SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3401.

SIMON DOROTTYA ANNA

Patobiokémia

című program

Programvezető: Dr. Csala Miklós, egyetemi tanár

Témavezető: Dr. Welker Ervin, tudományos tanácsadó

DEVELOPMENT OF METHODS TO FACILITATE THE APPLICATION OF PRIME EDITING

PhD thesis

Dorottya Anna Simon

Semmelweis University Doctoral College

Molecular Medicine Division

Supervisor: Ervin Welker, D.Sc

Official reviewers: Tamás Arányi, Ph.D
Sándor Spisák, Ph.D

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

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

Budapest
2026

Table of Contents

List of Abbreviations

BEAR: base editor activity reporter

Cas: CRISPR-associated

CRISPR: clustered regularly interspaced short palindromic repeats

crRNA: CRISPR-RNA

CYP: cytochrome P450

dCas9: nuclease inactive "dead" Cas9

DSB: double-strand DNA breaks

engRNA: essential nicking guide RNA

HDR: homology directed repair

nCas9: nickase Cas9

NGS: next generation sequencing

NHEJ: non-homologous end joining

PAM: protospacer adjacent motif

PBS: primer binding site

PE: prime editor

PEAR: prime editor activity reporter

pegRNA: prime editing guide RNA

proPE: prime editing with prolonged editing window

RT: reverse transcriptase

RTT: reverse transcription template

sgRNA: single guide RNA

SNP: single nucleotide polymorphism

tpgRNA: template providing guide RNA

tracrRNA: trans-activating CRISPR-RNA

1. Introduction

1.1 The CRISPR/Cas9 system

The story of CRISPR (clustered regularly interspaced short palindromic repeats) / Cas (CRISPR-associated) systems began in 1987 with a report of repetitive sequence arrays in a bacterial genome [1]. In the following decades scientists discovered that they serve as prokaryotic adaptive immune defence mechanisms against viruses and other mobile genetic elements. During the immunisation phase in response to an infection, bacteria integrate short sequences from the invading genome into their own. The name CRISPR itself refers to the fact that these fragments, called spacers, are clustered in the bacterial genome, separated by short, repeated segments. These are transcribed into CRISPR-RNAs (crRNA) that, when complexed with a Cas protein, can recognise their complementary sequence (protospacer) in the corresponding viral genome. In the case of a repeated viral infection, this recognition activates a nuclease involved in the immune response, which cleaves the foreign DNA, thereby eliminating the virus from the host. [2,3].

Besides the spacer–protospacer complementarity, an additional requirement of DNA cleavage is the presence of a short sequence called PAM (protospacer adjacent motif) next to the protospacer. PAMs vary in length and sequence depending on the host species and allow the differentiation of self and foreign DNA so that they are not found in the CRISPR array of the bacterial genome. Thus, complementarity of the spacer with its own coding region does not lead to cleavage [4].

Successive discoveries have enabled the use of CRISPR systems for genetic engineering. To date, the most widely used Cas endonuclease is the *Streptococcus pyogenes* Cas9. In 2011 the research group of Emmanuelle Charpentier discovered the existence of *trans*-activating CRISPR-RNA (tracrRNA) and that it directs the maturation of crRNAs [5]. Further research in collaboration with Jennifer Doudna showed that tracrRNA is also essential for the cleavage activity of Cas9 which forms a ribonucleoprotein with the tracrRNA:crRNA duplex. They also showed that by replacing the spacer sequence of the crRNA, Cas9 can be effectively re-targeted to sequence-specifically cleave double-stranded DNA targets *in vitro*. To further simplify the use of

Cas9 they also demonstrated that the tracrRNA:crRNA duplex can be engineered to a single guide RNA (sgRNA). The sgRNA is functionally divided into two parts: the first 20 nucleotides from the 5' end encode the spacer sequence, and the rest forms the scaffold with which the Cas9 protein interacts [6].

In 2013 several studies reported that the Cas9-sgRNA system is also functional in eukaryotic cells causing genetic modifications at the targeted genomic loci [7–9]. Subsequent structural research has revealed Cas9's mechanism of action: First, the enzyme scans the DNA for PAMs. If one is found, subsequent binding separates the DNA strands in the 5' direction. If the spacer and the corresponding strand of the DNA sequence are complementary, a spacer–protospacer heteroduplex is formed. This leads to structural rearrangements that activate the two nuclease domains of Cas9 (RuvC and HNH) each of which cleaves one strand of the DNA three base pairs away from the PAM, resulting in a blunt-ended DNA break [10,11] (Figure 1a). The widespread use of *S. pyogenes* Cas9 (hereafter Cas9) has been also facilitated by the fact that, unlike Cas9 from other species, it has a short and relatively simple 5'-NGG PAM. This occurs on average every 8 base pairs in the human genome thus the majority of sequences can be targeted, whereas longer PAMs can significantly reduce the number of available targets [8].

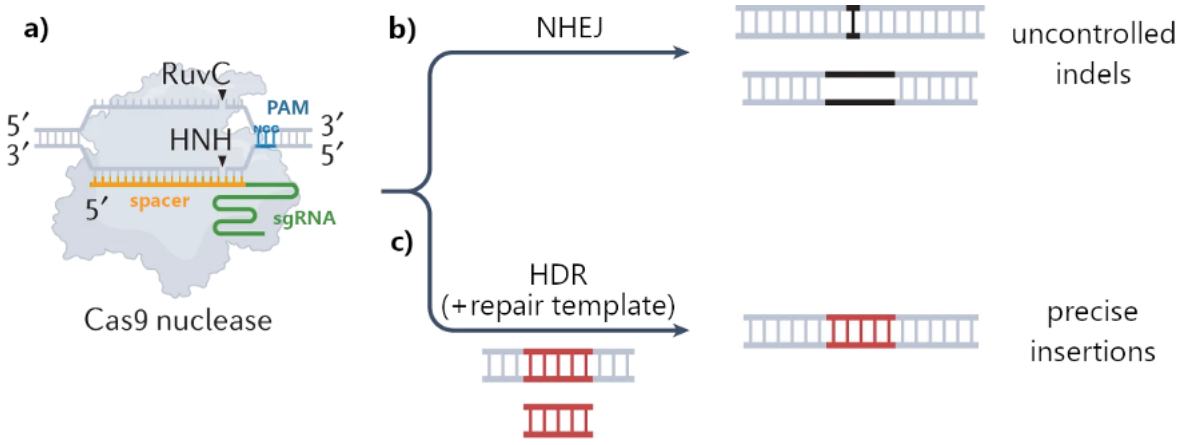

Figure 1: The Cas9 nuclease and the two main pathways of DSB-induced editing
a) Cas9 in complex with the DNA target. The spacer part of the sgRNA is yellow, the scaffold is green. The PAM on the target DNA is highlighted in blue. Black rectangles indicate the positions at which the RuvC and HNH domains cleave the DNA strands. **b)** NHEJ results in smaller insertions or deletions (black). **c)** HDR with a repair template allows the precise insertion of a selected DNA segment (red). Figure source: [12] edited.

1.2 First generation CRISPR editing tools

The first generation of CRISPR-based genetic engineering tools relied on the introduction of site-specific double-strand DNA breaks (DSB) into the genome. In eukaryotic cells these lesions are mainly repaired through end joining mechanisms like non-homologous end joining (NHEJ), or DNA donor dependent homology directed repair (HDR) pathways. As end joining mechanisms are often error-prone, in a significant number of cases the original sequence is not restored, but short insertions or deletions (indels) of a few base pairs are created [13] (Figure 1b). These can be used for gene knockout, as targeting the coding region of the gene with Cas9 can result in a frameshift mutation or the generation of an early stop codon, leading to loss of function [8]. Similarly, the disruption of cis- or trans-regulatory elements can be investigated [14]. Furthermore, by introducing two DSBs, larger deletions up to the megabase scale can be generated [15]. Of course, it is also possible that DNA ends are perfectly repaired. But then the restored sequence remains the substrate of Cas9, and the chances of indel formation increase with repeated cleavage [16].

HDR mechanisms, on the other hand, are usually error-free, but require a DNA template with a sufficient length of homologous sequence. By default, these are sister chromatids or, less frequently, homologous chromosomes, but for genome editing approaches they can be plasmids, linear double-stranded DNA templates or synthetic single-stranded oligonucleotides. The advantage of HDR is that the DNA sequence between the homology arms is freely adjustable, allowing the introduction of single base changes, smaller insertions or deletions and even whole gene fragments [13,17] (Figure 1c). Although HDR is a promising approach for precise genome editing, its two main limitations are its often low efficiency and the fact that it only has significant activity in the S and G2 phases of actively dividing cells [17]. Several methods have been published that aim to enhance the efficiency of HDR. These either optimise technical conditions such as cell cycle synchronisation [18] or the directing of the donor DNA to the DSB site [19]; or shift the HDR/NHEJ rate by blocking the NHEJ pathway or upregulating HDR. The latter can be achieved both by small molecule drugs [20,21] or the up or downregulation of key proteins [22].

1.3 Second generation CRISPR editing tools

The development of second-generation CRISPR editing tools has been driven by the need for more precise and safer gene editing, given the low efficiency of HDR and the potential genotoxicity of DSBs. Instead of using Cas9 as molecular scissors, they exploit its sequence-specific DNA targeting ability. To prevent the formation of DSBs, these methods rely on either a nuclease inactive "dead" Cas9 (dCas9) or a nickase Cas9 (nCas9). These enzymes are generated by introducing single amino acid changes to inactivate both or one of the nuclease domains of Cas9. The mutant proteins are still able to interact with the sgRNA and to sequence specifically bind the DNA target, but dCas9 cannot cleave it and nCas9s can only create a nick by cleaving only one strand [6,23]. In these systems editing is carried out by an additional DNA-modifying domain fused to Cas9. To date, the two most important technologies are base editing and prime editing. In base editors a nucleoside deaminase domain is used to introduce targeted single base changes. Initially, cytosine and adenine base editors were developed to catalyse C-to-T and A-to-G conversions, respectively [24,25]. Later, A-to-C [26], A-to-Y [27] and C-to-G [28] base editors were also reported. This way the base editing toolbox can cover all transition mutations and most of the transversion mutations. However, they are still unable to correct C-to-A transversions and smaller insertions and deletions. This is where the importance of prime editing lies, as it allows virtually any change within a shorter DNA segment.

1.3.1 Prime editing

In prime editor 2 (PE2), nCas9 is fused to an engineered Moloney murine leukaemia virus reverse transcriptase (RT) domain. It also has a 3' extended sgRNA, called the prime editing guide RNA (pegRNA), which both guides Cas9 to the DNA target and acts as a template for the RT - containing the desired mutation (Figure 2a). After target recognition, Cas9 nicks the PAM-containing DNA strand. The free 3' end of the DNA then hybridises to the complementary 3' end of the pegRNA, called the primer binding site (PBS). This primes reverse transcription using the rest of the pegRNA extension - also coding the desired edit - as a reverse transcription template (RTT) (Figure 2b). This results in a 3' DNA flap that competes with the unmodified 5' DNA flap for hybridisation to the DNA strand opposite the PAM. Although binding of the original 5' end, which does not contain

the mutation, is thermodynamically more favourable, cellular repair mechanisms can tip the balance towards hybridisation of the 3' overhang. The edit is permanently incorporated if mismatch repair favours the edited strand as a repair template when the DNA heteroduplex is resolved (Figure 2c). This can be promoted by adding a simple sgRNA to guide PE to nick the non-edited strand (PE3 approach). The second nick biases DNA repair mechanisms to correct this strand using the edited strand as a template (Figure 2d). Compared to PE2, PE3 often increases editing efficiency, but also results in higher levels of unwanted indels [29].

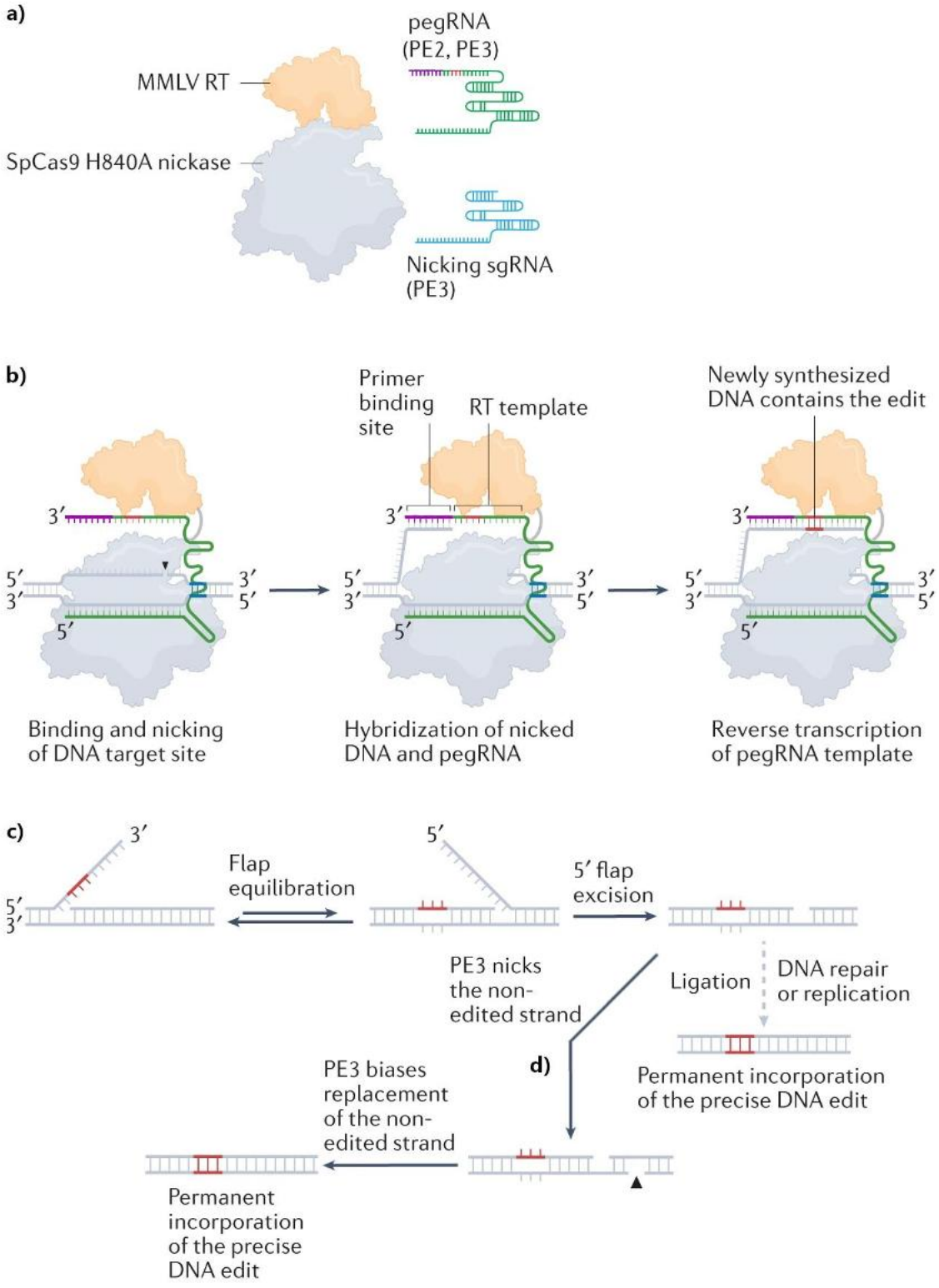

Figure 2: The prime editing system
In the PE protein, nCas9 is grey and the RT domain is orange. The pegRNA is green, with the PBS highlighted in purple and the intended modification in red. **a)** The components of prime editing. **b)** PE introduces the intended modification. **c)** Flap equilibration and resolution. **d)** PE3 biases repair mechanisms. Figure source: [12] edited.

The main advantage of prime editing is that the sequence of the RTT can be freely altered within certain limits. Firstly, this allows any substitution, small deletions or even insertions up to a few tens of nucleotides to be introduced. Secondly, its length is not fixed, so the number of editable positions is less limited by the availability of a PAM. Unfortunately, the widespread use of prime editing is hindered by its variable and often low editing efficiency. This is partly due to the fact that DNA repair mechanisms depend on the cell type and cell state. And partly because the length of the PBS and RTT, and the position of the second nick, vary with the target and intended modification. Therefore, extensive experimental optimisation is often required to achieve high editing efficiency.

1.4 The cytochrome P450 superfamily

Cytochrome P450 (CYP) enzymes are a large superfamily of haemoproteins, that play important roles in drug and cellular metabolism. Humans have 57 putatively functional genes, which are classified into families and subfamilies based on sequence identity. The CYP1-3 families are responsible for the biotransformation of around 80% of xenobiotics [30,31]. Some members are also involved in the metabolism of antitumour agents and have been investigated as potential therapeutical targets or biomarkers [31]. Individual genetic variations, among other factors, can influence the expression and function of CYP enzymes, leading to different responses to drug treatments in different individuals [30].

Selective targeting of CYPs using CRISPR tools can be challenging due to the high sequence similarity between homologous genes. This study assessed the targetability of the *CYP1A1*, *CYP1A2* and *CYP2B6* genes using prime editing. CYP1A1 is expressed in extrahepatic sites and peripheral blood cells. It has endogenous substrates such as hormones and fatty acids, as well as xenobiotics such as polycyclic aromatic hydrocarbons. Elevated levels of CYP1A1 have been associated with an increased risk of lung cancer for the latter [31,32]. CYP1A2, on the other hand, is exclusively expressed in the liver, where it metabolises various xenobiotics. It is the main metaboliser of caffeine and theophylline, for example [32]. CYP2B6 is primarily expressed in the liver and contributes to the metabolism of many drugs. It has several polymorphic variants, and it was suggested, that the CYP2B6 genotype should be taken into account when prescribing the HIV drug efavirenz [33].

2. Objectives

In my PhD work my aim was to:

- Develop a reporter system for an easy, fluorescence-based detection of prime editing outcomes that allows maximum flexibility for the target sequences and pegRNA designs to be tested on it.
- To verify whether the same factors influence prime editing efficiency on the PEAR plasmid as in genomic context.
- Find out whether prime edited mammalian cells could be identified and enriched by using a plasmid-based surrogate marker for chromosomal DNA modifications.
- To demonstrate that PEAR can be applied to facilitate the development of new prime editing systems.
- To compare the efficiency of proPE and PE on target-distal edits, using targets identified in the clinically relevant *CYP* gene family.

3. Methods

3.1 Molecular cloning

The following plasmids were acquired from the non-profit plasmid distribution service Addgene: pCMV-PE2 (#132775), pAT9624-BEAR-cloning (#162986), pAT9651-BEAR-GFP (#162989), pAT9752-BEAR-mScarlet (#162991), pAT9658-sgRNA-mCherry (#162987), pAT9679-sgRNA-BFP (#162988), pX330-Flag-wtSpCas9-H840A (#80453), pX330-Flag-dSpCas9 (#92113), and pX330-Flag-wtSpCas9 (#92353).

The following plasmids developed in this study are available from Addgene: pDAS12125_PEAR-GFP (#177178), pDAS12124_PEAR-GFP-preedited (#177179), pDAS12069-U6-pegRNA-mCherry (#177180), pDAS12222-U6-pegRNA-BFP (#177181), pDAS12230_pegRNA-PEAR-GFP(10PBS-24RT)-mCherry (#177182), pDAS12137_sgRNA-PEAR-GFP-nick(+17)-mCherry (#177183). pDAS12342_PEAR-GFP_2in1 (#177184), pDAS12395_PEAR-GFP-2.0-preedited (#177185), and pDAS12489_PEAR-GFP-2in1-2.0 (#177186).

For linker oligonucleotides and/or spacer, RTT and PBS sequences used in the study see reference [34] Supplementary file 3, or reference [35]Supplementary Tables 2-3.

PEAR target plasmids were constructed by Esp3I digestion of the pAT9624-BEAR-cloning plasmid [36], followed by one-pot cloning of the corresponding linker oligonucleotides.

For the construction of pegRNAs, first the one-step cloning protocol described by Anzalone et al. [29] was used, however, it resulted in frequent deletions and/or mutations in the constructs. Therefore, a two-step cloning method was developed: first, the spacer coding linkers were cloned into pDAS12069 or pDAS12222 plasmids between BpiI sites via one-pot cloning. In the second cloning step oligonucleotides bearing the PBS-RTT were cloned between Esp3I sites of the plasmids created in the first cloning step.

For second nicking sgRNAs targets were cloned into pAT9658-sgRNA-mCherry or pAT9679-sgRNA-BFP plasmids between BpiI sites via one-pot cloning, as described above.

In each case, one-pot cloning mixtures were incubated at 37°C for 30 min before being transformed into NEB5-alpha competent cells. The sequence of the cloned plasmids was verified by Sanger sequencing (Microsynth AG).

3.2 Cell culturing and transfection

HEK293T (ATCC, CRL-3216) and U2OS cells (ATCC, HTB-96), and the HEK-BEAR-GFP and HEK-BEAR-mScarlet cell lines were grown in DMEM. K562 cells (ATCC, CCL-243) were grown in RPMI 1640. Both media was supplemented with 10% heat-inactivated foetal bovine serum and 100 units/mL penicillin and 100 µg/mL streptomycin. Cells were cultured at 37°C in a humidified atmosphere of 5% CO_2.

In plasmid-based PEAR experiments, HEK293T cells were seeded on 48-well plates 1 day before transfection at a density of $5 \cdot 10^4$ cells/ well. A total of 565 ng DNA was used: 55 ng of PEAR target plasmid, 153 ng of pegRNA-mCherry (or pegRNA-BFP in the case of BEAR-mScarlet), 49 ng of sgRNA-mCherry (or sgRNA-BFP in the case of BEAR-mScarlet), and 308 ng of PE2 coding plasmid. These were mixed with 1 µl Turbofect reagent diluted in 50 µl serum-free DMEM, and the mixture was incubated for 30 min at RT before adding it to the cells. Each transfection was performed in three replicates. Cells were analysed by flow cytometry on day 3 from transfection.

The BEAR-GFP and BEAR-mScarlet cell lines were transfected with a total of 565 ng DNA: 170 ng of pegRNA-mCherry (or pegRNA-BFP in the case of BEAR-mScarlet), 54 ng of sgRNA-mCherry (or sgRNA-BFP in the case of BEAR-mScarlet), and 340 ng of PE2 coding plasmid, that was all mixed with 1 µl Turbofect reagent diluted in 50 µl serum-free DMEM. The mixture then was incubated for 30 min at room temperature before adding it to the cells.

In the enrichment experiments, where cell sorting was used, HEK293T cells were transfected with the above protocol. A total of 550 ng DNA was used: 40 ng of PEAR-GFP- 2in1 plasmid, 153 ng of pegRNA- BFP, 49 ng of sgRNA-BFP, and 308 ng of PE2 coding plasmid. In each parallel condition a total of 12 wells were transfected. Three days after transfection, cells were trypsinised and sorted directly into genomic lysis buffer, which was followed by genomic DNA extraction.

In the case of U2OS and K562 cells, 1000 ng total DNA; 603 ng PE2 coding plasmid, 300 ng pegRNA-BFP, 97 ng sgRNA-BFP, and 78 ng PEAR-2in1 plasmid was nucleofected in a final volume of 20 µl in a 16-well nucleocuvette strip. 3×10^5 U2OS and 5×10^5 K562 cells were transfected per well using an Amaxa 4D-Nucleofector (Lonza) according to the manufacturer's protocol. For U2OS, the SE Cell Line 4D-Nucleofector X Kit (program DN-100) for K562 the SF Cell Line 4D-Nucleofector X Kit (program FF-120) was used. Three days after transfection, U2OS and K562 cells were sorted directly into genomic lysis buffer, which was followed by genomic DNA extraction. ***The cell culturing and transfection of U2OS and K562 cells was carried out by Péter István Kulcsár.***

In proPE experiments HEK293T cells were seeded on 48-well plates 1 day before transfection at a density of $3\cdot10^4$ cells/ well and a total of 350 ng DNA was used. The inhibitory effect of degraded RNAs on PE and proPE was tested with the following amounts: 223 ng PE2 coding plasmid, 36 ng PEAR plasmid, 91.4 ng pegRNA plasmid for PE and 88.2 ng tpgRNA + 3.1 ng engRNA coding plasmid for proPE. Degraded RNAs decreased the amount of pegRNAs or tpgRNAs. In experiments on the CYP gene targets 248 ng prime editor coding plasmid was used with 102 ng pegRNA or engRNA+tpgRNA coding plasmid in case of PE2, and 62 ng pegRNA or engRNA+tpgRNA coding plasmid and 40 ng second nicking sgRNA plasmid in case of PE3.

3.3 Flow cytometry and cell sorting

Flow cytometry analysis was carried out using an Attune NxT Acoustic Focusing Cytometer (Applied Biosystems by Life Technologies). In all experiments, a 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) filter for emission. Attune Cytometric Software v.4.2 was used for data analysis.

Cell sorting was carried out on a FACSAria III cell sorter (BD Biosciences). The live single-cell fraction was acquired by gating based on side and forward light-scatter parameters. BFP or GFP signals were detected using the 405 or 488 nm diode laser for

excitation and the 450/50 or 530/30 nm filter for emission, respectively. To sort control (no enrichment) cells, live single cells were sorted regardless of any fluorescent markers. To sort transfection marker enriched cells, BFP-positive cells were sorted regardless of GFP fluorescence. To sort PEAR enriched cells GFP-positive cells were sorted regardless of BFP fluorescence. A minimum of 50,000 cells were sorted in all experiments. ***In every experiment, cell sorting was carried out by Dr. György Várady.***

3.4 Genomic DNA purification and genomic PCR

Genomic DNA from sorting experiments was extracted according to the Puregene DNA Purification protocol (Gentra Systems Inc). The purified genomic DNA was subjected to PCR analysis conducted with Q5 polymerase and locus-specific primers (see Table 1 on page 50). PCR products were gel purified via NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) and were subjected to next-generation sequencing.

To extract the genomic DNA, cells were centrifuged at 100 rcf for 10 minutes. After discarding the supernatant, 300 µl lysis buffer (10 mM Tris-HCl pH 8.0; 25 mM EDTA; 0.5 % SDS) containing 10 µg RNase A (Sigma-Aldrich) was added to of the cell pellet. Samples were incubated at 37 °C for 30 mins, then 200 µl solution of 5 M ammonium acetate was added, samples were mixed by shaking and centrifuged at 16,000 rcf for 10 minutes. 500 µl of the supernatant was transferred to 750 µl of cooled isopropanol (-20°C). The mixture was centrifuged for 10 minutes at 16,000 rcf, after which the supernatant was poured off. The precipitated DNA was then washed with 700 µl of 70% ethanol and centrifuged at 16,000 rcf for 10 minutes. The supernatant was discarded; tubes were placed in a thermoblock heated to 65 °C to evaporate the remaining ethanol. The genomic DNA was redissolved in 30 µl of 5 mM Tris-HCl buffer solution (pH 8.5) and incubated for 1 h at 65 °C.

3.5 Next generation sequencing analysis

3.5.1 PEAR enrichment experiments

Sequencing on an Illumina NextSeq instrument was performed by Delta Bio 2000 Ltd. Reads were aligned to the reference sequence using BBMap. Indels were counted computationally among the aligned reads that matched at least 75% to the first 20 bp of the reference amplicon. Indels without mismatches were searched at ± 2 bp around the

cut site with allowing indels of any size. For each sample, indel frequency was determined as (number of reads with an indel)/(number of total reads).

Frequency of substitution without indels generated by prime editing was determined as the percentage of (sequencing reads with the intended modification, without indels)/(number of total reads). By contrast, frequency of intended insertions or deletions generated by prime editing was determined as the percentage of (all sequencing reads with the intended modification)/(number of total reads). For these samples, the indel background was calculated from reads containing different types of indels, than the aimed edit. For NGS analysis, the following software were used: BBMap 38.08, samtools 1.8, BioPython 1.71, and PySam 0.13. To avoid falsely high specificity ratios on Figures 8c and 10c during calculations indels lower than 0.05% were assumed to be 0.05% as this amount is considered to be the resolution limit of NGS. The deep sequencing data have been submitted to the NCBI Sequence Read Archive under accession number PRJNA779199***. In all PEAR NGS experiments bioinformatic calculations were performed by Sarah Laura Krausz.***

3.5.2 Experiments on the *CYP* gene targets

For each *CYP* gene, gene-specific primers were used. Prior to alignment, primer-specific reads were sorted based on their first 8 nucleotides. During analysis, 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. Editing and indel values were calculated as described above (Section 3.5.1).

3.6 Statistics

Unless stated otherwise, differences between samples were tested using one-way ANOVA with Tukey's post hoc test for homoscedastic samples. Homogeneity of variances was tested by Brown-Forsythe test and normality of residuals was tested by D'Agostino-Pearson omnibus (K2) test. In cases where data did not pass normality but fulfilled the assumptions of Box-Cox transformation the transformed data were analysed as above. If not, Kruskal-Wallis test with Dunn's test was applied.

Figure 12: For each normalised 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.

Statistical tests were performed using GraphPad Prism 9.2.

4. Results

4.1 Development of the prime editor activity reporter (PEAR)

My first aim was to develop a reporter system for prime editing with the following key features:

- It should be a transient, plasmid-based system, so that it is quick and easy to use. This would allow the user to use it in multiple cell lines without the need to generate reporter cell lines.
- It should be based on a gain-of-function fluorescent signal with minimal background, so that even low prime editing efficiency can be detected in the timeframe of a transient system.
- It should provide unrestricted flexibility in the full length of the target sequence and the nucleotides surrounding the positions to be edited.

As we have previously developed a reporter system for base editing (named BEAR) that meets these requirements [36], my goal was to adapt it for prime editing. BEAR is based on a GFP protein whose coding sequence is split by the last intron of the mouse Vim gene. The sequence of the 5' splice site has been modified (from 5'-GT to 5'-AC) to abolish splicing and therefore the fluorescent signal. However, base editors can restore a functional splice site (Figure 3a). In order to adapt the assay to PEs, two main changes were required due to technical differences between the two gene editing methods. The first is that while base editors can edit in the 5' direction from the nick, PE can edit in the 3' direction. Therefore, to keep the PE target in the intron, I had to move the PAM to the complementary DNA strand. This was necessary to ensure that the target sequence could be easily swapped with minimal sequence constraints. The other difference is that prime editing can restore the original, canonical 5'-GT splice site, so the intended modification is a two-base substitution (Figure 3b).

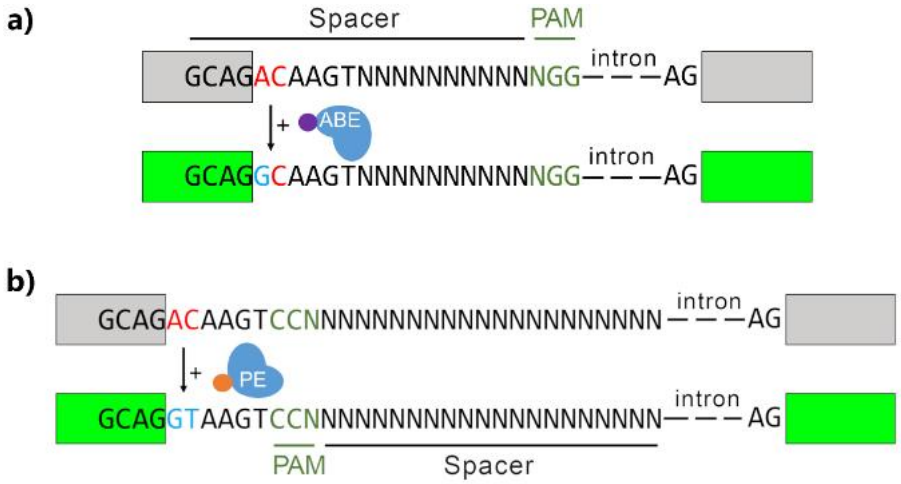

Figure 3:The schematic of BEAR and PEAR
Both reporters consist of a split GFP coding sequence separated by an intron with an inactive splice site, resulting in a dysfunctional protein (grey). Both editors can revert the splice site to a functional one and restore GFP expression (green). The altered bases in the splice site are shown in red, the edited bases are blue, the PAM sequence is dark green. **a)** Base editors (here ABE) act on the sense strand of the DNA. The variable bases in the sequence of the spacer are shown as 'N'-s, nCas9 is blue, and the fused tadA deaminase is purple. **b)** PE edits downstream of the cut site in the target. This enables the spacer sequence to be positioned within the intron and its entire length to be freely adjusted (shown as 'N'-s). nCas9 is blue and the fused reverse transcriptase is orange. The figure is from our publication [34].

Because there was no suitable PAM for prime editing at the 5' end of the intron, I installed one after the canonical AAGT motif flanking the splice donor site. For the target sequence, I modified the subsequent part of the intron to have a GC content of about 50%. To verify that GFP fluorescence would exclusively report on prime editing, I transfected a pegRNA - sgRNA combination together with either the PEAR-GFP plasmid, or its pair harbouring the active splice site, and either dCas9 or nCas9. As shown in Figure 4a, in both cases I only obtained fluorescent signal for the active splice site, indicating that nicking by nCas9 alone could not restore a functional splice site. The results also showed that sequence changes associated with establishing the prime editor target did not impair splicing activity.

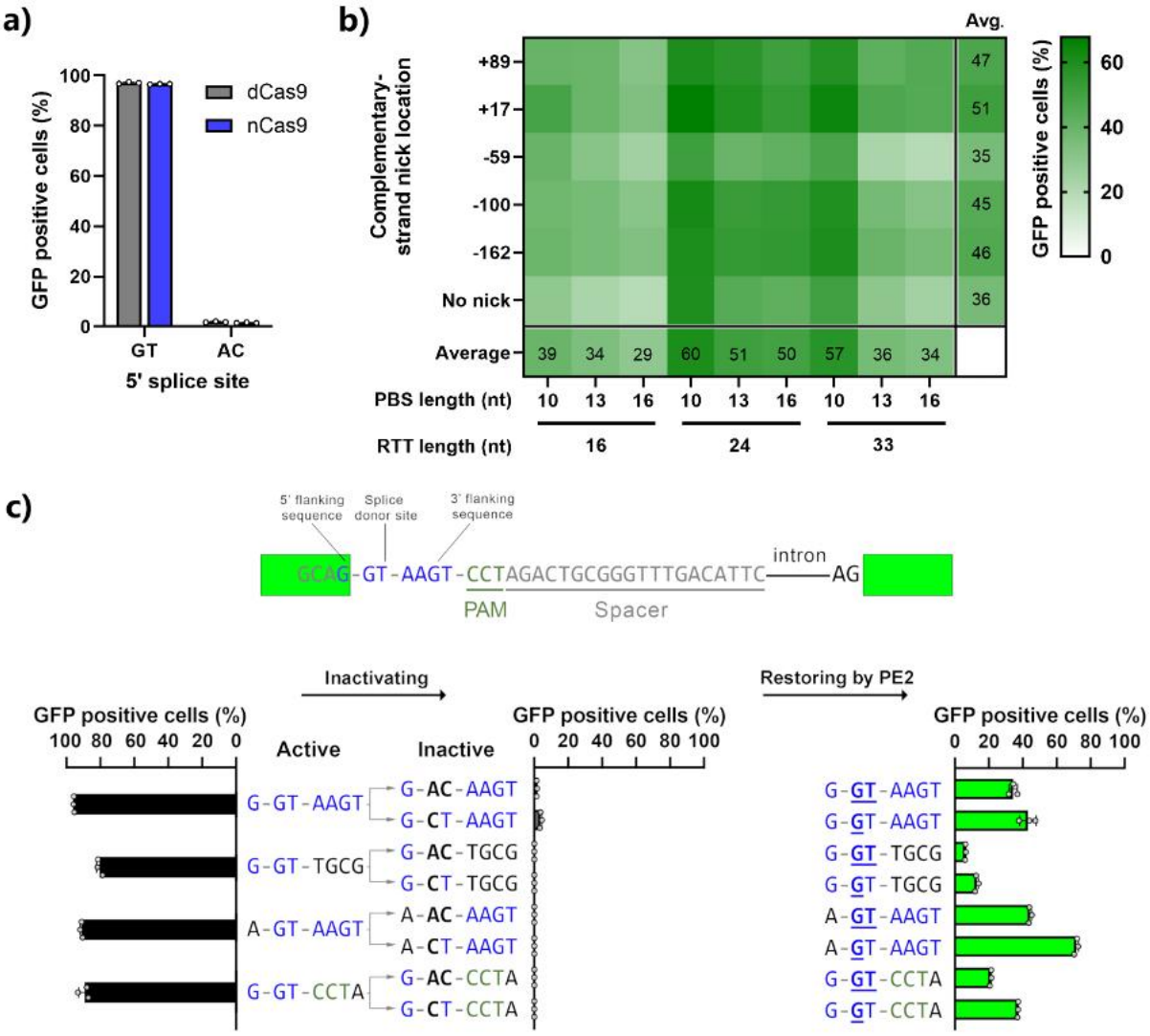

Figure 4: PEAR reports on prime editing efficiency

a) Flow cytometry measurements of GFP positive HEK293T cells co-transfected with a PEAR-GFP reporter plasmid harbouring a pre-edited active (GT) or an inactive (AC) 5' splice site, a pegRNA targeting the plasmid and nCas9 or dCas9. **b)** Optimisation of PBS, RTT, and complementary DNA strand nicking on the PEAR-GFP plasmid. The heatmap shows the average percentage of GFP-positive cells of three replicates of transfections with the PEAR-GFP plasmid in combination with different pegRNAs and sgRNAs for secondary nicking and the PE2 coding plasmid. The position of the second nick is given in relation to the first nick. Positive values indicate 3′, negative values indicate 5′ direction on the targeted DNA. When no second nick was introduced, it is indicated as 'no nick'. **c)** PE can restore various active splice sequences. Four additional active splice site variants were selected based on [36]. To generate suitable inactive plasmids, splicing was disrupted by systematically replacing the 5′-GT splice donor site to 5′-AC and 5′-CT (negative controls with inactive sequences, grey columns). With the appropriate pegRNAs, PE2 was able to restore GFP fluorescence in every case (green columns). Letters highlighted in blue indicate the bases of the canonical splice donor site and the flanking sequences which influence splicing the most: 5′-G-GT-AAGT-3′; the altered bases of the inactive sequences are bold, and bases in the active sequences that are reverted by PE2 are underlined. Columns represent means ± SD of three parallel transfections (white circles). Panel **b)** and **c)** are from our publication [34].

For the correction of the 5' splice site, I designed a total of nine pegRNAs by combining three PBS and RTT lengths, and five second nicking sgRNAs. I transfected these plasmids in all combinations into HEK293T cells together with the target plasmid (named PEAR-GFP) and the PE2 protein coding plasmid. I also compared the efficiency of PE2 and PE3 using a non-targeting sgRNA that has no target sites in the PEAR plasmid or the cellular genome. The number of GFP positive cells was measured by flow cytometry.

The heatmap in Figure 4b shows that several pegRNA - second nicking sgRNA combinations edited effectively the PEAR-GFP plasmid. The most efficient pegRNA was the one with a 10-nucleotide long PBS (PBS-10) and a 24-nucleotide long RTT (RTT-24). Overall, the most efficient second nicking sgRNA site was position +17 (the numbering of the nick positions starts from the pegRNA induced nick counting the first base 5' as position -1 and the first base 3' as position +1). The results also show that, as expected, editing was generally more efficient with PE3 than with PE2 (6th row, "no nick").

To demonstrate the versatility of PEAR regarding sequence modifications, I designed additional inactive splice site variants from three other active splice sites. These were also efficiently corrected by PE2 in most cases (Figure 4c). In this experiment the pegRNA target was the same as in Figure 4b, and the most efficient PBS and RTT lengths from Figure 4b were used (PBS-10 and RTT-24). The sequence of the RTTs differed as different active splice variants were restored.

4.2 PEAR in genomic context

To test whether the same factors influence prime editing on plasmids as on genomic targets, I compared editing efficiency on two cell lines previously constructed in our lab that contain an intron-disrupted GFP or mScarlet sequence with an inactive splice site (HEK-BEAR-GFP and HEK-BEAR-mScarlet, [36]) and plasmids coding the same disrupted proteins. Since these were designed for base editors, there is no suitable PAM for prime editing at the 5' end of the intron as mentioned above. I therefore searched for suitable PAM sites in the coding region of the fluorescent proteins adjacent to the 5' splice site. I identified one PAM for BEAR-GFP and two for BEAR-mScarlet. For these targets, I again designed nine pegRNAs by combining three PBS and RTT lengths. I also selected

three second nicking sgRNA sites in the intron, these could be used with all three targets as the intron is the same in BEAR-GFP and BEAR-mScarlet.

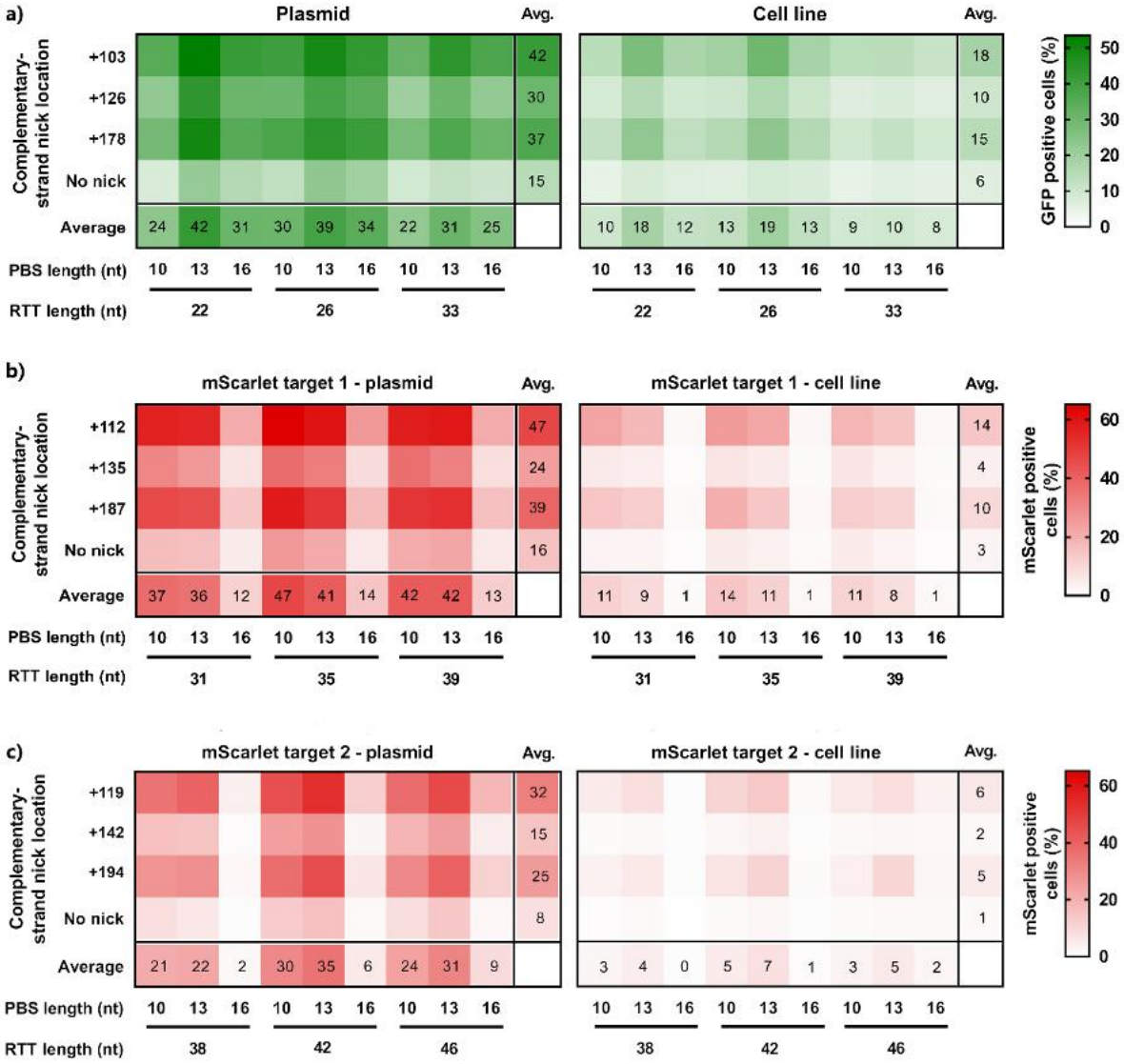

Figure 5: PEAR editing in a plasmid and in genomic context
Prime editing on **a)** BEAR-GFP, **b)** BEAR-mScarlet target 1 or **c)** BEAR-mScarlet target 2 either located in plasmids (Plasmid) or incorporated into the genome (Cell line). The heatmaps show the average percentage of **a)** GFP and **b-c)** mScarlet positive cells derived from three replicates. Cells were transfected with different pegRNAs and sgRNAs for secondary nicking, the PE2 coding plasmid, and in case of the 'Plasmid' experiments, the corresponding BEAR plasmid. The figure is from our publication [34].

In case of BEAR-GFP, both in the cell line and on the plasmid, the pegRNA with PBS-13 and RTT-22 or RTT-26 was the most efficient combined with the second nick site position +103 (Figure 5a). For BEAR-mScarlet target1, the highest editing was achieved with PBS-10, RTT-35 and for BEAR-mScarlet target2 with PBS-13, RTT-42.

The best second nick site was in both cases the same as for BEAR-GFP (its numbering is different because its distance from the pegRNA nick is different for every target) (Figure 5b-c).

For all three targets the heatmap pattern for plasmids and genomic sites is remarkably similar, and there is a strong correlation between the measured efficiencies as shown in Figure 6. The generally lower editing efficiency observed in the cell lines compared to the plasmids is likely due to the much higher copy number of the plasmids present in the cells. The observation that the sgRNAs used for the second nick had the same effect on editing efficiency for all three targets supports the hypothesis that the effect of the second nick is not primarily influenced by its exact distance from the pegRNA nick, but by other factors related to the targeted sequence and its environment. Together these results imply that our system reflects accurately the main features of prime editing and factors affecting its efficiency.

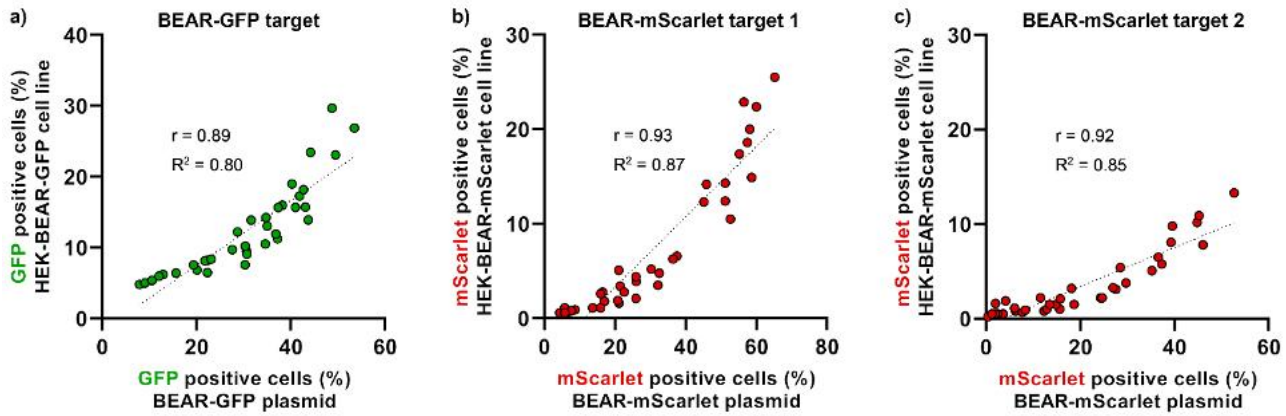

Figure 6: Correlation of editing efficiencies in the BEAR cell lines and the corresponding plasmids
a) Scatter plot of GFP-positive cells measured either when the BEAR-GFP plasmid or the HEK-BEAR-GFP cell line was edited by PE in various setups, Pearson's r=0.89 (data from Figure 5a). **b-c)** Scatter plot of mScarlet-positive cells measured either when the BEAR-mScarlet plasmid or the HEK-BEAR-mScarlet cell line was edited by **b)** PE targeting mScarlet target 1, Pearson's r=0.93 (data from Figure 5b) or **c)** PE targeting mScarlet target 2, Pearson's r=0.92 (data from Figure 5c). The figure is from our publication [34].

4.3 PEAR as a selection marker for prime edited cells

Previous work has shown that Cas9 edited cells can be successfully enriched with a plasmid-based surrogate marker that is subjected to the same type of genetic modification as the intended genomic edit (e.g. HDR or base editing) [36–38]. I tested

the applicability of the PEAR-GFP plasmid as a surrogate marker in the HEK-BEAR-mScarlet cell line.

Cells were transfected with the PEAR-GFP plasmid, two pegRNAs (also coding BFP to monitor transfection efficiency), one targeting the PEAR plasmid and the other targeting the mScarlet in the genome, a genomic second nicking sgRNA and the PE2 encoding plasmid. For PEAR-GFP, the two best performing pegRNAs from Figure 4b were tested. The ratio of mScarlet positive cells was measured by flow cytometry in: (i) all living cells, (ii) cells expressing BFP, i.e. transfection marker enrichment and (iii) cells expressing GFP for PEAR enrichment. Figure 7 shows that while both types of enrichment could significantly increase the percentage of the edited population, PEAR enrichment was more efficient, resulting in a 3 or 4-fold increase compared to the 1.4-fold increase of transfection enrichment.

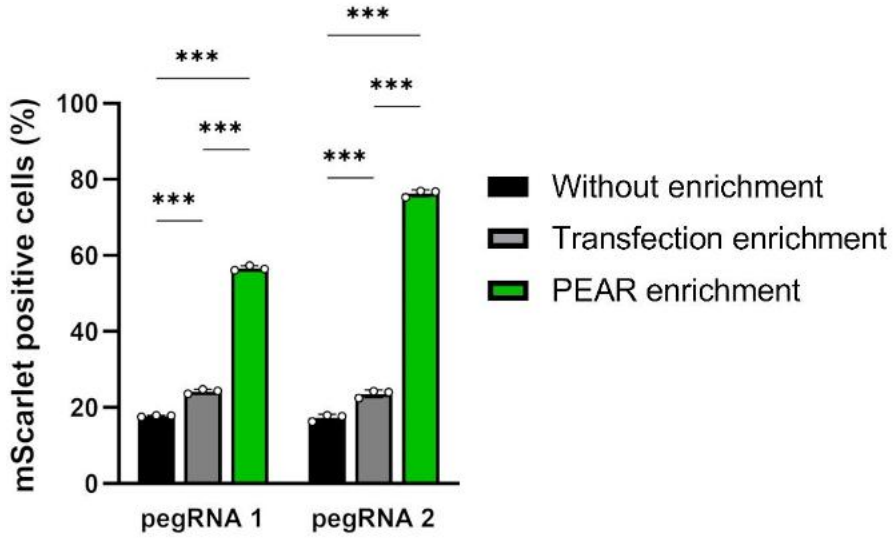

Figure 7: PEAR enrichment in the BEAR-mScarlet cell line
Flow cytometry measurements of prime edited (i.e., mScarlet positive) cells in the HEK-BEAR-mScarlet cell line. The PEAR-GFP plasmid was co-transfected with PE protein-coding plasmid, two pegRNAs targeting either the plasmid or the genomic PEAR sequences and an sgRNA (for complementary strand nicking at position +112). To edit the PEAR-GFP plasmid, two efficient pegRNAs were selected from the experiment in Figure 4b (pegRNA 1 and pegRNA 2 are those with RTT-24 and PBS-10, and RTT-33 and PBS-10, respectively). The mScarlet positive cell count was gated for either all live single cells (no enrichment, black bars); for BFP-positive cells (transfection enrichment, grey bars); and for GFP-positive cells (PEAR enrichment, green bars). Columns represent means ± SD of three parallel transfections (white circles). The figure is from our publication [34].

To investigate the effect of PEAR enrichment on endogenous targets, I selected five target sites (EMX1, RNF2, FANCF, HEK3, HEK4), at which single-nucleotide substitution prime editing has previously been optimised [29]. To facilitate transfection, the PEAR-GFP-2in1 plasmid was constructed, which encodes both the PEAR reporter and the best performing pegRNA from Figure 4b (pegRNA1 on Figure 7). In these experiments the pegRNA and sgRNA targeting the genome also coded BFP as a transfection marker. On the third day after transfection, cells were sorted into three fractions: (i) single living cells, i.e. no enrichment, (ii) cells expressing BFP - transfection marker enrichment and (iii) cells expressing GFP - PEAR enrichment. From the sorted populations, editing and indel formation at the genomic target sites was quantified by NGS.

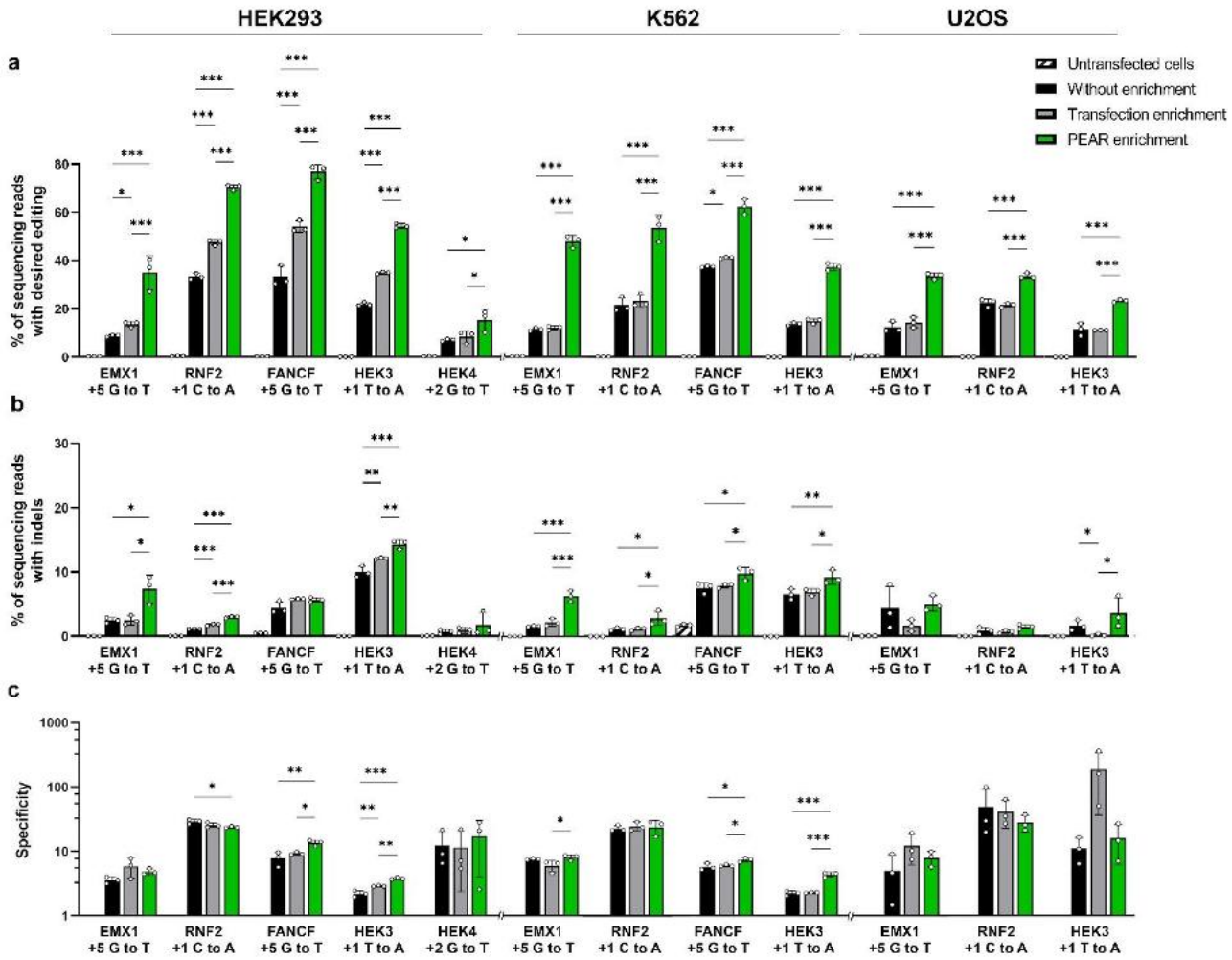

Figure 8: PEAR enrichment in HEK293T, K562 and U2OS cells
The PEAR-GFP-2in1 plasmid and endogenous genomic targets were co-edited in HEK293T, K562, and U2OS cells. Cells were enriched and analysed as described in the methods section. Results from cells without enrichment are shown in black, transfection enrichment in grey, PEAR enrichment in green, and untransfected cells in striped black and white. Precise prime editing (**a**) and unwanted indel formation (**b**) were quantified from the same samples. Specificity (prime editing%/indel%) was calculated separately for each sample (**c**). Columns represent means ± SD of three parallel transfections (white circles). When indel% was below the detection limit of NGS, specificity was calculated with 0.05% indel to avoid falsely high specificity values. Differences between samples were tested using one-way ANOVA. Only statistically significant differences are shown, differences to untransfected cells are not shown. *p<0.05, **p<0.01, ***p<0.001. The figure is from our publication [34].

PEAR enrichment significantly increased editing rates in all five cases, even compared to transfection marker enrichment (Figure 8a). However, for three targets it also increased unwanted indel formation (Figure 8b). To determine whether the increase in editing rates outweighed the increase in indels, specificity was calculated (as

edit%/indel%). As shown in Figure 8c, specificity decreased only slightly for one edit and increased for two.

HEK293T cells have a partially defective mismatch repair system, making them highly susceptible to prime editing [39,40]. To test PEAR in cell lines that typically have lower editing efficiencies, we also attempted enrichment in K562 and U2OS cells. As shown in Figure 8a, PEAR enrichment was effective in these cases, as well. Editing increased by an average of 2.6-fold in K562 cells, compared to transfection marker enrichment, and by 2.0-fold in U2OS cells. Some samples also exhibited increased levels of indels (Figure 8b), but PEAR enrichment did not compromise specificity in any case (Figure 8c).

Using the HEK293T samples, I also examined whether PEAR enrichment influenced off-target prime editing. For this I sequenced known Cas9 off-target sites of the EMX1, FANCF, HEK3 and HEK4 targets. Of the fifteen sites tested, there was only one site with detectable editing (Figure 9a) and indel formation (Figure 9b), which was also enriched by PEAR enrichment. For the remaining fourteen sites, no off-target modifications could be detected, regardless of the enrichment. Taken together, these data suggest that PEAR enrichment can significantly increase the efficiency of prime editing without altering either on-target or off-target specificity.

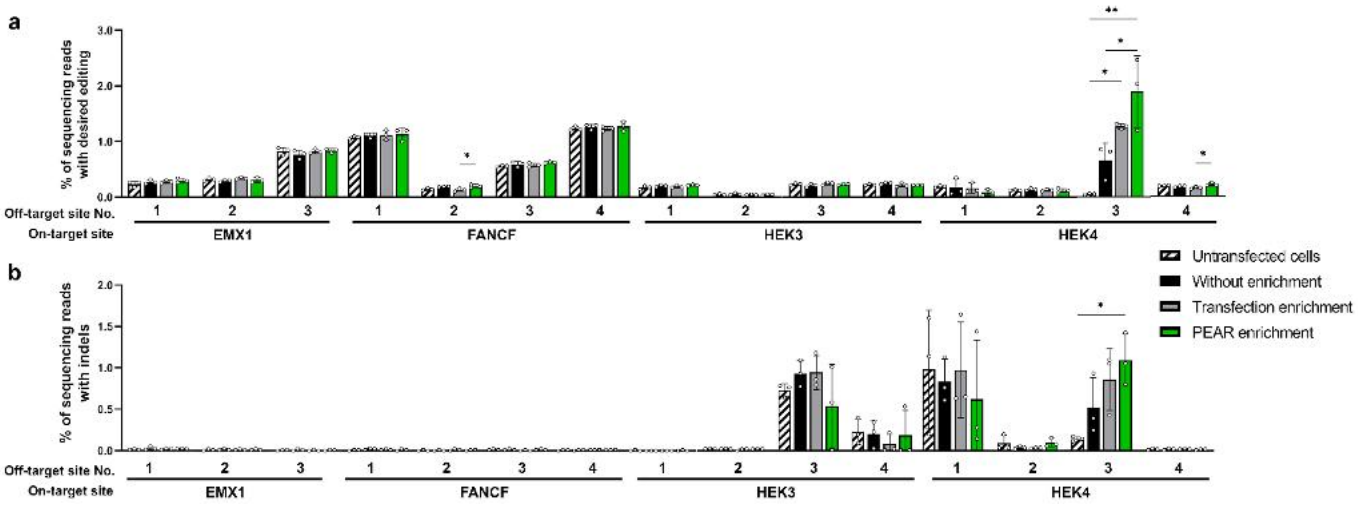

Figure 9: Off-target analysis of PEAR enrichment from HEK293T samples
Off-target prime editing (**a**) and indel formation (**b**) were analysed for known off-targets of target sites from (Figure 8a) and (Figure 8b) in HEK293T cells. Columns represent means ± SD of three parallel transfections (white circles). Differences between samples were tested using one-way ANOVA. Only statistically significant differences are shown. *p<0.05, **p<0.01, ***p<0.001. The figure is from our publication [34].

We also wanted to ensure that the pegRNA targeting the PEAR plasmid had minimal genome-wide off-target effect. However, it is difficult to assess the off-target sites of the pegRNA targeting the intron, as it lacks a reference on-target site in the human genome. Therefore, we decided to construct a new PEAR-GFP-2in1 plasmid with a suitable pegRNA target. For this, I chose the EGFP target site 2, which has no on-target site in the human genome, but its genome-wide off-target effect was previously evaluated by GUIDE-seq in HEK cells carrying a genome-integrated GFP [41]. As this EGFP target is also present in the PEAR reporter, I introduced silent mutations in the GFP coding sequence, creating six mismatches to ensure that the pegRNA will not target the GFP on the plasmid. I tested the enrichment efficiency of the new plasmid (PEAR-GFP-2in1-2.0) on nine target sites introducing insertions, deletions or substitutions. As Figure 10 shows, PEAR-GFP-2in1-2.0 can also efficiently enrich prime edited cells without compromising specificity.

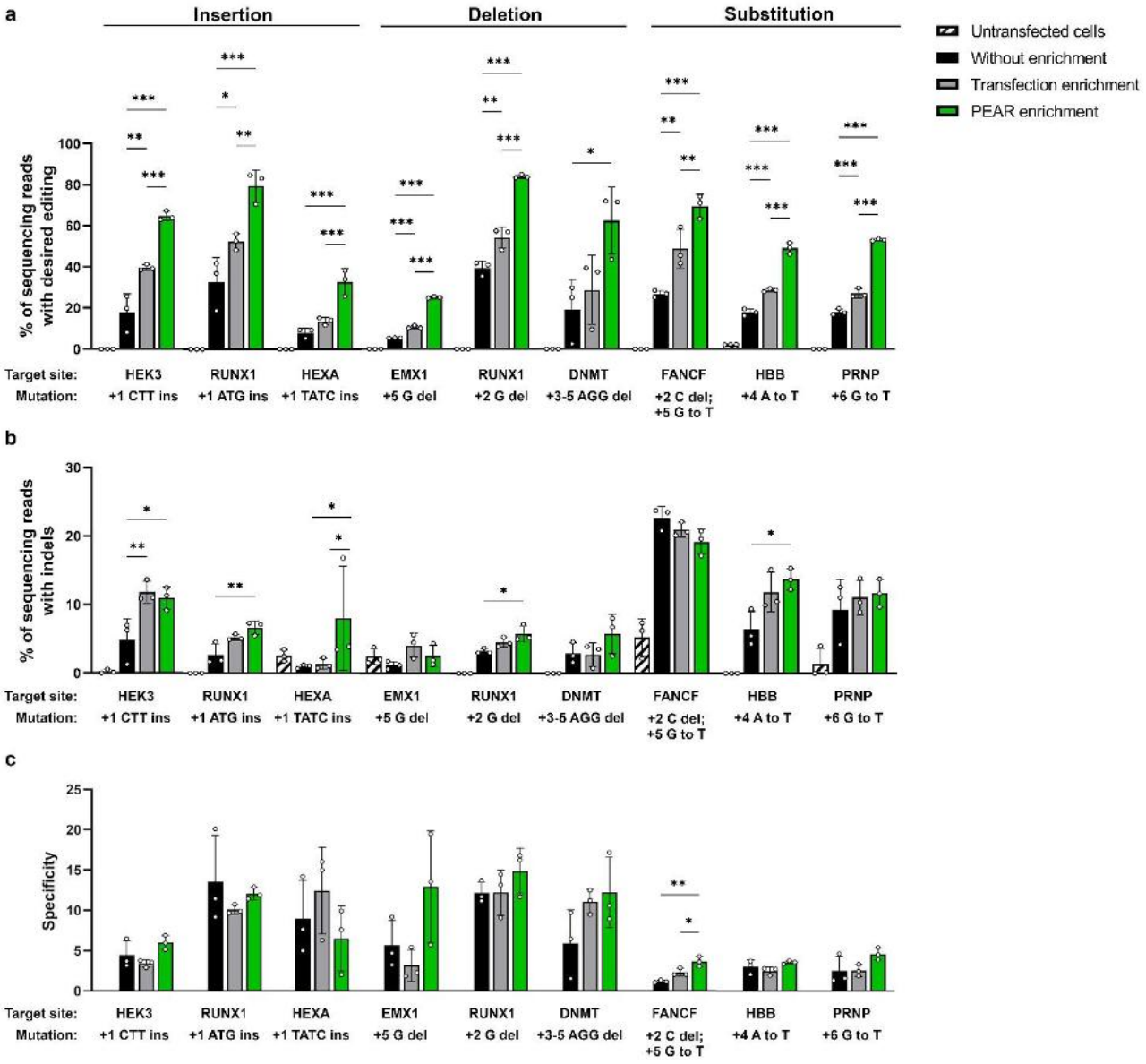

Figure 10: PEAR enrichment with PEAR-GFP-2in1-2.0
Results from cells without enrichment are shown in black, transfection enrichment in grey, PEAR enrichment in green, and untransfected cells in striped black and white. Prime editing (**a**) and indel formation (**b**) were quantified from the same samples. (**c**) Specificity (prime editing%/indel%) was calculated separately for each sample. Columns represent means ± SD of three parallel transfections (white circles). When indel% was below the detection limit of NGS, specificity was calculated with 0.05% indel to avoid falsely high specificity values. Differences between samples were tested using one-way ANOVA. Only statistically significant differences are shown, differences to untransfected cells are not shown. *p<0.05, **p<0.01, ***p<0.001. The figure is from our publication [34].

4.4 The application of PEAR in the development of a novel PE tool

ProPE (prime editing with prolonged editing window), a novel PE tool was recently developed in our research group. In this system, the two main functions of the pegRNA are carried out by two separate RNAs: The essential nicking guide RNA (engRNA), which is a standard sgRNA, is responsible for the targeting and subsequent nicking of the DNA at the selected target site. The template providing guide RNA (tpgRNA) is a pegRNA targeted to a nearby target site to provide the PBS and RTT sequences required for editing. It has a truncated spacer (11-15 nt), which ensures that it only binds to the DNA and does not nick it (Figure 11). During the development process, we used the PEAR system to determine the effective parameter ranges of proPE, and then to demonstrate some of its efficiency enhancing effects compared to PE [35].

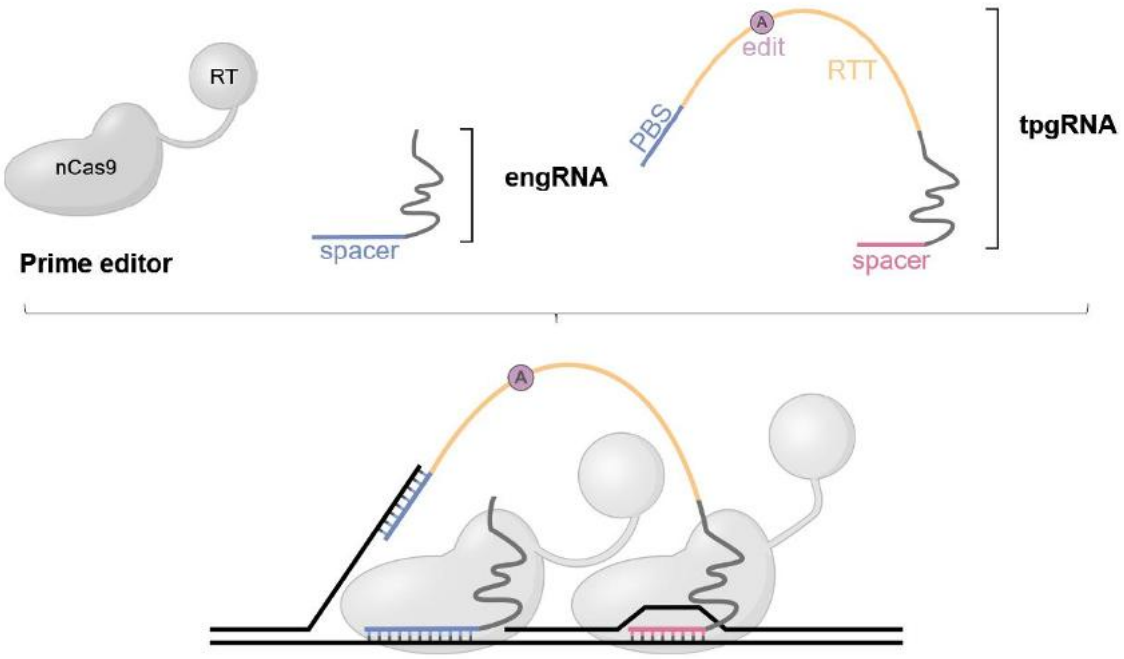

Figure 11: The schematic of proPE
The prime editor protein is grey, the spacer of the engRNA and the complementary sequence of the PBS of the tpgRNA are blue, the shorter spacer of the tpgRNA is pink, the scaffold part is dark grey, the RTT containing the desired mutation (purple) is yellow. The double stranded DNA is shown in black. PE complexed with an engRNA nicks the target DNA, the subsequent editing is carried out by the PE – tpgRNA complex directed to a nearby target site. Figure elements were designed by Viktória Faragó for our publication [35].

It was previously shown that degraded pegRNAs can significantly impair editing efficiency [42]. As well as that, when Cas9 forms an inactive, cleavage-incompetent complex with an sgRNA having a shorter spacer, its dwell time on the target DNA is

shorter [43–45]. Thus, one potential advantage of proPE is, that as PE complexed with a tpgRNA would dissociate more quickly, it would be less prone to the inhibitory effect of degraded pegRNAs. To test this hypothesis, I designed an experiment based on the assay used in the study by Nelson et al.[42], and the PEAR reporter system. To mimic the degradation of tpg/pegRNAs, the proportion of intact RNA coding plasmids was gradually replaced with plasmids coding PBS-less RNAs or sgRNAs (i.e. RTT-PBS-less RNAs). In control experiments, degraded RNAs with non-targeting spacers were used to measure the effect of the decrease in the amount of intact RNA alone. The assay was measured using the BEAR-mScarlet plasmid and a PEAR-GFP plasmid designed for proPE. Figure 12 shows, as an example, the results obtained with (RTT-PBS)-less RNAs on the PEAR-GFP plasmid. It is apparent that as the amount of intact RNA decreases, the efficiency decreases even for controls. Furthermore, PE appears to be more inhibited by targeting sgRNAs than proPE.

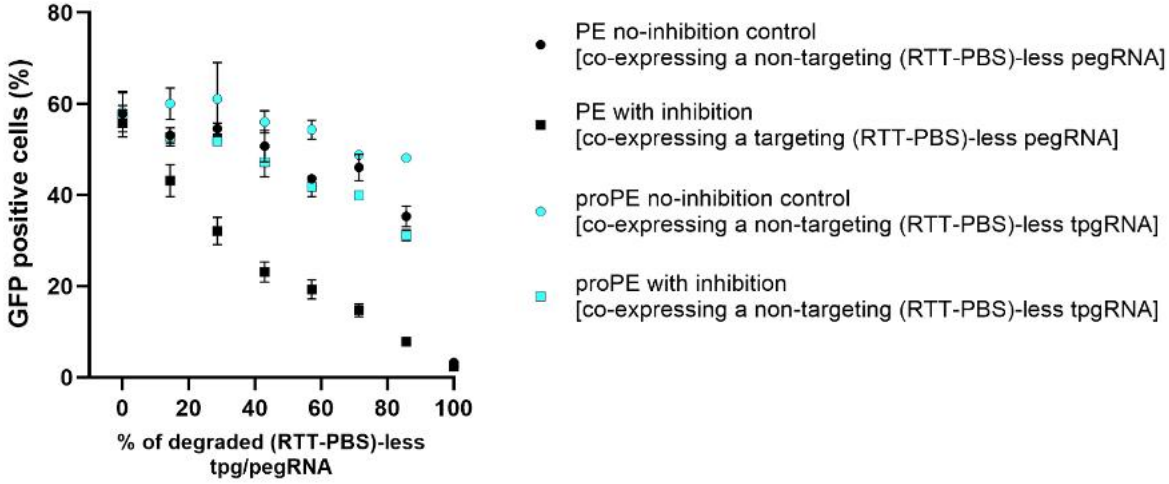

Figure 12: The inhibitory effect of (RTT-PBS)-less RNAs on the PEAR-GFP plasmid
PE and proPE editing were measured in the presence of an increasing proportion of (RTT-PBS)-less RNAs to mimic degraded RNAs. Data points represent means ± SD of three parallel transfections. In cases with inhibition the degraded RNA targets the same site as the pegRNA/tpgRNA, while in "no-inhibition control" cases the degraded RNA is targeted to a target site not relevant in the experimental setup. Data are from our publication [35].

To visualise the inhibitory effect without the efficiency-decreasing effect of the decreasing percentages of intact RNA, at each proportion, results with targeting RNAs were normalised to results with the corresponding non-targeting RNAs. For each dataset, a line was fitted to the data points, and statistical significance was assessed between the

slope of the fitted lines. As Figure 13 shows, degraded RNAs inhibited proPE less than PE in every case.

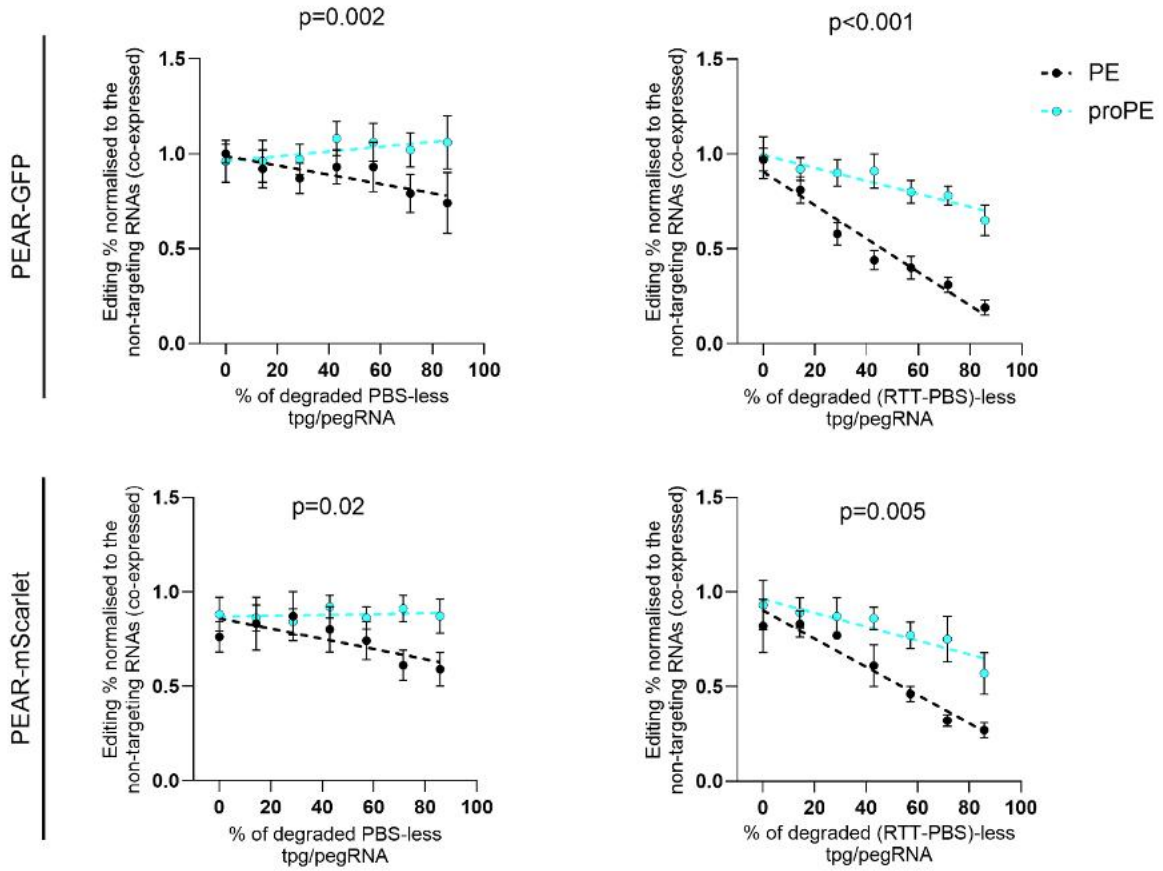

Figure 13: PE and proPE editing in the presence of degraded pegRNAs/tpgRNAs
PE and proPE editing was measured in the presence of an increasing proportion of PBS-less RNAs (left column) or (RTT-PBS)-less RNAs (right column) targeting the same pegRNA/tpgRNA site. The editing values are normalised to the 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. The figure is from our publication [35].

4.5 Comparing the efficiency of proPE and PE on target-distal edits

Unlike in previous PEAR experiments, early experiments on genomic targets showed that proPE was not significantly more efficient than PE. This inconsistency could be explained by the position of the edit. At that time, the majority of edits reported in the literature were within ten nucleotides from the pegRNA nick site. However, in the PEAR assays, edits were outside this 10-nucleotide window. Furthermore, data from the

literature [46] and our own data (not shown here) indicated that the efficiency of PE usually declined when the distance of the edit was 10 or more nucleotides away from the nick site (target-distal edits). To evaluate proPE's ability to enhance the efficiency of target-distal edits, I compared the efficiencies of PE and proPE in installing naturally occurring SNPs in the *CYP1A1* (two edits), *CYP1A2* (three edits) and *CYP2B6* (seven edits) genes, for which no prior data were available. For each SNP I designed two pegRNAs which differed in their target sites and/or PBS lengths, and one or two second nicking sgRNAs, depending on the availability of target sites. I also identified suitable tpgRNA targets to construct a corresponding proPE editing setup for each pegRNA. In total, this yielded 62-62 different editing combinations for PE and proPE (for detailed design see Table 2 on page 53). Editing efficiency in HEK293T cells was quantified by NGS. To minimise the effect of mispriming from homologous *CYP* genes, I designed gene-specific PCR primers using Primer-BLAST [47] and modified the NGS data analysing script (see Section 3.5.2). A summary of the results of the experiment is shown in Figure 14. In line with previous experiments, proPE could significantly increase the median editing efficiency from 1.2% to 5.2%, and the median specificity 4.1-fold.

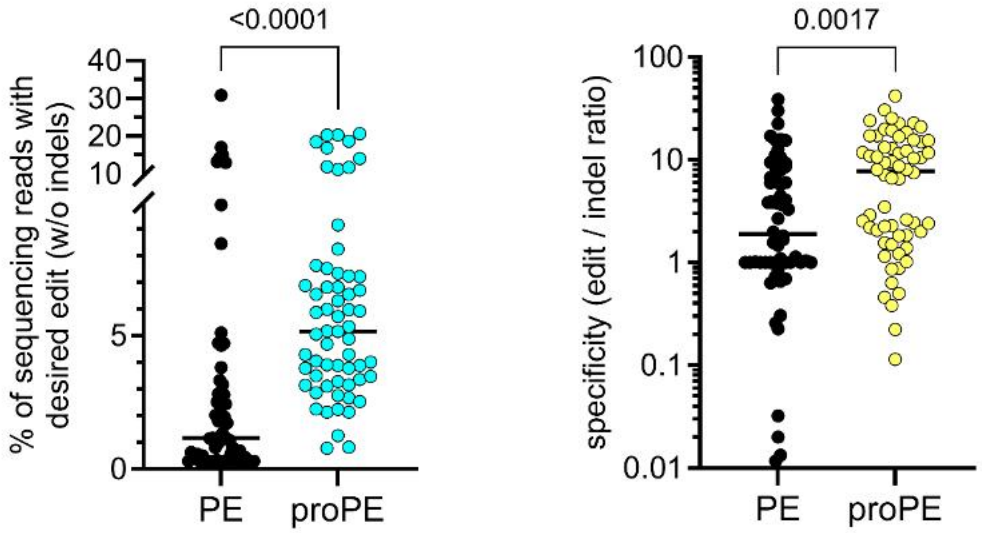

Figure 14: PE and proPE editing on the new *CYP* gene targets
Editing efficiency (left graph) and specificity (right graph) of PE and proPE on the *CYP* gene targets. Specificity was calculated as edit%/indel% from the same samples. Statistics: two-tailed Mann-Whitney test, in each data group N=62. Dots represent the average of three parallel samples, horizontal lines show the median of groups. Data are from our publication [35].

5. Discussion

Prime editing is a versatile CRISPR tool that enables the precise installation of substitutions and smaller insertions or deletions. Initially, its widespread use was hindered by its often low efficiency and the fact that the length of the PBS and the RTT, as well as the position of the second nick needs to be optimised for every intended modification.

The main goal of my PhD work was to develop methods that can help to overcome these problems and thus facilitate the application of prime editing. At the heart of this was the development of PEAR, a plasmid-based fluorescent assay that enables the easy monitoring of prime editing efficiency in mammalian cells. PEAR has several advantageous features that can help in the development of new prime editing approaches.

Firstly, it is insensitive to the indel background caused by the nicking of Cas9, while at the same time, it is sensitive enough to distinguish between the editing efficiencies of different prime editing setups (Figure 1a-b). Secondly, by design it provides unrestricted flexibility in the target sequence and with the tolerance of the splice donor site for substitutions, even the region of the edit is readily adjustable (Figure 4c). This makes it possible to create customised PEAR plasmids for different applications.

Lastly, as experiments with the BEAR-GFP and BEAR-mScarlet plasmids and the corresponding cell lines confirmed, the efficiency of prime editing to modify PEAR plasmids is governed by the same factors as prime editing in genomic context (Figure 5 and Figure 6). Thus, PEAR experiments can provide information that is also relevant when targeting genomic regions. Taken together, these features allow the user to examine the effect of sequence features and other factors affecting the efficiency of prime editing, while keeping all other factors constat. This makes possible such systematic investigations that could not be explored otherwise using genomic targets, making PEAR a suitable tool for the characterization of different PEs.

Beyond these, PEAR can also be applied as a surrogate marker to enrich prime edited cells. On average, in all enrichment experiments with cell sorting, PEAR enrichment increased the number of edited cells by 2.8-fold compared to samples without enrichment. With this, it was twice as efficient as transfection marker enrichment, which only resulted in a 1.4-fold increase (Figure 8a, 10a). To facilitate the use of PEAR as an

enrichment marker, the PEAR-GFP-2in1-2.0 plasmid was constructed, on which the pegRNA targeting the reporter has no detectable off-target sites in the human genome as determined by GUIDE-seq. This plasmid could enrich various types of edits with the same efficiency as the original PEAR-GFP-2in1 plasmid (Figure 10).

Although PEAR enrichment can only be used when cell sorting is feasible, it still has the potential to facilitate the more widespread use of prime editing. This type of selection with a transient marker can be advantageous, for example, when creating disease modelling cell lines, for which hard-to-edit cell lines or low-efficiency targets often have to be used.

At the time we developed PEAR, only a few reporter systems for prime editing had been published. Even these either showed a rather low signal [37] or were restricted to very few target sequences [48,49]. Since then, a number of other reporter systems, some of which can be used more broadly, have been published. PINE-TREE is based on a BFP-to-GFP conversion by prime editing the 66th amino acid in the BFP from histidine to tyrosine. Thus, the target sequence is relatively fixed, so it mainly can be used as an enrichment marker. Similarly to PEAR, PINE-TREE enrichment also surpasses the efficiency of transfection marker enrichment, and the authors even demonstrated that it can be used to generate clonal cell lines from human pluripotent stem cells [50]. The fluoPEER [51] and V-GEAR [52] systems have a similar design in that they both place a genomic target region of interest between two reporter genes, and this construct forms an expression cassette driven by a constitutive promoter The first reporter enables the identification and/or selection of cells containing the reporter construct. The target region is inserted in such a way that expression of the downstream reporter is abolished by a premature stop codon or a frameshift mutation, which can be resolved through the intended edit. Therefore, expression of the second reporter indicates successful prime editing. The advantage of this design is that the same pegRNA can edit both the reporter and the genomic target. Consequently, these systems can be used to optimise pegRNAs or compare the efficiency of different PEs on a genomic target site without using NGS. Furthermore, the enrichment of edited cells is also feasible using only one pegRNA [51,52]. The enrichment of prime edited cells has only been demonstrated for fluoPEER, which performed similarly to PEAR in HEK293T cells, achieving an average 1.8-fold enrichment compared to transfection marker enrichment [51]. Apart from fluorescent

markers, there are other enrichment strategies based on drug resistance. Levesque et al. developed a method in which the enrichment of editing at an endogenous target site is achieved by co-editing the endogenous Na+/K+ ATPase (ATP1A1), thereby conferring ouabain resistance. Introducing different point mutations into the ATP1A1 and increasing ouabain concentration makes multiple rounds of selection feasible, too. Using this method, they were able to enrich prime editing by up to 83% in K562 cells. While this method does not require an exogenous marker, it should be noted that the enriched cells will carry the ATP1A1 mutations as well as the modifications at the target locus [53].

When compared to the reporter systems just described, PEAR has its own strengths and weaknesses. Unlike PINE-TREE it offers unrestricted flexibility in the target sequence and its flanking nucleotides, however fluoPEER and V-GEAR give even more freedom to the user, as they can be used with genomic segments of up to a few hundred base pairs. Nevertheless, the design of a new reporter plasmid for fluoPEER or V-GEAR can be more complex, because the genomic DNA sequence has to be placed such, that the edit reverts a stop codon or a frameshift mutation. Over the method of Levesque et al., PEAR has the advantage that it does not leave additional genomic modifications behind, so in theory it can be used to enrich the same cell population multiple times.

The development of new PE systems can also expand the targeting scope of prime editing. One such system, proPE, was recently developed in our group [35]. During the development process, PEAR was used in numerous experiments. One of these was the experiment presented in Section 4.4, which I designed to assess the efficiency-decreasing effect of degraded RNAs on PE and proPE (Figure 12). The results showed that: (i) PEAR is sensitive enough to detect the difference in the efficiency-decreasing effect of degraded RNAs; and (ii) proPE is less susceptible to this effect. In addition, I also showed that proPE can rescue the efficiency of PE on target-distal edits exploring genomic targets in the *CYP1A1*, *CYP1A2* and *CYP2B6* genes.

6. Conclusions

- PEAR, the reporter system developed here, exclusively reports on prime editing and is not sensitive to potentially caused indels. However, it is sufficiently sensitive to detect differences in the editing efficiencies of different pegRNA – second nicking sgRNA combinations.
- PEAR can effectively restore various active splice sequences and can be used with different target sequences.
- Experiments with the BEAR-GFP and BEAR-mScarlet plasmids and cell lines have shown that the efficiency of prime editing on the PEAR plasmid is affected by the same factors as in the genomic context.
- PEAR can be applied as a surrogate marker to enrich prime edited cells in multiple cell lines. The PEAR-GFP-2in1 plasmid could enrich the edited population in HEK293T, K562 and U2OS cells, resulting in up to 76% editing, without compromising specificity. Off-target analysis of the HEK293T samples showed that PEAR enrichment only minimally affects the levels of off-target prime editing.
- To minimise the genome-wide off-target effect of the PEAR-GFP-2in1 plasmid, PEAR-GFP-2in1-2.0 was constructed. This plasmid could enrich prime edited HEK293T cells similarly to PEAR-GFP-2in1.
- The inhibitory effect of degraded RNAs on PE and proPE was assessed using the PEAR system. Based on the results, proPE is less affected by the presence of PBS-less and (RTT-PBS)-less RNAs.
- The efficiency-enhancing effect of proPE on target-distal edits was demonstrated by installing naturally occurring SNPs in the *CYP1A1*, *CYP1A2* and *CYP2B6* genes.

7. Summary

Prime editing is a CRISPR tool with great therapeutic potential. However, there are still many questions to be answered before it can be used routinely, including those related to its efficacy and mechanism. The aim of this work was to facilitate the application of prime editing by increasing its efficiency and expanding its targeting scope.

In the first part of the study, PEAR, a flexible fluorescent reporter system was developed to simplify the monitoring of prime editing efficiency in mammalian cells. Due to its versatility, PEAR can be used in multiple ways. Its sensitivity allows it to be used to compare the editing efficiencies of different prime editing setups (e.g. different pegRNA–second nicking sgRNA combinations), or to characterise the performance of different prime editors under the same editing conditions. PEAR experiments are generally less time- and cost-intensive than NGS-based approaches, so this can help to accelerate the development process of new prime editing systems. PEAR can also be applied as an enrichment marker to identify and select individual cells with high prime editing activity. PEAR can efficiently enrich all three main editing categories – insertions, deletions and substitutions – in multiple cell types. In 22 out of the 23 enrichment experiments performed, PEAR enrichment surpassed transfection marker enrichment, increasing the edited cell population by up to 84%.

The second part of the study contributed to the development of a new PE system, called proPE. ProPE can expand the targeting scope of prime editing by alleviating various bottlenecks that can reduce the efficiency of the original PE. This work assessed the effect of degraded RNAs on editing efficiency. It was shown using the PEAR system that degraded RNAs inhibit proPE to a lesser extent than PE. The efficiency enhancing effect of proPE on edits lying outside the target sequence was also evaluated exploring a new set of prime editing targets in the *CYP1A1*, *CYP1A2* and *CYP2B6* genes.

Together, the findings presented in the dissertation could help to promote the more widespread use of PEs by providing means to enhance editing efficiency and study PE variants.

8. References

1. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakatura A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol. 1987;169(12):5429–33.

2. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. Science (1979). 2007 Mar 23;315(5819):1709–12.

3. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008 Dec 19;322(5909):1843–5.

4. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (N Y). 2009 Mar 1;155(3):733–40.

5. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E, Designed EC. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011 Mar 31;471(7340):602–7.

6. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (1979). 2012 Aug 17;337(6096):816–21.

7. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-Guided Human Genome Engineering via Cas9. Science (1979). 2013 Feb 15;339(6121):823–6.

8. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (1979). 2013 Feb 15;339(6121):819–23.

9. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Elife. 2013 Jan 29;2(2).

10. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell. 2014 Feb;156(5):935–49.

11. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. Science (1979). 2014 Mar 14;343(6176):1215–6.

12. Chen PJ, Liu DR. Prime editing for precise and highly versatile genome manipulation. Nature Reviews Genetics 2022 24:3. 2022 Nov 7;24(3):161–77.

13. Nambiar TS, Baudrier L, Billon P, Ciccia A. CRISPR-based genome editing through the lens of DNA repair. Mol Cell. 2022 Jan 20;82(2):348–88.

14. Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han R, Myacheva K, Zwart W, Elkon R, Agami R. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat Biotechnol. 2016 Feb 1;34(2):192–8.

15. Essletzbichler P, Konopka T, Santoro F, Chen D, Gapp B V., Kralovics R, Brummelkamp TR, Nijman SMB, Bürckstümmer T. Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. Genome Res. 2014 Dec 1;24(12):2059–65.

16. Brinkman EK, Chen T, de Haas M, Holland HA, Akhtar W, van Steensel B. Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. Mol Cell. 2018 Jun 7;70(5):801-813.e6.

17. Xue C, Greene EC. DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome Editing. Trends in Genetics. 2021 Jul 1;37(7):639–56.

18. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife. 2014;3:e04766.

19. Aird EJ, Lovendahl KN, St. Martin A, Harris RS, Gordon WR. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. Commun Biol. 2018 Dec 1;1(1).

20. Riesenberg S, Maricic T. Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. Nat Commun. 2018 Dec 4;9(1):2164.

21. Riesenberg S, Chintalapati M, Macak D, Kanis P, Maricic T, Pääbo S. Simultaneous precise editing of multiple genes in human cells. Nucleic Acids Res. 2019 Nov 4;47(19):e116–e116.

22. Riesenberg S, Kanis P, Macak D, Wollny D, Düsterhöft D, Kowalewski J, Helmbrecht N, Maricic T, Pääbo S. Efficient high-precision homology-directed repair-dependent genome editing by HDRobust. Nature Methods 2023 20:9. 2023 Jul 20;20(9):1388–99.

23. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell. 2013 Feb;152(5):1173–83.

24. Komor AC, Kim B, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533.

25. Gaudelli NM, Komor AC, Rees HA, Packer michael S, Badran AH, Bryson DI, Liu DR. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature. 2017 Nov 23;551(7681):464–71.

26. Chen L, Hong M, Luan C, Gao H, Ru G, Guo X, Zhang D, Zhang S, Li C, Wu J, Randolph PB, Sousa AA, Qu C, Zhu Y, Guan Y, Wang L, Liu M, Feng B, Song G, Liu DR, Li D. Adenine transversion editors enable precise, efficient A•T-to-C•G base editing in mammalian cells and embryos. Nature Biotechnology 2023 42:4. 2023 Jun 15;42(4):638–50.

27. Tong H, Wang X, Liu Y, Liu N, Li Y, Luo J, Ma Q, Wu D, Li J, Xu C, Yang H. Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. Nature Biotechnology 2023 41:8. 2023 Jan 9;41(8):1080–4.

28. Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, Grünewald J, Joung JK. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nat Biotechnol. 2020 Jul 20;

29. Anzalone A V., Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149–57.

30. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. Pharmacol Ther. 2013 Apr 1;138(1):103–41.

31. Zhao M, Ma J, Li M, Zhang Y, Jiang B, Zhao X, Huai C, Shen L, Zhang N, He L, Qin S. Cytochrome p450 enzymes and drug metabolism in humans. Int J Mol Sci. 2021 Dec 1;22(23):12808.

32. Kwon YJ, Shin S, Chun YJ. Biological roles of cytochrome P450 1A1, 1A2, and 1B1 enzymes. Archives of Pharmacal Research 2021 44:1. 2021 Jan 23;44(1):63–83.

33. Desta Z, El-Boraie A, Gong L, Somogyi AA, Lauschke VM, Dandara C, Klein K, Miller NA, Klein TE, Tyndale RF, Whirl-Carrillo M, Gaedigk A. PharmVar GeneFocus: CYP2B6. Clin Pharmacol Ther. 2021 Jul 1;110(1):82–97.

34. Simon DA, Tálas A, Kulcsár PI, Biczók Z, Krausz SL, Várady G, Welker E. PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells. Elife. 2022 Feb 1;11.

35. Krausz SL, Simon DA, Bartos Z, Biczók Z, Varga É, Huszár K, Kulcsár PI, Tálas A, Ligeti Z, Welker E. ProPE expands the prime editing window and enhances gene editing efficiency where prime editing is inefficient. Nature Catalysis 2025. 2025 Oct 10;1–17.

36. 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.

37. Katti A, Foronda M, Zimmerman J, Diaz B, Zafra MP, Goswami S, Dow LE. GO: a functional reporter system to identify and enrich base editing activity. Nucleic Acids Res. 2020;48(6):2841–52.

38. Yan N, Sun Y, Fang Y, Deng J, Mu L, Xu K, Mymryk JS, Zhang Z. A Universal Surrogate Reporter for Efficient Enrichment of CRISPR/Cas9-Mediated Homology-Directed Repair in Mammalian Cells. Mol Ther Nucleic Acids. 2020 Mar 6;19:775–89.

39. Trojan J, Zeuzem S, Randolph A, Hemmerle C, Brieger A, Raedle J, Plotz G, Jiricny J, Marra G. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. Gastroenterology. 2002 Jan 1;122(1):211–9.

40. Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen PF, Chen C, Nelson JW, Newby GA, Sahin M, Osborn MJ, Weissman JS, Adamson B, Liu DR. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell. 2021 Oct 28;184(22):5635-5652.e29.

41. Kulcsár PI, Tálas A, Tóth E, Nyeste A, Ligeti Z, Welker Z, Welker E. Blackjack mutations improve the on-target activities of increased fidelity variants of SpCas9 with 5′G-extended sgRNAs. Nature Communications 2020 11:1. 2020 Mar 6;11(1):1–14.

42. Nelson JW, Randolph PB, Shen SP, Everette KA, Chen PJ, Anzalone A V., An M, Newby GA, Chen JC, Hsu A, Liu DR. Engineered pegRNAs improve prime editing efficiency. Nat Biotechnol. 2022 Mar 1;40(3):402–10.

43. Ma H, Tu LC, Naseri A, Huisman M, Zhang S, Grunwald D, Pederson T. CRISPR-Cas9 nuclear dynamics and target recognition in living cells. Journal of Cell Biology. 2016 Aug 29;214(5):529–37.

44. Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci Adv. 2017 Aug 1;3(8).

45. 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.

46. Yu G, Kim HK, Park J, Kwak H, Cheong Y, Kim D, Kim J, Kim J, Kim HH. Prediction of efficiencies for diverse prime editing systems in multiple cell types. Cell. 2023 May 11;186(10):2256–72.

47. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134.

48. Sürün D, Schneider A, Mircetic J, Neumann K, Lansing F, Paszkowskirogacz M, Hänchen V, Leekirsch MA, Buchholz F. Efficient generation and correction of mutations in human iPS cells utilizing mRNAs of CRISPR base editors and prime editors. Genes (Basel). 2020 May 1;11(5).

49. Lin Q, Zong Y, Xue C, Wang S, Jin S, Zhu Z, Wang Y, Anzalone A V., Raguram A, Doman JL, Liu DR, Gao C. Prime genome editing in rice and wheat. Nat Biotechnol. 2020 May 16;38(5):582–5.

50. Frisch C, Kostes WW, Galyon B, Whitman B, Tekel SJ, Standage-Beier K, Srinivasan G, Wang X, Brafman DA. PINE-TREE enables highly efficient genetic modification of human cell lines. Mol Ther Nucleic Acids. 2023 Sep 12;33:483–92.

51. Schene IF, Joore IP, Baijens JHL, Stevelink R, Kok G, Shehata S, Ilcken EF, Nieuwenhuis ECM, Bolhuis DP, van Rees RCM, Spelier SA, van der Doef HPJ, Beekman JM, Houwen RHJ, Nieuwenhuis EES, Fuchs SA. Mutation-specific reporter for optimization and enrichment of prime editing. Nature Communications 2022 13:1. 2022 Mar 1;13(1):1–10.

52. Kleinboehl EW, Laoharawee K, Lahr WS, Jensen JD, Peterson JJ, Bell JB, Webber BR, Moriarity BS. Development and testing of a versatile genome editing application reporter (V-GEAR) system. Mol Ther Methods Clin Dev. 2024 Jun 13;32(2):101253.

53. Levesque S, Mayorga D, Fiset JP, Goupil C, Duringer A, Loiselle A, Bouchard E, Agudelo D, Doyon Y. Marker-free co-selection for successive rounds of prime

editing in human cells. Nature Communications 2022 13:1. 2022 Oct 7;13(1):1–14.

9. Bibliography of the candidate's publications

9.1 Publications related to the thesis

- S. L. Krausz, D. A. Simon, Z. Bartos, Z. Biczók, É. Varga, K. Huszár, P. I. Kulcsár, A. Tálas, Z. Ligeti, and E. Welker, "ProPE expands the prime editing window and enhances gene editing efficiency where prime editing is inefficient," *NATURE CATALYSIS*, vol. 8, pp. 1100–1116, 2025.
- D. A. Simon, A. Tálas, P. I. Kulcsár, Z. Biczók, S. L. Krausz, G. Várady, and E. Welker, "PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells," *ELIFE*, vol. 11, 2022.
- D. Simon, A. Tálas, P. I. Kulcsár, G. Várady, Z. Biczók, and E. Welker, "PEAR: a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells," in *Hungarian Molecular Life Sciences 2021*, 2021, p. 201., poster
- D. Simon, A. Tálas, P. I. Kulcsár, and E. Welker, "PEAR: A FLEXIBLE FLUORESCENT REPORTER FOR THE IDENTIFICATION AND ENRICHMENT OF SUCCESSFULLY PRIME EDITED CELLS." GENOME ENGINEERING: CRISPR FRONTIERS 2021 [Online conference, international], poster

9.2 Publications not related to the thesis

- Z. Bartos, Z. Biczók, D. A. Simon, B. Csoma, A. Tálas, P. Tóth, B. Fehér, and E. Welker, "GRAND-PLUS enables the insertion of longer sequences with base pair precision." 2025. Genome Engineering: CRISPR Frontiers 2025-08-12 [Cold Spring Harbor (NY), US], poster
- D. Simon and A. Tálas, "BEAR reveals that increased fidelity variants can successfully reduce the mismatch tolerance of adenine but not cytosine base editors," in *Abstract book of Annual Meeting of the Hungarian Biochemical Society 2022*, 2022, p. 5., talk
- Tálas, D. A. Simon, P. I. Kulcsár, É. Varga, S. L. Krausz, and E. Welker, "BEAR reveals that increased fidelity variants can successfully reduce the mismatch tolerance of adenine but not cytosine base editors," *NATURE COMMUNICATIONS*, vol. 12, no. 1, 2021.

- D. Simon, A. Tálas, and E. Welker, “A Cas9 endonukleáz szabályzása a RUSH rendszer segítségével,” in *XIII. Szent-Györgyi Albert Konferencia - Konferenciakiadvány*, 2019, p. 40., talk
- D. Simon, “A Cas9 endonukleáz szabályozása a RUSH rendszer segítségével.” 2018. 8th Gárdos György Szimpózium 2018 [Mátraháza, Magyarország], talk

10. Acknowledgements

I would like to thank my supervisor Dr. Ervin Welker for his help, support and encouragement throughout my PhD studies.

I would like to thank my BSc and MSc supervisor Dr. András Tálas for teaching me most of the techniques I used in my thesis work; and for his help and mentorship as co-first author in the publication of the PEAR system.

I would like to thank and highlight the work of my co-authors and colleagues, whose work I could present in my dissertation to elevate its standards. Dr. András Tálas and Dr. Péter István Kulcsár helped to clone 51 of the 107 plasmids used in the PEAR publication. Dr. Péter István Kulcsár also did the nucleofection and maintenance of the K562 and U2OS cells. Dr. György Várady patiently sorted the cells for all the PEAR enrichment experiments. Sarah Laura Krausz performed bioinformatic calculations for the PEAR NGS experiments and conceived the idea of proPE. Last, but not least Viktória Faragó designed the schematic figures for proPE. Greatest thanks to Ildikó Szűcsné Pulinka, Judit Szűcs, Vivien Karl, Viktória Faragó and Veronika Csonka for their excellent technical assistance.

I am grateful to every member of the Welker group for the supportive and friendly work atmosphere and for laughing and/or crying together as the experimental results required.

I am grateful for the help and support of my family, especially my husband and my mother. And for the fact that my daughter just loves nursery, so I could focus on my thesis with a light heart.

I am thankful for the grant support of the Cooperative Doctoral Programme Doctoral Scholarship provided by the Ministry of Culture and Innovation of Hungary through the National Research, Development and Innovation Fund (KDP-2021_C1773316).

11. Appendix

Table 1: List of NGS primers used in the study

Primer ID	Primer sequence (5'→3')
HEK3_for-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGTGGGCTGCCTAGAAAGG
HEK3_rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCAGCCAAACTTGTCAACC
RNF2_for-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACGTCTCATATGCCCCTTGG
RNF2_rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACGTAGGAATTTTGGTGGGACA
FANCF2-for-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTGCTGACGTAGGTAGTGC
FANCF2-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACACGGATAAAGACGCTGGG
EMX1_for-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGCTCAGCCTGAGTGTTGA
EMX1_rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCGTGGGTTTGTGGTTGC
HEK4_for-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAACCCAGGTAGCCAGAGAC
HEK4_rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTTTCAACCCGAACGGAG
HEK3_OT1-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCCTGTTGACCTGGAGAA
HEK3_OT1-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACTGTACTTGCCCTGACCA
HEK3_OT2-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGGTGTTGACAGGGAGCAA
HEK3_OT2-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGAGATGTGGGCAGAAGGG
HEK3_OT3-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAGAGGGAACAGAAGGGCT
HEK3_OT3-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCCAAAGGCCCAAGAACCT
HEK3_OT4-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCTAGCACTTTGGAAGGTCG
HEK3_OT4-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTCATCTTAATCTGCTCAGCC
HEK4_OT1-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCATGGCTTCTGAGACTCA
HEK4_OT1-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCTCCCTTGCACTCCCTGTCTTT
HEK4_OT2-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTGGCAATGGAGGCATTGG

HEK4_OT2-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAAGAGGCTGCCCATGAGAG
HEK4_OT3-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTCTGAGGCTCGAATCCTG
HEK4_OT3-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGTGGCCTCCATATCCCTG
HEK4_OT4-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTCCACCAGAACTCAGCCC
HEK4_OT4-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTCGGTTCCTCCACAACAC
EMX1_OT1-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGGGGAGATTTGCATCTGTGGAGG
EMX1_OT1-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTTTTATACCATCTTGGGGTTACAG
EMX1_OT2-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAATGTGCTTCAACCCATCACGGC
EMX1_OT2-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCATGAATTTGTGATGGATGCAGTCTG
EMX1_OT3-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGAAGGAGGTGCAGGAGCTAGAC
EMX1_OT3-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATCCCGACCTTCATCCCTCCTGG
FANCF_OT1-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCGGGCAGTGGCGTCTTAGTCG
FANCF_OT1-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCTGGGTTTGGTTGGCTGCTC
FANCF_OT2-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCCTTGCCGCCCAGCCGGTC
FANCF_OT2-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACTGGGGAAGAGGCGAGGACAC
FANCF_OT3-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGTGTTTCCCATCCCCAACAC
FANCF_OT3-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAATGGATCCCCCCCTAGAGCTC
FANCF_OT4-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGGCCCACAGGTCCTTCTGGA
FANCF_OT4-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCACACGGAAGGCTGACCACG
HBB_ON-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGGTTGGCCAATCTACTCCC
HBB_ON-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCTTCTCTGTCTCCACATGCC
PRNP_ON-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCAGTGGAACAAGCCGAGT
PRNP_ON-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTTGGTTGGGGTAACGGTG

HEXA_ON-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATACAGGTGTGGCGAGAGG
HEXA_ON-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGCCTCCTTTGGTTAGCA
DNMT_ON-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACAACAGCTTCATGTCAGCC
DNMT_ON-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACGTTAATGTTTCCTGATGGTCC
RUNX1-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGTGAGGCTGAAACAGTGACC
RUNX1-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGAACTGGCAGGCACCGAGG
CYP1A1-ON-fwd-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGGTCCCCAAAGGCCTGAAGAAT
CYP1A1-ON-rev-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGGGTGAAGGTGTAGAGGTCG
CYP1A2-ON-fwd-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGTCCCCAAAGGCCTGAAAAGT
CYP1A2-ON-rev-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGAGGGTGGAGGTGTAGAGGTCAG
CYP2B6-ON-fwd1-i5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAGACCAGGACCATGGAACTCAGC
CYP2B6-ON-rev1-i7	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAAGCTTCCCCAAGTACCAAGGC

Table 2: Detailed information on the *CYP* gene edits

SNP ID	gene	PBS (nt)	RTT (nt)	edit pos.	engRNA / pegRNA spacer (5'->3')	tpgRNA spacer (5'->3')	tpgRNA / pegRNA RTT-PBS (5'->3')	second nicking sgRNA "A" spacer (5'->3')	second nicking sgRNA "B" spacer (5'->3')
rs34883432	CYP2B6	13	41	31	gagggcggggccctggtggg	atggatagaag	ctcctggttctgcgccaccctaacacccatgaccgcctccaccagggccccgc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
		17	44	34	gcagagggcggggccctggt	atggatagaag	ctcctggttctgcgccaccctaacacccatgaccgcctccaccagggccccgccctctgc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
rs8192709	CYP2B6	17	39	29	gagggcggggccctggtggg	atggatagaag	cctggttcagtgccaccctaacacccatgaccgcctccaccagggccccgccctc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
		13	42	32	gcagagggcggggccctggt	atggatagaag	cctggttcagtgccaccctaacacccatgaccgcctccaccagggccccgccct	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
rs33973337	CYP2B6	17	26	17	gagggcggggccctggtggg	atggatagaag	caccctaactcccatgaccgcctcccaccagggccccgccctc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
		13	30	21	ggcagagggcggggccctgg	atggatagaag	caccctaactcccatgaccgcctcccaccagggccccgccctc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
rs33980385	CYP2B6	17	19	13	gcagagggcggggccctggt	atggatagaag	cccatggccgcctcccaccagggccccgccctctgc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
		13	23	17	aggggcagagggcggggccc	atggatagaag	cccatggccgcctcccaccagggccccgccctctgc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct

SNP ID	gene	PBS (nt)	RTT (nt)	edit pos.	engRNA / pegRNA spacer (5'->3')	tpgRNA spacer (5'->3')	tpgRNA / pegRNA RTT-PBS (5'->3')	second nicking sgRNA "A" spacer (5'->3')	second nicking sgRNA "B" spacer (5'->3')
rs33926104	CYP2B6	13	16	8	gagggcggggccctggtggg	atggatagaag	cccatgacagcctcccacc agggccccgc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
		17	20	12	ggcagagggcggggccctgg	atggatagaag	cccatgacagcctcccacc agggccccgccctctgcc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
rs34284776	CYP2B6	17	19	10	gcagagggcggggccctggt	atggatagaag	cccatgaccacctcccacca gggccccgccctctgc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
		13	23	14	aggggcagagggcggggccc	atggatagaag	cccatgaccacctcccacca gggccccgccctctgc	gcctactcaaatcctttctg	gggtgagctgcttcttgcct
rs35303484	CYP2B6	17	36	21	gccccgccctctgcccctttt	atcatgggtgtt	ggcctcttctatccaactgca gaaggtttcccaaaagggg cagagggcggggc	gctgtgaggagtgcaaggaag	-
		13	36	21	gccccgccctctgcccctttt	gggtgttaggg	ggcctcttctatccaactgca gaaggtttcccaaaagggg cagagggcg	gctgtgaggagtgcaaggaag	-
rs35035798	CYP1A1	17	16	7	acccgcacctggcactgtca	atgcccaatcag	gctggctcacccttgacagt gccaggtgcgggt	gggcctgccggatggtgtcc	-
		13	16	7	acccgcacctggcactgtca	ggcctctgatt	gctggctcacccttgacagt gccaggtgc	gggcctgccggatggtgtcc	-
rs17861094	CYP1A1	13	49	37	gggcctgccggatggtgtcc	gatgatttcaa	tgcagatccgaactggctcc acacccgtggtggtgctga gcggcctggacaccatccg gcag	acccgcacctggcactgtca	gcatggggctggcctctgatt
		17	49	37	gggcctgccggatggtgtcc	acgatgatttca	tgcagatccgaactggctcc acacccgtggtggtgctga gcggcctggacaccatccg gcaggccc	acccgcacctggcactgtca	gcatggggctggcctctgatt

SNP ID	gene	PBS (nt)	RTT (nt)	edit pos.	engRNA / pegRNA spacer (5'->3')	tpgRNA spacer (5'->3')	tpgRNA / pegRNA RTT-PBS (5'->3')	second nicking sgRNA "A" spacer (5'->3')	second nicking sgRNA "B" spacer (5'->3')
rs45565238	CYP1A2	17	29	22	acccgcacctggcactgtca	ggcccttgctc	acgtccctgtagcgctggctcatccttgacagtgccaggtgcgggt	gggcctgccggatggtgtcc	-
		13	29	22	acccgcacctggcactgtca	atgcccgagca	acgtccctgtagcgctggctcatccttgacagtgccaggtgc	gggcctgccggatggtgtcc	-
rs138652540	CYP1A2	17	38	28	gggcctgccggatggtgtcc	gacgatttcaa	attggctccatgcccgtgctggtgctgagccgcctggacaccatccggcaggccc	acccgcacctggcactgtca	gcatggggctggcccttgctc
		13	38	28	gggcctgccggatggtgtcc	acgacgatttca	attggctccatgcccgtgctggtgctgagccgcctggacaccatccggcag	acccgcacctggcactgtca	gcatggggctggcccttgctc
rs34067076	CYP1A2	13	31	21	ggacaccatccggcaggccc	atggagccaatg	cccttgaaattgtcgccctgccgcaccagggcctgccggatggt	ggctcatccttgacagtgcc	-
		17	31	21	ggacaccatccggcaggccc	gtgcggatctgc	cccttgaaattgtcgccctgccgcaccagggcctgccggatggtgtcc	ggctcatccttgacagtgcc	-