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

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. However, the use of nucleases can lead to chromosomal rearrangements and on-target aberrations, which are detrimental to therapeutic applications. Moreover using a donor template for precise modification through homology-directed repair, can also cause template integration 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. Prime editing represents the most recent and is considered one of the most promising advancement among these nickase-based editing systems. 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. The released 3' DNA end hybridizes to the complementary PBS sequence of the pegRNA and functions as a primer for the reverse transcriptase. The reverse transcriptase then extends the non-target DNA strand along the RTT template, thereby incorporating the desired edit into the DNA. This newly synthesized strand forms a 3′ DNA flap that equilibrates with the original, unedited 5′ flap. The 5′ flap is preferably cleaved by an endogenous flap endonuclease, then the edited strand is

ligated. Finally, the introduced edit may become fixed by mismatch repair or during replication.

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.

2. Objectives

My first objective is to develop a python script for the bioinformatic analysis of Illumina NGS data obtained from prime editing experiments.

My second objective is to develop a prime editing tool, proPE (Prime Editing with Prolonged Editing Window) which overcomes the spacer-PBS intramolecular interaction limiting prime editing activity.

My third objective is to explore its potential for therapeutic applications.

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

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.

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), created by Tálas and colleagues. The linkers coding the targets were cloned into pAT9624 plasmid between Esp3I sites.

The following plasmids were obtained from the non-profit plasmid distribution service Addgene: pCMV-PE2 (#132775) created by Anzalone et al. (the original prime editor), pCMV-PEmax-P2A-hMLH1dn (#174828) created by Chen et al. (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. Several PE-expressing plasmids were constructed using NEB HiFi Assembly.

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.

3.3 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.

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.

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. 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). Reads with the intended modifications were identified by searching for a sequence stretch containing the desired edit, flanked by 5-5 matching nucleotides.

3.6 Statistics

For data sets with a normal distribution statistical significance was assessed by a two-tailed unpaired t-test (two groups) or one-way ANOVA (more groups) or RM one-way ANOVA (more groups, in the case of paired comparison). In cases where the data did not follow a normal distribution, significance was assessed using a two-tailed Mann-Whitney test (two groups) or the Kruskal-Wallis test (more groups) or the Friedman test (more groups, paired comparisons). Statistical tests were performed using GraphPad Prism 9.2.

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.

To analyze prime editing activity at the genomic sites using data obtained from Illumina next-generation sequencing (NGS) in the enrichment experiments, 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). 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.

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.

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). 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.

4.3 Characterization of proPE

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.

4.4 ProPE enhances efficiency on low-performing edits

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). 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. Notably, proPE significantly improved the overall editing efficiency in the target-distal group, while also enhancing editing specificity.

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.

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 T substitution at position 2 in the HEK4 site. ProPE exhibited increased specificity and reduced off-target editing.

4.6 Comparison of proPE with other split pegRNA systems

While we were developing proPE, two related methods, sPE-petRNA and SnPE, were published. Both approaches separate the RTT-PBS sequence from the

targeting spacer. 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.

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. Applying AAV-delivered proPE to HEK293 cells resulted in high editing efficiency.

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; the spacer–PBS intramolecular interactions, the inhibitory effects of 3′-degraded pegRNAs, the presence of the 3′-truncated DNA flaps and DNA re-nicking by the prime editor complex.

5. Conclusions

We developed a Python script for the bioinformatic analysis of Illumina NGS data from prime editing experiments. Building on this, we established an easy-to-use, plasmid-based assay for the systematic evaluation of prime editing outcomes.

We developed and characterized 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.

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. 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.

6. Bibliography of the candidate's publications

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 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.