without the PBS (PBS-less), therefore this tpg/pegRNA cannot install the edit on its own. However, when the flap containing the first edit is prolonged by the PBS-less tpg/pegRNA, it can result in double editing (Fig. 18a). We observed a higher amount of double editing in the case of proPE, while all corresponding positive controls showed lower or the same activity as with PE (Fig. 18b). This confirms that the differences observed here are not due to the higher activity of proPE on these edits.

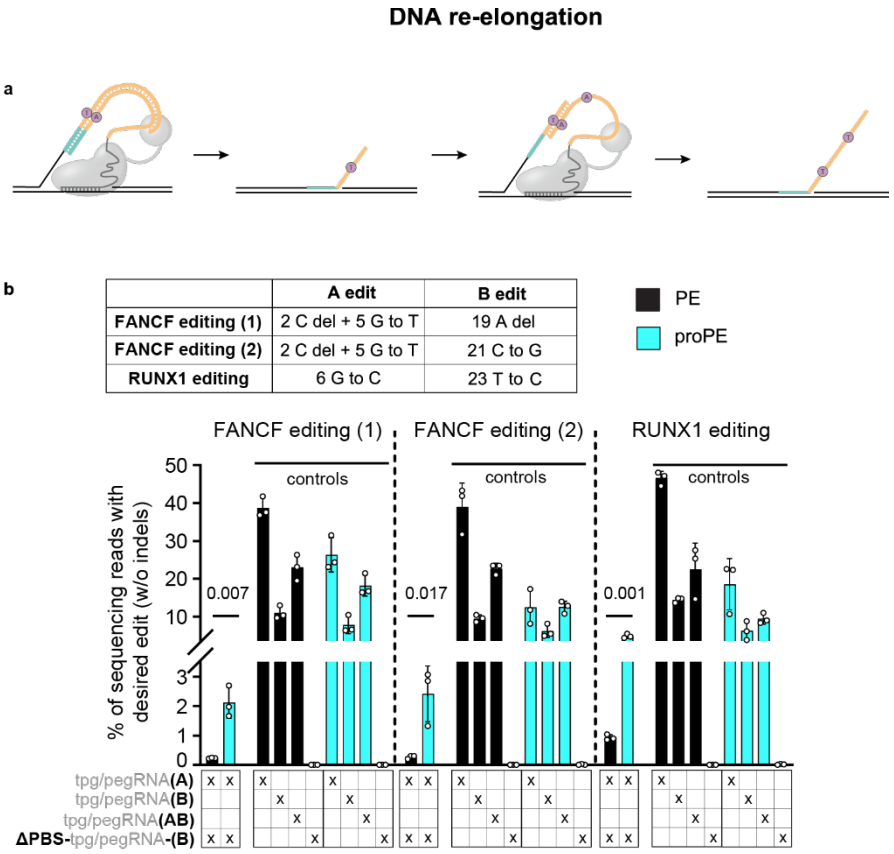

Figure 18.: ProPE is more effective than PE at elongating truncated flaps. a, Schematic representation of the experimental design to examine truncated flap elongation. Double editing could occur only if the short flap (generated by a tpg/pegRNA harboring a within-target edit) is re-elongated by a tpg/pegRNA containing the second edit but lacking the PBS sequence. **b**, The graph presents three separate examples demonstrating that proPE exhibits a significantly higher propensity for re-elongation. Data (mean ± s.d.) were calculated from triplicates. Statistical significance was assessed by a multiple unpaired t-test.

4.8.4 Bottleneck-d caused by re-nicking

Re-nicking of the target sequence by the editor limit the efficiency of PE. We demonstrate that proPE successfully mitigates this limitation and propose that proPE is less sensitive to these limitations because its nicking activity can be adjusted independently of the templating function.

When the edit is not within the target sequence, the target site remains accessible in the edited DNA and can be re-nicked. Re-nicking the already edited DNA can lead to indel formation and reduce the amount of the clear edits (edits without indels). We show that reducing the amount of engRNA by more than an order of magnitude can increase editing efficiency (Fig. 9, 19a) and concomitant reduction in indel formation could occur (Fig. 19a). These findings indicate that an optimal level of nicking activity exists, and that excessive nicking can limit PE efficiency. We also present examples of individual pegRNAs where proPE substantially increased editing over PE when lower engRNA plasmid amount was used (Fig. 19b).

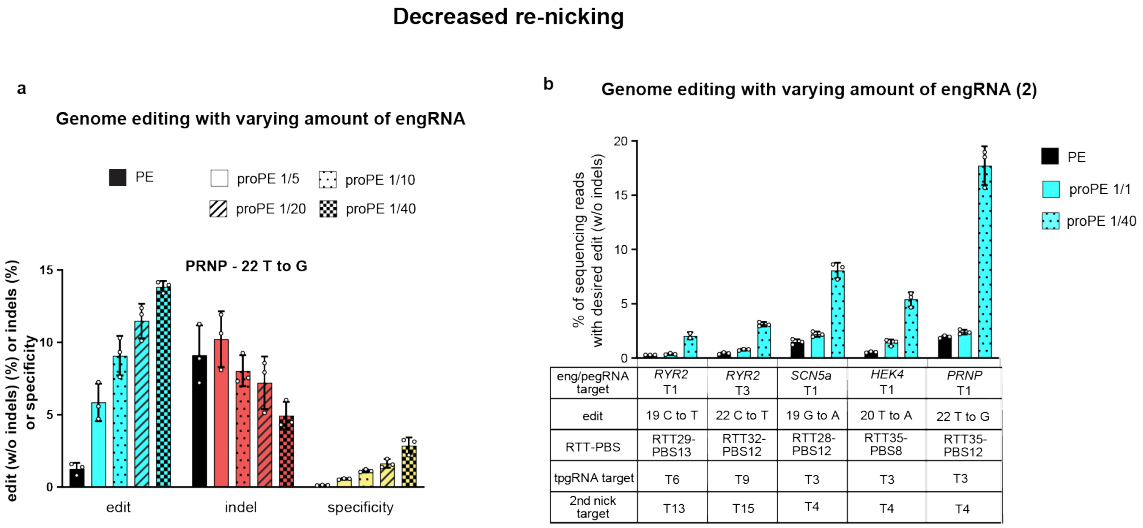

Figure 19.: ProPE could reduce the inhibitory effect of re-nicking. Reducing the amount of engRNA can increase editing efficiency (**a**) and decrease the indel background (**b**). Amplicon sequencing results of introducing edit(s) into the genome using PE and proPE with varying amounts of engRNA. The specific engRNA amounts used are described in the Methods section. All data points were calculated as the mean of triplicates.

5. Discussion

Prime editing has revolutionized the field of genome engineering; however, its inconsistent efficiency remains a major challenge for its broader application. Many efforts have been made to increase the usability of prime editing, but each method comes with its own limitations, and further improvements are still needed.

Significant advances have been made, including the protection of the pegRNA's 3' end, the use of paired pegRNAs to synthesize complementary DNA strands including the edit, introducing a second nicking sgRNA to bias mismatch repair, or co-expressing MLH1dn to transiently suppress mismatch repair. Co-editing the PAM site can also improve outcomes by resulting in edits that are no longer substrates of mismatch repair and by reducing re-nicking.

Despite these advances, limitations persist. Complete protection of the pegRNA's 3' extension has not been fully achieved. A new approach applying a fusion prime editor provide enhanced 3' end protection but exceed the packaging capacity of adeno-associated viruses (AAVs). Additionally, using a second nicking sgRNA often leads to a substantial increase in indel formation. Suppressing mismatch repair via MLH1dn raises concerns, as it alters the genome-wide mismatch landscape, affects microsatellites, and the mechanism for resolving mismatch-caused heteroduplexes in MMR-deficient cells remains unknown(39). Moreover, due to its size, it is also incompatible with AAV design. Co-editing the PAM is a viable option to reduce re-nicking, though an additional edit may influence with the pegRNA folding, potentially influencing editing outcomes. Importantly, PAM editing is not always feasible due to sequence constraints at the site where the PAM is located.

The development of proPE offers an alternative or complementary solution to overcome many of these bottlenecks. ProPE could mitigate the inhibitory effects of PBS–spacer intramolecular interaction, provide a solution against the inhibitory effects of 3' RNA degradation, and reduces indel formation by decreasing the frequency of re-nicking. Notably, it uniquely addresses the inhibition caused by 3'-truncated DNA flaps, thanks to its unique templating properties, which enable the complex carrying the template not to re-bind directly to the nick and facilitate a more dynamic template exchange, an issue that no other method currently claims to resolve.

Although we have highlighted several critical bottlenecks and proposed corresponding corrective mechanisms by proPE, we do not exclude the possibility that additional factors and mechanisms may also contribute to the improved performance of proPE.

An important finding of this study shows that the position of the intended edit has a striking impact on editing efficiency, and we were surprised that it has not been apparent earlier. However, a brief review of several prominent prime editing papers may provide an explanation as they all focus on edits within a narrow window (positions: 1-10 nt) (Table 2.).

Table 2.: Several high-profile studies developing enhancement methods for prime editing focus on edits within a specific window.

Publication	All samples for edits in positions **1-10 nt**	All samples for edits in positions **>10 nt**
Anzalone et al., Nature(9)	~410	9
Chen et al., Cell.(67)	~2000	0
Nelson et al., Nature Biotechnology(68)	~1000	0
Koeppel et al., Nature Biotechnology(29)	~3604	0
Lin et al., Nature Biotechnology(69)	~101	0
Petri et al., Nature Biotechnology(12)	~54	0
Lin et al., Nature Biotechnology(10)	~184	0

The identified inhibitory effects can be classified based on their impact on different edit groups. We propose that the PBS-spacer intramolecular interaction (bottleneck-a) and the presence of 3’ degraded pegRNAs (bottleneck-b) influence editing outcomes across all edit positions. In contrast, re-nicking (bottleneck-d) affects edits at target-proximal and target-distal positions, as the target can be re-nicked by the PE complex if the target remains unmodified after editing. Presence of truncated DNA flaps (bottleneck-c) and its correction may be especially relevant for target-distal edits, where both the inhibitory effects and also the corrective mechanism of proPE are likely to be more pronounced. This can be attributed to the observation that 3’ DNA flap degradation

is more pronounced for longer flaps(29) and longer flaps are most likely to retain an extended priming sequence, enabling proPE to initiate re-elongation. Therefore, the combined influence of these bottlenecks is likely greatest for target-distal edits, where proPE also demonstrates the highest potential for correction (Fig. 10e,f). This is especially important, as a large proportion of human pathogenic SNPs fall into the target-distal category (Fig. 20). To support the potential future therapeutic applicability of proPE, we also demonstrated its compatibility with AAV-based delivery systems (Fig. 14) and its reduced off-target effect (Fig. 12). Although the development of active, PAM-flexible prime editors may reduce the need for targeting distal edits by enabling more SNPs to be addressed as within-target edits, their editing efficiency remains quite low.

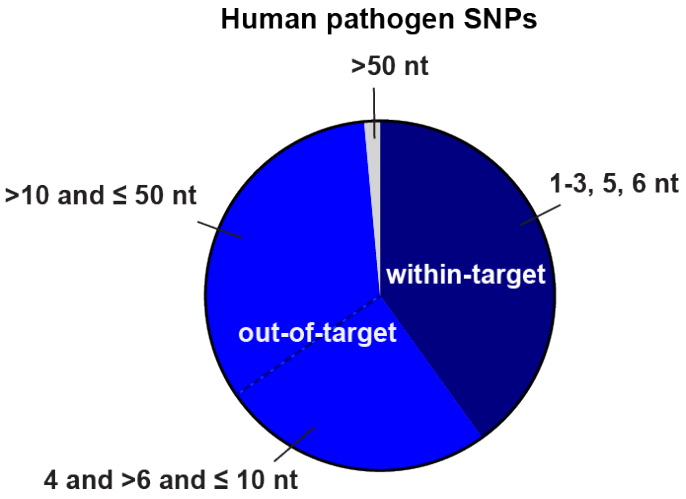

Figure 20: Distribution of the human pathogenic SNPs (total: 75259) based on their distance from the nearest targe site, showing that a large portion of these SNPs are classified as target-distal edits (positions >10).

Comparison with other split prime editor systems demonstrated that proPE achieves higher editing efficiency than similar methods. We attribute this improvement to targeting the template in close proximity to the nick, instead of no targeting or targeting to the nicking complex. Different targeting strategies may influence both the initiation of prime editing and the effectiveness of corrective mechanisms ‘b’ and ‘c’, which likely contribute to proPE’s enhanced performance. Overcoming the PBS-spacer intramolecular interaction (bottleneck-a) is inherently achieved by these similar methods, whereas re-nicking (bottleneck-d) can be addressed by all of these systems through the application of a reduced amount of nicking sgRNA, as demonstrated during the optimization of proPE. Overall, we believe proPE outperforms other split systems due to its unique targeting approach.

As the field of prime editing is still in its early stages, additional bottlenecks will likely be identified, and further solutions developed. We believe proPE is a valuable advancement in the prime editing toolkit, addressing multiple bottlenecks and offering promising potential for therapeutic applications.

6. Conclusions

First, we developed a Python script for the bioinformatic analysis of Illumina NGS data from prime editing experiments, enabling accurate quantification of editing efficiency and detailed characterization of associated indel profiles. Building on this, we established an easy-to-use, plasmid-based assay for the systematic evaluation of prime editing outcomes.

Second, we developed a novel prime editing system, proPE, which overcomes the limitations imposed by spacer–PBS interactions by delivering the spacer and PBS on two separate CRISPR complexes. The complex with the essential nicking gRNA (engRNA) nicks the target site recognized by its spacer sequence, while the complex with the template providing gRNA (tpgRNA) provides the primer and template for reverse transcription. Our results indicate that both the engRNA and tpgRNA are required for proPE editing, as no editing was observed in the absence of either RNA. We also systematically characterized its working parameters in respect to the position of the two target sites, the spacer length of the tpgRNA and the amount of the engRNA. Additionally, we found that the PAM-flexible SpRY prime editor is not suitable for efficient editing in either the PE or proPE format.

Third, we explored the therapeutic potential of proPE by identifying editing contexts in which it significantly improves both efficiency and specificity compared to conventional prime editing. ProPE is more likely to improve the editing efficiency and specificity on target-distal edits and in cases where conventional PE fails to reach high efficiency. We further demonstrated its compatibility with AAV-based delivery systems and its potential to reduce off-target effects due to its design employing two distinct target sites. We observed that other split pegRNA approaches did not exhibit the same therapeutic advantages as proPE.

Finally, we investigated the mechanisms underlying proPE's improved performance and identified key factors contributing to its enhanced editing efficiency and specificity. These include overcoming spacer–PBS interactions, reducing the inhibitory effects of 3′-degraded pegRNAs, re-elongating the 3′-truncated DNA flaps that arise during prime editing, and decreasing harmful DNA re-nicking.

Together, these achievements establish proPE as a robust and versatile prime editing tool, with broad potential for both fundamental research and therapeutic genome editing applications.

7. Summary

Prime editing (PE) is a versatile and precise genome editing technique; however, its broad application has been limited by variable efficiency and specificity, often requiring extensive optimization for each target. To overcome some of these limitations, we developed proPE (prime editing with prolonged editing window), a PE tool that employs two sgRNA; one to introduce the nick and another non-cleaving sgRNA to position the reverse transcriptase template in close vicinity of the nick.

ProPE demonstrates significantly improved performance, particularly in editing contexts where standard PE is inefficient. For 76 editings where PE achieves less than 5% efficiency, proPE enhances efficiency up to 29.3% and improves both efficiency and specificity by 6.0 and 3.8-fold, respectively. By addressing key limitations of PE, proPE expands the range of targetable edits, including many pathogenic SNPs.

By extending the editing window and reducing the off-target effect, proPE offers a powerful tool for therapeutic applications.

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.

2. Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol. 2020 Jul 1;38(7):824–44.

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

4. Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol. 1994 Dec;14(12):8096–106.

5. Klermund J, Rhiel M, Kocher T, Chmielewski KO, Bischof J, Andrieux G, et al. On- and off-target effects of paired CRISPR-Cas nickase in primary human cells. Mol Ther. 2024 May;32(5):1298–310.

6. Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat Methods. 2014 Apr;11(4):399–402.

7. 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 19;533(7603):420–4.

8. 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 5;576(7785):149–57.

9. 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 5;576(7785):149–57.

10. 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.

11. 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 Res. 2023 May 29;gkad456.

12. 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.

13. 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;186(10):2256-2272.e23.

14. Hwang GH, Jeong YK, Habib O, Hong SA, Lim K, Kim JS, et al. PE-Designer and PE-Analyzer: web-based design and analysis tools for CRISPR prime editing. Nucleic Acids Res. 2021 Jul 2;49(W1):W499–504.

15. Hsu JY, Grünewald J, Szalay R, Shih J, Anzalone AV, Lam KC, et al. PrimeDesign software for rapid and simplified design of prime editing guide RNAs. Nat Commun. 2021 Feb 15;12(1):1034.

16. 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.

17. 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;186(10):2256-2272.e23.

18. 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 [Internet]. 2023 Jan 16 [cited 2023 Apr 6]; Available from: https://www.nature.com/articles/s41587-022-01613-7

19. 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.

20. Zhang W, Petri K, Ma J, Lee H, Tsai CL, Joung JK, et al. Enhancing CRISPR prime editing by reducing misfolded pegRNA interactions. eLife. 2024 Jun 6;12:RP90948.

21. 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.

22. 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.

23. Li X, Wang X, Sun W, Huang S, Zhong M, Yao Y, et al. Enhancing prime editing efficiency by modified pegRNA with RNA G-quadruplexes. Li J, Li J, editors. J Mol Cell Biol. 2022 Aug 2;14(4):mjac022.

24. 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.

25. 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 18;628(8008):639–47.

26. Liu Y, Kao HI, Bambara RA. Flap Endonuclease 1: A Central Component of DNA Metabolism. Annu Rev Biochem. 2004 Jun;73(1):589–615.

27. Suck D. DNA recognition by structure-selective nucleases. Biopolymers. 1997;44(4):405–21.

28. Perrino FW, Harvey S, McMillin S, Hollis T. The Human TREX2 3′ → 5′-Exonuclease Structure Suggests a Mechanism for Efficient Nonprocessive DNA Catalysis. J Biol Chem. 2005 Apr;280(15):15212–8.

29. Koeppel J, Weller J, Peets EM, Pallaseni A, Kuzmin I, Raudvere U, et al. Prediction of prime editing insertion efficiencies using sequence features and DNA repair determinants. Nat Biotechnol [Internet]. 2023 Feb 16 [cited 2023 Mar 18]; Available from: https://www.nature.com/articles/s41587-023-01678-y

30. 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.

31. Zhuang Y, Liu J, Wu H, Zhu Q, Yan Y, Meng H, et al. Increasing the efficiency and precision of prime editing with guide RNA pairs. Nat Chem Biol. 2022 Jan;18(1):29–37.

32. Kweon J, Hwang HY, Ryu H, Jang AH, Kim D, Kim Y. Targeted genomic translocations and inversions generated using a paired prime editing strategy. Mol Ther. 2023 Jan;31(1):249–59.

33. Wang J, He Z, Wang G, Zhang R, Duan J, Gao P, et al. Efficient targeted insertion of large DNA fragments without DNA donors. Nat Methods. 2022 Mar;19(3):331–40.

34. Tao R, Wang Y, Jiao Y, Hu Y, Li L, Jiang L, et al. Bi-PE: bi-directional priming improves CRISPR/Cas9 prime editing in mammalian cells. Nucleic Acids Res. 2022 Jun 24;50(11):6423–34.

35. Jiang T, Zhang XO, Weng Z, Xue W. Deletion and replacement of long genomic sequences using prime editing. Nat Biotechnol. 2022 Feb;40(2):227–34.

36. Choi J, Chen W, Suiter CC, Lee C, Chardon FM, Yang W, et al. Precise genomic deletions using paired prime editing. Nat Biotechnol. 2022 Feb;40(2):218–26.

37. Anzalone AV, Gao XD, Podracky CJ, Nelson AT, Koblan LW, Raguram A, et al. Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat Biotechnol. 2022 May;40(5):731–40.

38. Tao R, Wang Y, Hu Y, Jiao Y, Zhou L, Jiang L, et al. WT-PE: Prime editing with nuclease wild-type Cas9 enables versatile large-scale genome editing. Signal Transduct Target Ther. 2022 Apr 20;7(1):108.

39. 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;184(22):5635-5652.e29.

40. 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.

41. 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.

42. Zou RS, Marin-Gonzalez A, Liu Y, Liu HB, Shen L, Dveirin RK, et al. Massively parallel genomic perturbations with multi-target CRISPR interrogates Cas9 activity and DNA repair at endogenous sites. Nat Cell Biol. 2022 Sep;24(9):1433–44.

43. Zou RS, Liu Y, Wu B, Ha T. Cas9 deactivation with photocleavable guide RNAs. Mol Cell. 2021 Apr;81(7):1553-1565.e8.

44. Vítor AC, Huertas P, Legube G, De Almeida SF. Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox. Front Mol Biosci. 2020 Feb 21;7:24.

45. 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 [Internet]. 2024 Jun 21 [cited 2024 Dec 17]; Available from: https://www.nature.com/articles/s41587-024-02268-2

46. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul;523(7561):481–5.

47. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature. 2018 Apr;556(7699):57–63.

48. Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science. 2018 Sep 21;361(6408):1259–62.

49. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science. 2020 Apr 17;368(6488):290–6.

50. Böck D, Rothgangl T, Villiger L, Schmidheini L, Matsushita M, Mathis N, et al. In vivo prime editing of a metabolic liver disease in mice. Sci Transl Med. 2022 Mar 16;14(636):eabl9238.

51. 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.

52. Grünewald J, Miller BR, Szalay RN, Cabeceiras PK, Woodilla CJ, Holtz EJB, et al. Engineered CRISPR prime editors with compact, untethered reverse transcriptases. Nat Biotechnol. 2023 Mar;41(3):337–43.

53. Zheng C, Liang SQ, Liu B, Liu P, Kwan SY, Wolfe SA, et al. A flexible split prime editor using truncated reverse transcriptase improves dual-AAV delivery in mouse liver. Mol Ther. 2022 Mar;30(3):1343–51.

54. Gao Z, Ravendran S, Mikkelsen NS, Haldrup J, Cai H, Ding X, et al. A truncated reverse transcriptase enhances prime editing by split AAV vectors. Mol Ther. 2022 Sep;30(9):2942–51.

55. Zong Y, Liu Y, Xue C, Li B, Li X, Wang Y, et al. An engineered prime editor with enhanced editing efficiency in plants. Nat Biotechnol. 2022 Sep;40(9):1394–402.

56. 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. eLife. 2022 Feb 23;11:e69504.

57. 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.

58. Liu P, Liang SQ, Zheng C, Mintzer E, Zhao YG, Ponnienselvan K, et al. Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. Nat Commun. 2021 Apr 9;12(1):2121.

59. Davis JR, Banskota S, Levy JM, Newby GA, Wang X, Anzalone AV, et al. Efficient prime editing in mouse brain, liver and heart with dual AAVs. Nat Biotechnol [Internet]. 2023 May 4 [cited 2023 Jun 7]; Available from: https://www.nature.com/articles/s41587-023-01758-z

60. 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. 2021 Aug 26;6(2):181–94.

61. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014 Mar;32(3):279–84.

62. 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.

63. 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;184(22):5635-5652.e29.

64. Feng Y, Liu S, Mo Q, Xiao X, Liu P, Ma H. Enhancing Prime Editing Efficiency and Flexibility with Tethered and Split pegRNAs. Protein Cell. 2022 Jul 15;pwac014.

65. Aldag P, Welzel F, Jakob L, Schmidbauer A, Rutkauskas M, Fettes F, et al. Probing the stability of the SpCas9–DNA complex after cleavage. Nucleic Acids Res. 2021 Dec 2;49(21):12411–21.

66. Josephs EA, Kocak DD, Fitzgibbon CJ, McMenemy J, Gersbach CA, Marszalek PE. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. Nucleic Acids Res. 2015 Oct 15;43(18):8924–41.

67. 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;184(22):5635-5652.e29.

68. 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.

69. Lin Q, Zong Y, Xue C, Wang S, Jin S, Zhu Z, et al. Prime genome editing in rice and wheat. Nat Biotechnol. 2020 May 1;38(5):582–5.

9. Bibliography of the candidate's publications

Peer reviewed publications relevant to the dissertation:

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 10;8(10):1100–16.

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. eLife. 2022 Feb 23;11:e69504.

Other peer reviewed publications:

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 Sept 16;14(1):5746.

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 June 23;51(11):5847–63.

Kulcsár PI, Tálas A, Ligeti Z, Krausz SL, Welker E. SuperFi-Cas9 exhibits remarkable fidelity but severely reduced activity yet works effectively with ABE8e. Nat Commun. 2022 Nov 11;13(1):6858.

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.

Sangeetham SB, Engelke AD, Fodor E, Krausz SL, Tatzelt J, Welker E. The G127V variant of the prion protein interferes with dimer formation in vitro but not in cellulo. Sci Rep. 2021 Feb 4;11(1):3116.

Tóth E, Varga É, Kulcsár PI, Kocsis-Jutka V, Krausz SL, Nyeste A, et al. Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases. Nucleic Acids Res. 2020 Apr 17;48(7):3722–33.

Tóth E, Czene BC, Kulcsár PI, Krausz SL, Tálas A, Nyeste A, et al. Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. Nucleic Acids Res [Internet]. 2018 Sept 20 [cited 2025 Nov 12]; Available from: https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky815/5103950

Tálas A, Kulcsár PI, Weinhardt N, Borsy A, Tóth E, Szebényi K, et al. A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a ‘self-cleaving’ GFP-expression plasmid. DNA Res. 2017 Dec 1;24(6):609–21.

10. Acknowledgements

First, I would like to thank my supervisor, Ervin Welker, for his guidance, which taught me a lot about scientific reasoning, made this project possible, and gave me the freedom to explore and test my own ideas. I am also grateful for his dedication to research and to this work, which led to many late-night discussions that helped us overcome several bottlenecks in the project. Finally, I would like to thank him for his efforts in maintaining the financial background for our research during several challenging periods over the past ten years.

I would like to thank all co-authors of the *proPE* publication, the amazing technical assistant help, and everyone who contributed to this work. In particular, I am very grateful to Dorottya Anna Simon, whose significant contributions improved the quality of the *proPE* publication, and on whom I could always rely.

I would like to thank Eszter Tóth, András Tálas and Péter Kulcsár from whom I learned many skills, from the beginning of my Master's studies until the end of my PhD, including technical skills, experimental design, data evaluation, as well as project organization and self-management.

I would also like to thank Zsombor Welker for deepening my knowledge of Python scripting, which enabled me to write my own scripts for the evaluation of prime editing, GUIDE-seq oligo integration, etc.

To the Welker group, thank you for providing such a supportive, cheerful, and inspiring environment. Thank you for supporting both my work and my well-being.

I am grateful to György Várady and Edit Szabó for their constant help with FACS measurements.

Finally, I would like to thank my husband for his endless emotional support and his patience during these stressful periods. I am also grateful to my mother and father for their support and for resisting the urge to constantly ask, "When are you going to finish your PhD?" I would like to thank my sister for her emotional and scientific support, and my uncle and aunt for their valuable advice and survival tips. I would also like to thank my friends for being there for me in difficult periods, and for accepting that even weekends were many times devoted to *proPE*. Lastly, I thank my grandfather, who played a significant role in inspiring me to pursue a scientific career.