Deletion and replacement of long genomic sequences using prime editing

Tingting Jiang[1], **Xiao-Ou Zhang**[2,3], **Zhiping Weng**[2], **Wen Xue**[1,4,5,6,*]

[1]RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

[2]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA

[3]School of Life Sciences and Technology, Tongji University, Shanghai 200092, China

[4]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

[5]Department of Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

[6]Li Weibo Institute for Rare Diseases Research, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA, 01605

Abstract

Genomic insertions, duplications and insertion/deletions (indels), which account for ~14% of human pathogenic mutations, cannot be accurately or efficiently corrected by current gene editing methods, especially those that involve larger alterations (>100 bp). Here, we optimize prime editing tools for creating precise genomic deletions and direct the replacement of a genomic fragment ranging from ~1-kb to ~10-kb with a desired sequence in the absence of an exogenous DNA template. By conjugating Cas9 nuclease to reverse transcriptase (PE-Cas9) and combining it with two prime editing guide RNAs (pegRNAs) targeting complementary DNA strands, we achieve precise and specific deletion and repair (PEDAR) of target sequences. PEDAR outperformed other genome editing methods in a reporter system and at endogenous loci, efficiently creating large and precise genomic alterations. In a mouse model of tyrosinemia, PEDAR removed a 1.38-kb pathogenic insertion within the *Fah* gene and precisely repaired the deletion junction to restore FAH expression in liver.

Editorial summary

Prime editing is expanded to deletions and replacements of genomic sequences of up to 10kb.

Genetic insertions, duplications, and indels (<u>ins</u>ertion/<u>del</u>etion) account for ~14% of 60,008 known human pathogenic variants[1] (Fig. 1a). Many of these abnormal insertions and duplications involve larger DNA fragments (>100 bp). Indeed, retrotransposon element insertions, ranging from 163 to 6000 bp[2, 3], disrupt the normal expression and function of genes[4] thereby causing genetic diseases like cystic fibrosis, hemophilia A, X-linked dystonia-parkinsonism, and inherited cancers[4–7]. Precise genome editing technologies that simultaneously delete the inserted or duplicated DNA sequences and repair the disrupted genomic site might provide a way to treat a wide range of diseases.

The CRISPR/Cas9 system is a powerful gene editing tool for correcting pervasive pathogenic gene mutations. Using dual single guide RNAs (sgRNA), Cas9 can induce two double-strand breaks (DSBs). The two cut ends are then ligated through the non-homologous end joining (NHEJ) repair pathway, leading to 5-Mb target fragment deletion in vitro[8, 9] and in vivo[10–12]. However, the random indels generated by NHEJ lower the editing accuracy of this method. When a donor DNA template is present, CRISPR/Cas9 can insert a desired sequence at the cut site to more accurately repair the deletion junction through homology directed repair (HDR)[13, 14]. Nevertheless, the repair efficiency of CRISPR-mediated HDR is hindered by the exogenous DNA donor and is limited in post-mitotic cells[15, 16]. To further expand the gene editing toolbox, a novel CRISPR-associated gene editor – called prime editing (PE)[17] – was developed by conjugating an engineered reverse transcriptase (RT) to a catalytically-impaired Cas9 'nickase' (Cas9[H840A]) that cleaves only one DNA strand. An extension at the 3' end of the prime editing guide RNA (pegRNA) functions as an RT template, allowing the nicked site to be precisely repaired[17, 18]. Thus, PE can mediate small deletion, insertion, and base editing without creating DSBs or requiring donor DNA[17], and holds great promise for correcting human genetic diseases[19–22]. Yet, PE has not been applied to delete larger DNA fragments. Here, we engineer a <u>PE</u>-Cas9-based <u>d</u>eletion <u>a</u>nd <u>r</u>epair (PEDAR) method enabling accurate deletion of a larger genomic fragment and concurrent insertion of a desired sequence without requiring a repair template.

Results

PEDAR strategy

To achieve accurate and efficient large fragment deletion and simultaneous insertion without requiring a DNA template (Supplementary Fig. 1a), we modified the prime editing system to employ a pair of pegRNAs (hereafter referred to as pegF and pegR) rather than one pegRNA and one nicking guide RNA. We reasoned that using two pegRNAs would enable concurrent targeting of both DNA strands. The 3' extension of each pegRNA is a reverse-complementary RT template, which encodes the sequences for desired insertion. In theory, this newly-engineered system could mediate accurate deletion-repair through the following steps (Fig. 1b, left side): (i) prime editor recognizes the 'NGG' PAM sequence, binds, and nicks the two complementary strands of DNA on either side of the large fragment[8]; (ii) the desired insertion sequences are reverse transcribed into the target site using the RT template linked to the pegRNAs; (iii) the complementary DNA strands containing the edits are annealed; (iv) the original DNA strands (i.e., 5' flaps) are excised; and (v) DNA

is repaired. However, Cas9 nickase cannot effectively mediate larger target deletions with paired guide RNAs[23, 24]. Indeed, PE applications reported in the literature are limited to programing deletions of less than 100 bp, raising the concern that PE cannot generate long genomic deletions[18].

Fully active Cas9 nuclease has been used to program larger deletions with dual sgRNAs[14]. Therefore, we conjugated an active Cas9 nuclease, instead of Cas9 nickase, to the RT[17] to create "PE-Cas9" (Supplementary Fig. 1b). With a single pegRNA[17], PE-Cas9 and PE generated similar rates of 3-bp CTT insertion at the cut/nicking site of an endogenous locus (Supplementary Fig. 1c), indicating that Cas9 nuclease activity does not affect prime editing efficiency. We hypothesized that, with the guidance of two pegRNAs targeting both complementary strands of DNA, PE-Cas9 can introduce two DSBs and delete the intervening DNA fragment between the two cut sites. Concurrently, the desired edits are incorporated at target sites using the RT template at the 3' extension of the pegRNAs. The two complementary edits then function as homologous sequences to direct the ligation and repair of the deletion junction. We term this method "PE-Cas9-based deletion and repair" or PEDAR (Fig. 1b, right side).

Compare PE, Cas9, and PEDAR in programming deletion-insertion

We compared the efficiency of PEDAR, PE, and Cas9 systems in coupling large target deletion and accurate insertion at the endogenous *HEK3* genomic locus in HEK293T cells. We designed two pegRNAs with an offset of 979 bp (distance between the two 'NGG' PAM sequences) to program a 991bp-deletion/18bp-insertion at the *HEK3* site. The RT template at the 3' extension of the pegRNAs encodes an I-SceI recognition sequence (18-bp), which will be reversed transcribed and integrated into the target site (Supplementary Fig.1d). Paired pegRNAs along with PE, PE-Cas9, or Cas9 were transfected into cells. Delivery of PE-Cas9 with or without single pegRNA was used as a negative control. Three days post transfection, we amplified the target genomic site and found that the treatment with either PE-Cas9 or active Cas9, but not PE, led to a ~450-bp deletion amplicon. This amplicon was ~1-kb shorter than the amplicon without deletion (Fig. 1c). We digested the deletion amplicon with I-SceI endonuclease, and observed that only the PE-Cas9-treated group showed cut bands of expected size (~251bp and ~199bp), indicating insertion of the I-SceI recognition sequence (Fig. 1d). Using real-time quantitative PCR, we found that PE-Cas9 generates an accurate deletion-insertion frequency of 2.67±0.839% in total genomic DNA, whereas Cas9 seldom generated accurate editing (0.0112±0.00717%, Fig. 1e). To further verify editing accuracy, we purified the deletion amplicon (~450-bp band in Fig.1c) and performed deep sequencing analysis. We found that PE-Cas9 mediates 27.0±1.83% accurate editing of total deletion events (Fig. 1f). Taken together, our findings suggest that, PEDAR outperforms prime editing and Cas9 editing in programming accurate large fragment deletion and simultaneous insertion.

PEDAR also generated unintended edits, classified as: (i) other deletion/insertion, including direct deletion without insertion and imperfect deletion-insertion, and (ii) small indels generated by individual pegRNA at the two cut sites, hereafter referred to as cut site_F and cut site_R. We measured the incidence of these events in total genomic DNA by

real-time quantitative PCR, and observed that PE-Cas9 and Cas9 generated comparable rates of unintended edits (Fig. 1e). Of all the deletion events, PE-Cas9 generated 38.0±4.15% imperfect deletion-insertions caused by imprecise DNA repair or pegRNA scaffold insertion[17] and a significantly lower rate of direct deletion without insertion than that mediated by active Cas9 (35.0±4.80% and 88.8±1.58%, respectively) (Fig. 1f). PE-Cas9-mediated unintended deletion edits with the highest sequencing reads are listed in Supplementary Fig. 2a and Supplementary Table 3. PE-Cas9 or Cas9 also introduced indels at the two cut sites without generating the desired deletion. Sanger sequencing of the amplicon without deletion (~1.4-kb band in Fig.1c) reveals no significant difference in small indels caused by PE-Cas9 and Cas9 (Supplementary Fig. 2b).

To explore the potential repair mechanism underlying PEDAR-mediated editing, we delivered PE-Cas9 with one pegRNA and one sgRNA targeting the *HEK3* locus into cells. And PE-Cas9 with paired pegRNAs serves as a positive control (Supplementary Fig. 2c). Although PE-Cas9 generated a ~450-bp deletion amplicon using one pegRNA and one sgRNA (Supplementary Fig. 2d), this amplicon failed to be digested into two distinct bands by I-Sce1 endonuclease (Supplementary Fig. 2e). Deep sequencing revealed that minimal accurate deletion-insertion (0.716±0.0868%) in the cells transfected with one pegRNA and one sgRNA, as compared to a 26.5±1.12% accurate editing rate in the cells treated with PEDAR (Supplementary Fig. 2f). This result demonstrates that the reverse-complementary sequences introduced by paired pegRNAs at the two cut sites are essential for directing accurate repair, resembling the annealing and ligation process in the MMEJ or SSA repair pathway[25, 26].

We also investigated how design of the pegRNAs, namely the length of the primer binding site (PBS) and the design of RT template, might affect editing efficiency of PEDAR. Our original PEDAR system used paired pegRNAs with 13-nt PBS. We designed two additional paired pegRNAs with 10-nt or 25-nt PBS targeting the *HEK3* locus as comparisons. Although all paired pegRNAs supported ~1-kb deletion (Supplementary Fig. 3a) and simultaneous insertion of the I-Sce1 recognition sequence (Supplementary Fig. 3b), the shorter and longer PBS lengths significantly impaired the accurate editing rate identified by deep sequencing (Supplementary Fig. 3c). To determine the effect of RT template design on editing efficiency, we designed an alternative pegRNA (pegRNA_alt) – similar to the pegRNA used in PE2[17]– by extending the RT template with a 14-nt sequence homologous to the region after the other cut site (Supplementary Fig. 3d). After transfecting the newly-designed paired pegRNAs with PE or PE-Cas9 into cells, we identified a deletion amplicon of the expected size (Supplementary Fig. 3e), and insertion of I-Sce1 recognition sequence was detected in the deletion amplicon (Supplementary Fig. 3f). Deep sequencing of the deletion amplicon reveals that pegRNA_alt significantly decreased PE-Cas9-mediated accurate editing rate compared to the original pegRNAs (Supplementary Fig. 3g). Surprisingly, co-transfection of PE and pegRNA_alt greatly improved the purity of deletion product (85.9±0.644% accurate editing in deletion amplicon, Supplementary Fig. 3g). However, the absolute accurate editing rate in total genomic DNA was comparable between PE/pegRNA_alt and PE-Cas9/pegRNA groups (Supplementary Fig. 3h), potentially due to the limited ability of Cas9 nickase to introduce larger deletion[23, 24]. Based on the collective findings, we elected to use pegRNAs with a 13-nt PBS and an RT template

without adding the sequence homologous to target site after incision in the subsequent studies.

To assess the efficiency of PEDAR-mediated deletion-insertion at endogenous locus other than *HEK3* site, we targeted *DYRK1* locus for deleting a 995-bp DNA fragment and simultaneously inserting I-Sce1 recognition sequence. Treatment of HEK293T cells with PEDAR could lead to a ~507-bp deletion band (Supplementary Fig. 4a), and this amplified product could be digested by I-Sce1 endonuclease (Supplementary Fig. 4b). Deep sequencing of the deletion amplicon identified a 2.18±0.552% accurate editing efficiency (Supplementary Fig. 4c). We reasoned that the low G/C contents at the primer binding sequences of the two pegRNAs targeting *DYRK1* locus (23% of pegF and 31% of pegR) restricted the integration of the desired DNA fragment, which is consistent with a report showing poor PE efficiency when the GC content in PBS is less than 30%[27].

PEDAR enables larger deletion and insertion

To further understand the robustness of the PEDAR system, we explored its limits with respect to insertion size and deletion size. First, we set out to insert the I-Sce1 recognition sequence together with either Flag epitope tag (44bp total) or Cre recombinase *LoxP* site (60bp total) into the *HEK3* locus after deletion of a ~1-kb DNA fragment. Two paired pegRNAs were designed with either a 44-nt RT template or a 60-nt RT template, and the pegRNAs with 18-nt RT template serve for comparison (Fig. 2a). For all paired pegRNAs, PE-Cas9 generated the expected deletion (Fig. 2b) and inserted the desired sequence (Fig.2c) at the target site in cells. Deep sequencing revealed 13.7±1.51% (44bp-insertion) and 12.4±2.88% (60bp-insertion) accurate deletion-insertion rates within total deletion edits, which are significantly lower than the 22.6±0.267% accurate editing efficiency of PE-Cas9 when inserting a shorter sequence (18bp) (Fig.2d). To investigate the maximum deletion size generated by PEDAR, we designed two distinct paired pegRNAs with an offset of ~8kb or ~10kb to target the *CDC42* locus (Fig.2e). Using the indicated primers to amplify the corresponding target site, we observed the expected deletion amplicon (Fig.2f). After I-Sce1 endonuclease treatment, two digested bands were detected in the PE-Cas9-treated group (Fig.2g). Deep sequencing revealed 18.4±2.07% (8kb-del/18bp-ins) and 6.97±1.00% (10kb-del/18bp-ins) accurate deletion-insertion rates within the deletion amplicon (Fig.2h). In all, these data demonstrate the robustness and flexibility of PE-Cas9 in generating >10-kb larger deletion and up to 60-bp insertion.

PEDAR restores gene expression by programming in-frame deletion

Next, we asked whether PEDAR could generate large in-frame deletions and accurately repair genomic coding regions to restore gene expression. To answer this question, we used a HEK293T traffic light reporter (TLR) cell line[28, 29], which contains a GFP sequence with an insertion and an mCherry sequence separated by a T2A (2A self-cleaving peptides) sequence. The disrupted GFP sequence causes a frameshift that prevents mCherry expression (Fig. 3a). We hypothesized that PEDAR could restore mCherry signal by accurately deleting the disrupted GFP and T2A sequence (~800 bp in length). We designed two pegRNAs targeting the promoter region before the start codon of GFP and the site immediately after T2A, respectively. In this approach, part of the Kozak sequence and start

codon are unintentionally deleted due to the restriction of the PAM sequence. However, we designed the RT template at the 3' end of pegRNAs to encode the Kozak sequence and start codon to ensure their insertion into the target site by reverse transcription (Fig. 3a).

We treated TLR reporter cells with dual pegRNAs (pegF+pegR) and either PE-Cas9, PE, or Cas9, and used flow cytometry to assess the mCherry signal. The frequency of mCherry positive cells was significantly higher in the PE-Cas9-treated group (2.12±0.105%) compared to PE- or Cas9-treated groups (Fig. 3b and Supplementary Fig. 5a, b). The mCherry positive cell rate was limited in all three replicates, likely because the cleavage efficiency of pegRNA at cut site_R (pegR) are very low (~1.8%; Supplementary Fig. 5c). Thus, we designed another pegRNA (pegR2) with a ~10.3% cleavage rate (Fig. 3a and Supplementary Fig. 5c) and assessed its efficiency in restoring mCherry expression. Indeed, the newly-designed paired pegRNAs significantly improved the mCherry positive cell rate (2.99±0.166%, Fig. 3b and Supplementary Fig. 5d). Alternatively, to enhance the editing rate, we explored the possibility of improving the expression level of gene editing agents in cells[30]. Co-transfection of cells with a fluorescent protein-expressing plasmid, followed by FACS sorting, would enrich for cells with high levels of transgene expression[31, 32]. Thus, a GFP-expressing plasmid was co-transfected with PE-Cas9 and paired pegRNAs into TLR cells as an indicator of transfection efficiency. We observed a ~1.42-fold increase in mCherry positive cell rate after selection of cells with high GFP expression (Fig. 3c and Supplementary Fig. 5e). These results indicate that the editing efficiency of PEDAR largely relies on the efficiency of pegRNA and the expression level of gene editing components. To verify that PEDAR restored mCherry expression via accurate deletion-insertion, we sorted mCherry positive cells in PE-Cas9-treated groups and amplified the target sequence. In all three replicates, we detected a deletion amplicon that is ~800-bp shorter than the amplicon in untreated control cells (Fig. 3d). Further, we assessed the accurate editing rate by deep sequencing analysis of the ~300-bp deletion amplicon. The results revealed a 16.2±2.58% accurate deletion-insertion rate (Fig. 3e). The most common imperfect editing event across the three replicates restores mCherry open reading frame but the inserted sequence lacks three nucleotides compared to the intended insertion (Supplementary Fig. 5f). These data demonstrate that PEDAR can repair genomic coding regions that are disrupted by large insertions.

PEDAR corrects the disrupted *Fah* gene *in vivo*

Furthermore, to test the in vivo application of PEDAR, we utilized a Tyrosinemia I mouse model, referred to as Fah [Exon5]. This Tyrosinemia I model is derived by replacing a 19-bp sequence with a ~1.3-kb neo expression cassette[33] at exon 5 of the *Fah* gene[34] (Fig. 4a). This insertion disrupts the *Fah* gene to cause FAH protein deficiency and liver damage. To maintain body weight and survival, these mice are given water supplemented with NTBC [2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione], a tyrosine catabolic pathway inhibitor. We hypothesized that PEDAR can correct the causative Fah [Exon5] mutation by deleting the large insertion and simultaneously inserting the 19-bp fragment back to repair exon 5 (Fig. 4b). We engineered two pegRNAs targeting the genomic region before and after the inserted neo expression cassette, respectively. At the 3' end of pegRNAs, a 22-bp RT template encoding the deletion fragment (19bp) plus a 3-bp sequence that is unintentionally

deleted by PE-Cas9 was designed. PE-Cas9 and the two pegRNAs were delivered to the livers of mice (n=4) via hydrodynamic injection. Mice (n=2) treated with Cas9/pegRNAs serve as negative control. Mice were kept on NTBC water after treatment. One week later, the mice were euthanized, and immunochemical staining was performed on liver sections with FAH antibody. We detected FAH-expressing hepatocytes on PE-Cas9-treated liver sections (Fig. 4c), with a 0.76±0.25% correction rate (Fig. 4d). FAH expression was not detected in Cas9-treated mouse liver (Fig. 4c).

Hepatocytes with corrected FAH protein will gain a growth advantage and eventually repopulate the liver[35]. Therefore, we delivered PE-Cas9 and the two pegRNAs via hydrodynamic injection to mice (n=4) and subsequently removed the NTBC supplement to allow repopulation. Untreated Fah [Exon5] mice (on or off NTBC water) were used as controls. Forty days later, widespread FAH patches were observed in PE-Cas9-treated mouse liver sections, and the corrected hepatocytes showed normal morphology (Fig. 4e, Supplementary Fig. 6a). To understand the editing events in mouse liver, we amplified the target site by using PCR primers spanning exon 5, and identified the ~300-bp deletion amplicon in treated mice, indicating deletion of the ~1.3-kb insertion fragment (Fig. 4f). Deep sequencing of the ~300-bp deletion amplicon uncovered that accurate deletion-insertion constitutes 78.2±3.17% of total deletion events (Fig. 4g). We reasoned that, in this mouse model, hepatocytes with corrected FAH protein will outgrow cells with unintended editing, imposing a positive selection for desired editing. The average indel rates caused by each pegRNA at the *Fah* locus were 9.6% (cut site_F) and 0.14% (cut site_R) (Supplementary Fig. 6b). Although one mouse had a much higher average indel rate (27.7%) at cut site_F (Mouse 1 in Supplementary Fig. 6b), it did not negatively affect FAH protein expression (Mouse 1 in Fig. 4e). Overall, our data demonstrate the potential of using PEDAR in vivo to repair pathogenic mutations caused by large insertions.

Discussion

Here, we expanded the application scope of prime editing by developing a PE-Cas9-based deletion and repair method – PEDAR – that can correct mutations caused by larger genomic rearrangements. Our PEDAR system is similar to a recently-developed paired prime editing method, called *PRIME-Del*, that can introduce 20- to 700-bp target deletions and up to 30-bp insertions[36]. Compared to *PRIME-Del*, PEDAR seems to be more error-prone, introducing higher fractions of direct deletion and imperfect deletion-insertion (Supplementary Fig. 3g); however, both editors exhibited comparable absolute accurate rates in total genomic DNA (Supplementary Fig. 3h). Importantly, we show that PEDAR is able to introduce >10-kb target deletions and up to 60-bp insertions in cells, both of which are larger than what primer editors can generate[17, 36]. Moreover, PEDAR can program target deletion-insertion editing in quiescent hepatocytes in mouse liver, where HDR is not favorable[37].

Despite the relative editing efficiency and accuracy of PE-Cas9 being higher than PE and Cas9, the absolute editing efficiency of PEDAR is limited, possibly due to the cleavage activity, PBS length, and RT template length of the paired pegRNAs. Designing and comparing multiple paired pegRNA sequences could improve PEDAR efficiency. PEDAR

efficiency is also rendered by imperfect deletion-insertion edits due to partial insertion of pegRNA scaffold sequence[17] (Supplementary Fig. 2a). Optimizing the prime editing system to eliminate these unintended editing might improve the editing purity of PEDAR. Finally, PEDAR efficiency might be restricted by competition of distinct repair pathways at the DSBs. Given that PEDAR might employ a similar mechanism with MMEJ or SSA during repairing the DSB (Supplementary Fig. 2f), NHEJ pathway and MMEJ or SSA pathway might compete for repairing the DSB introduced by Cas9. Previous reports demonstrate that inhibition of NHEJ could enhance homology-mediated precise editing rate[38, 39]; and thus, this approach might improve PEDAR editing rate.

Finally, we propose that PEDAR could also be used to correct genome duplications (Supplementary Fig. 7a), which constitute ~10% of all human pathogenic mutations according to the ClinVar database[1]. One such genome duplication of high clinical significance is the trinucleotide CAG repeat expansion in the HTT gene – the root cause of Huntington disease[40]. Future studies should investigate whether PEDAR could accurately remove this expansion to reduce CAG repeat length. Thus, our findings have potential implications for the gene therapy field. The significance of PEDAR also extends to basic biology, where it could be used for protein function studies (Supplementary Fig. 7b). Previous studies introduce in-frame deletions by a "tiling CRISPR" method to explore the functional domain of specific genomic-coding or long non-coding regions[41, 42].

Methods

Cell Culture and Transfection

Human embryonic kidney (HEK293T) cells (ATCC) and HEK293T-TLR cells[1, 2] were maintained in Dulbecco's Modified Eagle's Medium (Corning) supplemented with 10% fetal bovine serum (Gibco) and 1% Penicillin/Streptomycin (Gibco). Cells were seeded at 70% confluence in 12-well cell culture plate one day before transfection. 1.5 μg PE-Cas9, and 1 μg paired pegRNAs (0.5 μg each) was transfected with Lipofectamine 3000 reagent (Invitrogen).

PegRNA design and clone

Sequences for pegRNAs are listed in Supplementary Table 1. Plasmids expressing pegRNAs were constructed by Gibson assembly using BsaI-digested acceptor plasmid (Addgene #132777) as the vector.

Mouse experiments

All animal study protocols were approved by the UMass Medical School IACUC. Fah[Exon5] mice[3] were kept on 10mg/L NTBC water. 30μg PE-Cas9 or Cas9 plasmid and 15μg paired pegRNA expressing plasmids were injected into 9-week-old mice. One week later, NTBC supplemented water was replaced with normal water, and mouse weight was measured every two days. As per our guidelines, when the mouse lost 20% of its body weight relative to the first day of measurement (day when NTBC water was removed), the mouse was supplemented with NTBC water until the original body weight was achieved. After 40 days, mice were euthanized according to guidelines.

Immunohistochemistry

Portion of livers were fixed with 4% formalin, embedded in paraffin, sectioned at 5 μm and stained with hematoxylin and eosin (H&E) for pathology. Liver sections were de-waxed, rehydrated, and stained using standard immunohistochemistry protocols[4]. The following antibody was used: anti-FAH (Abcam, 1:400). The images were captured using Leica DMi8 microscopy.

Genomic DNA extraction, amplification, and digestion

To extract genomic DNA, HEK293T cells (3 days post transfection) were washed with PBS, pelleted, and lysed with 50μl Quick extraction buffer (Epicenter) and incubated in a thermocycler (65°C 15 min, and 98°C 5 min). PureLink Genomic DNA Mini Kit (Thermo Fisher) was used to extract genomic DNA from two different liver lobes (~10 mg each) per mouse. Target sequences were amplified using Phusion Flash PCR Master Mix (Thermo Fisher) with the primers listed in Supplementary Table 2. PCR products were analyzed by electrophoresis in a 1% agarose gel, and target amplicons were extracted using DNA extraction kit (Qiagen). 10 ng of purified PCR products were incubated with I-SceI endonuclease (NEB) according to manufacturer's instruction. One-hour post incubation, the product was visualized and analyzed by electrophoresis in 4–20% TBE gel (Thermo).

Tracking of Indels by Decomposition (TIDE) analysis to calculate indel rates at two cut sites

The sequences around the two cut sites of the target locus were amplified using Phusion Flash PCR Master Mix (Thermo Fisher) with the primers as listed in Supplementary Table 2. Sanger sequencing was performed to sequence the purified PCR products, and the trace sequences were analyzed using TIDE software (https://tide.nki.nl/). The alignment window of left boundary was set at 10-bp.

Quantification of total genomic DNA to determine absolute editing rate of PEDAR

Real-time quantitative PCR (qPCR) was used to calculate the absolute editing rate in total genomic DNA at the *HEK3* locus. Quantitative PCR was performed with SsoFast EvaGreen Supermix (Bio-rad). Primers within the deletion region (P1 and P2), spanning the deletion region (P3 and P4), or across the deletion-insertion junction (P5 and P6) were designed (Supplementary Fig. 8a). Two 250-bp DNA fragments (referred to as WT and Edited) of the same sequence with unedited or accurately edited target site were designed and serially diluted, serving as standard templates (Supplementary Fig. 8b). Using indicated primers and templates to perform quantitative PCR, three standard curves were generated, reflecting the correlation between qPCR cycle number and the concentration of DNA without 991-bp deletion (Supplementary Fig. 8c), with 991-bp deletion (Supplementary Fig. 8d), or with accurate 991-bp deletion/18-bp insertion (Supplementary Fig. 8e). Finally, three rounds of quantitative PCR were performed using the edited genomic DNA as template and corresponding primer pairs (P1+P2, P3+P4, or P5+P6). The standard curves were applied to calculate the absolute copy number of genomic DNA with deletion, without deletion, or with accurate deletion-insertion.

The absolute rates of each type of editing introduced by PEDAR were calculated as follows: (1) Accurate deletion-insertion editing rate = copy number of DNA with accurate deletion-insertion / copy number of DNA with and without deletion. (2) Other deletion-insertion rate = (copy number of DNA with deletion - copy number of DNA with accurate deletion-insertion) / copy number of DNA with and without deletion. (3) Absolute rate of small indels at two cut sites = copy number of DNA without deletion × indel rate at distinct cut site calculated by TIDE / copy number of DNA with and without deletion

Flow Cytometry analysis

To assess mCherry recovery rate, post-editing HEK293T-TLR cells were trypsinized and analyzed using the MACSQuant VYB Flow Cytometer. Untreated HEK293T-TLR cells were used as a negative control for gating. To select cells with high transfection efficiency, 0.25 μg GFP plasmid was co-transfected with PE-Cas9 and paired pegRNAs into TLR cells. Three days post transfection, cells were trypsinized and analyzed using the MACSQuant VYB Flow Cytometer. Cells transfected with GFP plasmid alone were used as a negative control for gating. Cells with high expression level of GFP (~20% of total population) were selected for analyzing mCherry signal. All data were analyzed by FlowJo10.0 software.

High throughput DNA sequencing of genomic DNA samples

Genomic sites of interest were amplified from genomic DNA using specific primers containing illumina forward and reverse adaptors (listed in Supplementary Table 2). To quantify the percentage of desired deletion-insertion by PE-Cas9 or Cas9, we amplified the fragment containing deletions (~200 bp in length) from total genomic DNA to exclude length-dependent bias during PCR amplification. 20 μL PCR1 reactions were performed with 0.5 μM each of forward and reverse primer, 1 μL of genomic DNA extract or 300ng purified genomic DNA, and 10 μL of Phusion Flash PCR Master Mix (Thermo Fisher). PCR reactions were carried out as follows: 98°C for 10s, then 20 cycles of [98°C for 1 s, 55°C for 5 s, and 72°C for 10 s], followed by a final 72°C extension for 3 min. After the first round of PCR, unique Illumina barcoding reverse primer was added to each sample in a secondary PCR reaction (PCR 2). Specifically, 20 μL of a PCR reaction contained 0.5 μM of unique reverse Illumina barcoding primer pair and 0.5 μM common forward Illumina barcoding primer, 1 μL of unpurified PCR 1 reaction mixture, and 10 μL of Phusion Flash PCR Master Mix. The barcoding PCR2 reactions were carried out as follows: 98 °C for 10s, then 20 cycles of [98°C for 1 s, 60°C for 5 s, and 72°C for 10 s], followed by a final 72 °C extension for 3 min. PCR 2 products were purified by 1% agarose gel using a QIAquick Gel Extraction Kit (Qiagen), eluting with 15 μL of Elution Buffer. DNA concentration was measured by Bioanalyzer and sequenced on an Illumina MiSeq instrument (150bp, paired-end) according to the manufacturer's protocols. Paired-end reads were merged with FLASh[5] (v1.2.11) with maximum overlap length equal to 150 bp. Alignment of amplicon sequence to the reference sequence was performed using CRISPResso2[6] (v2.0.32). To quantify accurate deletion-insertion edits, CRISPResso2 was run in HDR mode using the sequence with desired deletion-insertion editing as the reference sequence. The editing window is set to 15-bp. Editing yield was calculated as: [# of HDR aligned reads] ÷ [total aligned reads].

ClinVar data analysis

The ClinVar variant summary was obtained from NCBI ClinVar database (accessed Dec 31,2020). Variants with pathogenic significance were filtered by allele ID to remove duplicates. All pathogenic variants were categorized according to mutation type. The fractions of distinct mutation types were calculated using GraphPad Prism8.

Statistics and Reproducibility

In Figs 1c–d, 2b–c, 2f–g, 3d, 4f and Supplementary Figs 2d–e, 3a–b, 3e–f, 4a–b, 5b, 5d, three biological repeats were performed with similar results. In Figs 4c, 4e and Supplementary Fig 6a, four mice (Fig 4c: PE-Cas9, Figs 4e, and Supplementary Fig 6a) or two mice (Fig 4c: Cas9) were employed in the experiment with similar results.

Data Availability

A reporting summary for this article is available as a Supplementary Information file. The raw gel images underlying Figs. 1c–d, 2b–c, 2f–g, 3d,4f and Supplementary Figs. 2d–e, 3a–b, 3e–f, 4a–b are provided as a Source Data File and an additional supplementary data file, respectively. NCBI Clinvar database is accessible through the indicated link: https://www.ncbi.nlm.nih.gov/clinvar/. The raw DNA sequencing data are available at the NCBI Sequence Read Archive database under PRJNA746292 and PRJNA746489.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Landrum MJ et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42, D980–985 (2014). [PubMed: 24234437]

2. Cordaux R & Batzer MA The impact of retrotransposons on human genome evolution. Nat Rev Genet 10, 691–703 (2009). [PubMed: 19763152]

3. Chen JM, Stenson PD, Cooper DN & Ferec C A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet 117, 411–427 (2005). [PubMed: 15983781]

4. Hancks DC & Kazazian HH Roles for retrotransposon insertions in human disease. Mobile DNA 7, 9 (2016). [PubMed: 27158268]

5. Wang L, Norris ET & Jordan IK Human Retrotransposon Insertion Polymorphisms Are Associated with Health and Disease via Gene Regulatory Phenotypes. Front Microbiol 8, 1418 (2017). [PubMed: 28824558]

6. Hancks DC & Kazazian HH Jr. Active human retrotransposons: variation and disease. Curr Opin Genet Dev 22, 191–203 (2012). [PubMed: 22406018]

7. Qian Y et al. Identification of pathogenic retrotransposon insertions in cancer predisposition genes. Cancer Genet 216–217, 159–169 (2017).

8. Ran FA et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389 (2013). [PubMed: 23992846]

9. Cong L et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823 (2013). [PubMed: 23287718]

10. Kato T et al. Creation of mutant mice with megabase-sized deletions containing custom-designed breakpoints by means of the CRISPR/Cas9 system. Sci Rep 7, 59 (2017). [PubMed: 28246396]

11. Hara S et al. Microinjection-based generation of mutant mice with a double mutation and a 0.5 Mb deletion in their genome by the CRISPR/Cas9 system. J Reprod Dev 62, 531–536 (2016). [PubMed: 27396308]

12. Wang L et al. Large genomic fragment deletion and functional gene cassette knock-in via Cas9 protein mediated genome editing in one-cell rodent embryos. Sci Rep 5, 17517 (2015). [PubMed: 26620761]

13. Yeh CD, Richardson CD & Corn JE Advances in genome editing through control of DNA repair pathways. Nat Cell Biol 21, 1468–1478 (2019). [PubMed: 31792376]

14. Zheng Q et al. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. Biotechniques 57, 115–124 (2014). [PubMed: 25209046]

15. Cox DB, Platt RJ & Zhang F Therapeutic genome editing: prospects and challenges. Nature medicine 21, 121–131 (2015).

16. Liu M et al. Methodologies for Improving HDR Efficiency. Front Genet 9, 691 (2018). [PubMed: 30687381]

17. Anzalone AV et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149–157 (2019). [PubMed: 31634902]

18. Matsoukas IG Prime Editing: Genome Editing for Rare Genetic Diseases Without Double-Strand Breaks or Donor DNA. Front Genet 11, 528 (2020). [PubMed: 32582281]

19. Liu P et al. Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. Nat Commun 12, 2121 (2021). [PubMed: 33837189]

20. Jang H et al. Prime editing enables precise genome editing in mouse liver and retina. bioRxiv, 2021.2001.2008.425835 (2021).

21. Schene IF et al. Prime editing for functional repair in patient-derived disease models. Nat Commun 11, 5352 (2020). [PubMed: 33097693]

22. Jiang YY et al. Prime editing efficiently generates W542L and S621I double mutations in two ALS genes in maize. Genome biology 21, 257 (2020). [PubMed: 33023639]

23. Song X, Huang H, Xiong Z, Ai L & Yang S CRISPR-Cas9(D10A) Nickase-Assisted Genome Editing in Lactobacillus casei. Appl Environ Microbiol 83 (2017).

24. Cho SW et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res 24, 132–141 (2014). [PubMed: 24253446]

25. Sfeir A & Symington LS Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? Trends Biochem Sci 40, 701–714 (2015). [PubMed: 26439531]

26. Bhargava R, Onyango DO & Stark JM Regulation of Single-Strand Annealing and its Role in Genome Maintenance. Trends Genet 32, 566–575 (2016). [PubMed: 27450436]

27. Kim HK et al. Predicting the efficiency of prime editing guide RNAs in human cells. Nat Biotechnol 39, 198–206 (2021). [PubMed: 32958957]

28. Mir A et al. Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing. Nat Commun 9, 2641 (2018). [PubMed: 29980686]

29. Certo MT et al. Tracking genome engineering outcome at individual DNA breakpoints. Nat Methods 8, 671–676 (2011). [PubMed: 21743461]

30. Zhan H, Li A, Cai Z, Huang W & Liu Y Improving transgene expression and CRISPR-Cas9 efficiency with molecular engineering-based molecules. Clin Transl Med 10, e194 (2020). [PubMed: 33135339]

31. Chen R et al. Enrichment of transiently transfected mesangial cells by cell sorting after cotransfection with GFP. Am J Physiol 276, F777–785 (1999). [PubMed: 10330060]

32. Homann S et al. A novel rapid and reproducible flow cytometric method for optimization of transfection efficiency in cells. PloS one 12, e0182941 (2017). [PubMed: 28863132]

33. Pham CT, MacIvor DM, Hug BA, Heusel JW & Ley TJ Long-range disruption of gene expression by a selectable marker cassette. Proceedings of the National Academy of Sciences of the United States of America 93, 13090–13095 (1996). [PubMed: 8917549]

34. Grompe M et al. Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. Genes & development 7, 2298–2307 (1993). [PubMed: 8253378]

35. Paulk NK et al. Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo. Hepatology 51, 1200–1208 (2010). [PubMed: 20162619]

36. Choi J et al. Precise genomic deletions using paired prime editing. bioRxiv, 2020.2012.2030.424891 (2021).

37. VanLith CJ et al. Ex Vivo Hepatocyte Reprograming Promotes Homology-Directed DNA Repair to Correct Metabolic Disease in Mice After Transplantation. Hepatol Commun 3, 558–573 (2019). [PubMed: 30976745]

38. Dutta A et al. Microhomology-mediated end joining is activated in irradiated human cells due to phosphorylation-dependent formation of the XRCC1 repair complex. Nucleic Acids Research 45, 2585–2599 (2016).

39. Aida T et al. Gene cassette knock-in in mammalian cells and zygotes by enhanced MMEJ. BMC Genomics 17, 979 (2016). [PubMed: 27894274]

40. Warby SC et al. CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. Am J Hum Genet 84, 351–366 (2009). [PubMed: 19249009]

41. Wang Y et al. Identification of a Xist silencing domain by Tiling CRISPR. Sci Rep 9, 2408 (2019). [PubMed: 30787302]

42. He W et al. De novo identification of essential protein domains from CRISPR-Cas9 tiling-sgRNA knockout screens. Nat Commun 10, 4541 (2019). [PubMed: 31586052]

References

1. Mir A et al. Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing. Nat Commun 9, 2641 (2018). [PubMed: 29980686]

2. Certo MT et al. Tracking genome engineering outcome at individual DNA breakpoints. Nat Methods 8, 671–676 (2011). [PubMed: 21743461]

3. Grompe M et al. Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. Genes & development 7, 2298–2307 (1993). [PubMed: 8253378]

4. Xue W et al. Response and resistance to NF-kappaB inhibitors in mouse models of lung adenocarcinoma. Cancer discovery 1, 236–247 (2011). [PubMed: 21874163]

5. Magoc T & Salzberg SL FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963 (2011). [PubMed: 21903629]

6. Clement K et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol 37, 224–226 (2019). [PubMed: 30809026]
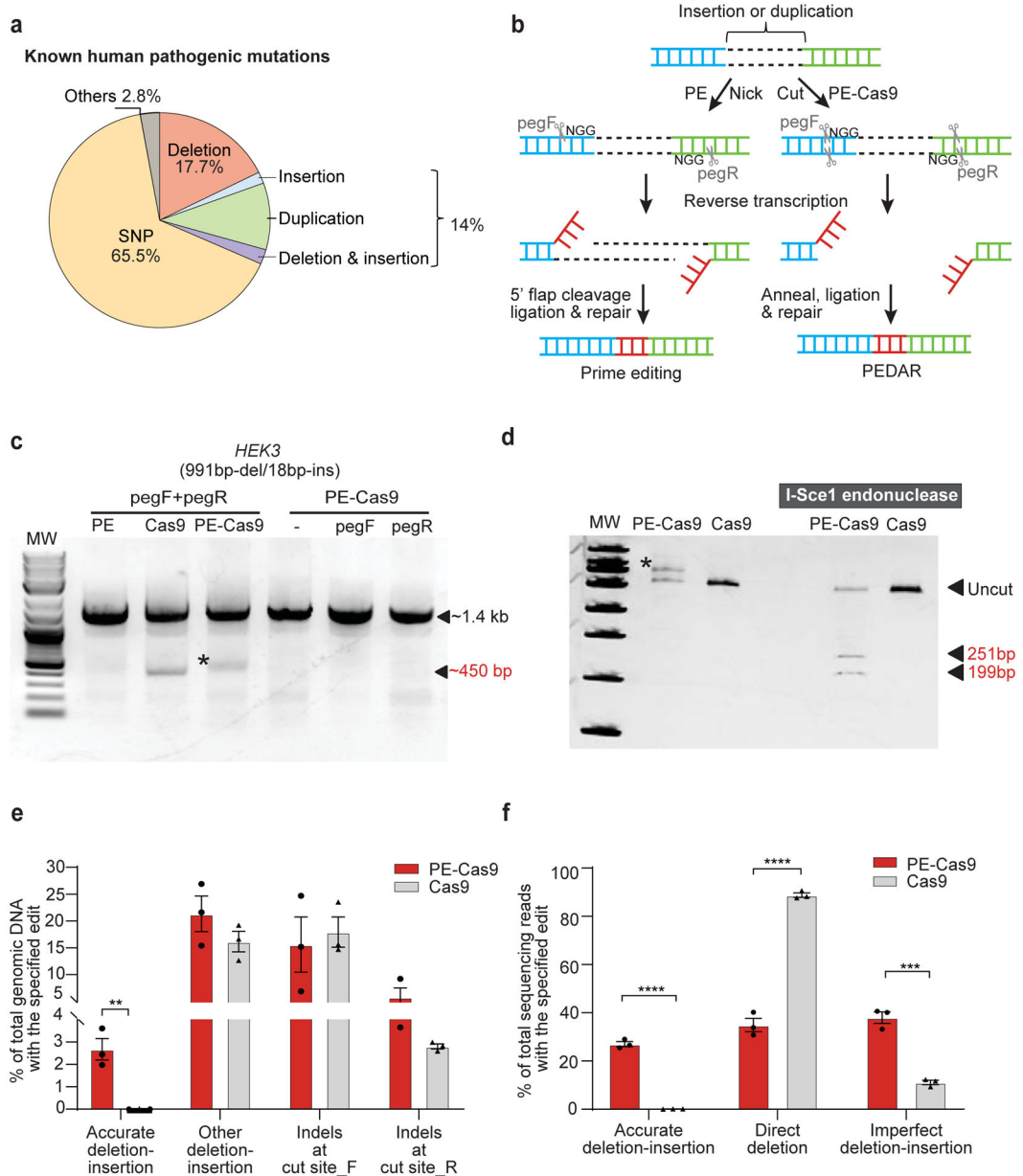
Fig 1. PEDAR mediates large target deletion and simultaneous insertion at an endogenous genomic locus.

a, Classification of the 60,008 known human pathogenic genetic variants reported in the ClinVar database[1]. **b,** Overview of using prime editing (left) and PEDAR (right) to generate accurate deletion-insertion. **c,** Deleting a 991-bp DNA fragment and simultaneous insertion of I-Sce1 recognition sequence (18bp) at the *HEK3* locus (Chr9:107422166–107423588). Target genomic region was amplified using primers that span the cut sites. The paired pegRNAs targeting complementary DNA strand are denoted as pegF and pegR. HEK293T cells were transfected with PE, Cas9, or PE-Cas9 with or without single or paired pegRNAs. The ~450-bp band is the expected deletion amplicon (denoted with *), and the ~1.4-kb band is the amplicon without deletion. **d,** Deletion amplicons from Cas9- or PE-Cas9-treated groups shown in (c) were incubated with or without I-SceI endonuclease and analyzed in

4–20% TBE gel. Digested products are marked by arrows with expected sizes. Original amplicon is marked as "uncut". The band with insertion of i-Sce1 recognition sequence is denoted with *. **e,** PE-Cas9 or Cas9-mediated absolute rates of all the editing events in total genomic DNA at *HEK3* site. Data represent mean ± SEM (n = 3 biologically independent samples). P= 0.0053 (**), two-tailed t-test. **f,** Deep sequencing of deletion amplicons shown in (c). Bar chart shows distribution of all deletion events, including accurate deletion-insertion, direct deletion (deletion without any insertions), or imperfect deletion-insertion. Editing rate = the reads with indicated editing / total deletion events. Data represent mean ± SEM (n = 3 biologically independent samples). P=0.0000140 (****), 0.002 (***), two-tailed t-test.
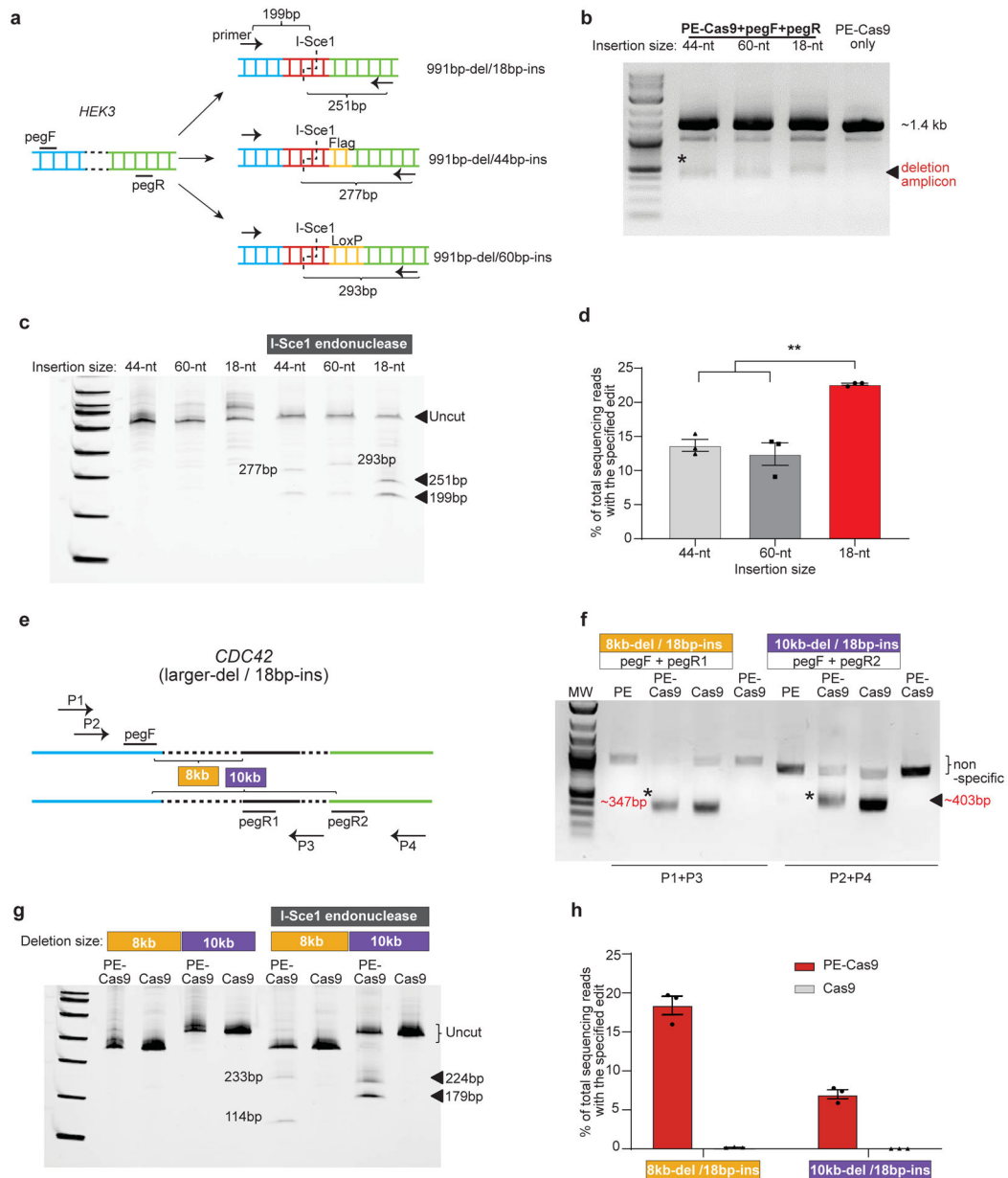
Fig 2. Flexibility of PEDAR in programming larger deletion and insertion in HEK293T cells.
a, Insert DNA fragments of variable lengths (18-bp, 44-bp, and 60-bp) to the target site
at *HEK3* locus. pegRNAs and primers for amplifying the target site are as shown. The
expected sizes of digestion products after I-Sce1 treatment are shown. **b,** Amplification of
target genomic region using primers spanning the cut sites at *HEK3* locus. The deletion
amplicons are denoted with *. Cells transfected with PE-Cas9 alone serves as negative
control. **c,** Deletion amplicons from groups shown in (b) were incubated with or without
I-SceI endonuclease and analyzed in 4–20% TBE gel. Digested products are marked by
arrows with expected sizes. The original amplicon is marked as "uncut". **d,** Deep sequencing
of deletion amplicons shown in (b). Bar chart shows accurate deletion-insertion rate. Data
represent mean ± SEM (n = 3 biologically independent samples). P=0.0024 (**), two-tailed

t-test. **e,** Test the efficiency of PEDAR in mediating larger deletions. Paired pegRNAs spaced ~8-kb (pegF+pegR1) or 10-kb (pegF+pegR2) apart were designed as indicated to target the *CDC42* locus. Primers used to amplify the target genomic regions are as marked (P1+P3 and P2+P4). **f,** Target genomic region was amplified using the primers indicated in (e). Cells transfected with PE-Cas9 alone serve as negative control. The deletion amplicons are marked with expected sizes (denoted with *). **g,** Deletion amplicons from Cas9- or PE-Cas9-treated groups shown in (f) were incubated with or without I-SceI endonuclease and analyzed in 4–20% TBE gel. Digested products are marked with expected sizes. The original amplicon is marked as "uncut". **h,** Deep sequencing of deletion amplicons shown in (f). Bar chart shows rate of accurate deletion-insertion events. Data represent mean ± SEM (n = 3 biologically independent samples).
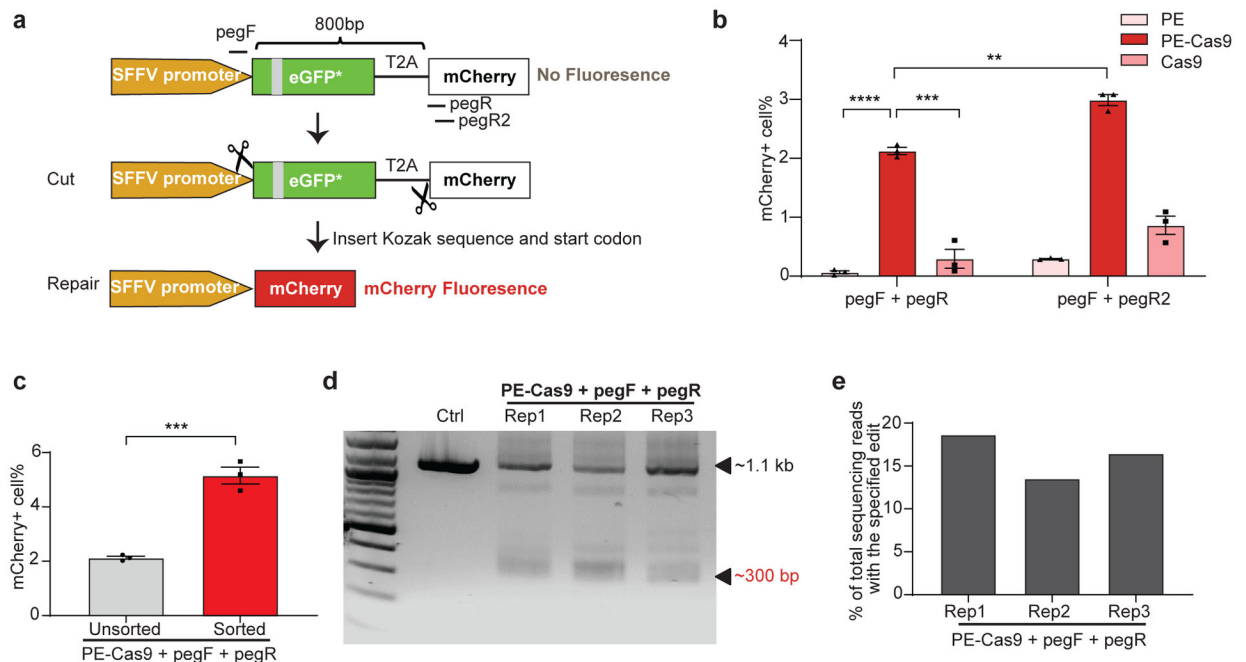
Fig 3. PEDAR generates in-frame deletion to restore mCherry expression in TLR reporter cells.
a, Diagram of the TLR reporter system. GFP sequence is disrupted by an insertion (grey).
Deleting the disrupted GFP sequence and inserting Kozak sequence and start codon will
restore mCherry protein expression. **b,** TLR reporter cells transfected with indicated paired
pegRNAs along with PE, PE-Cas9, or Cas9 were analyzed by flow cytometry, and the
percentage of mCherry positive cells are shown among different groups. Data represent
mean ± SEM (n = 3 biologically independent samples). P=0.00000636 (****), =0.0004
(***), =0.0015 (**), two-tailed t-test. **c,** mCherry positive cell rate before and after sorting of
cells with high transfection level. A plasmid expressing GFP was co-transfected with paired
pegRNAs and PE-Cas9 into TLR cells. Three days later, cells with high GFP expression
were selected for analyzing mCherry signal by flow cytometry. Data represent mean ±
SEM (n = 3 biologically independent samples). P=0.0007 (***), two-tailed t-test. **d,** TLR
reporter cells edited by PEDAR were selected by flow cytometry (for mCherry signal) and
subjected to PCR amplification using primers spanning the two cut sites. The amplicon
with the desired deletion is ~300 bp compared to a ~1.1-kb PCR products in control group.
Rep: replicate; Ctrl: untreated TLR reporter cells. **e,** Efficiency of accurate deletion-insertion
in three PEDAR-edited replicates (Rep 1–3) measured by deep sequencing of the deletion
amplicons shown in (d).
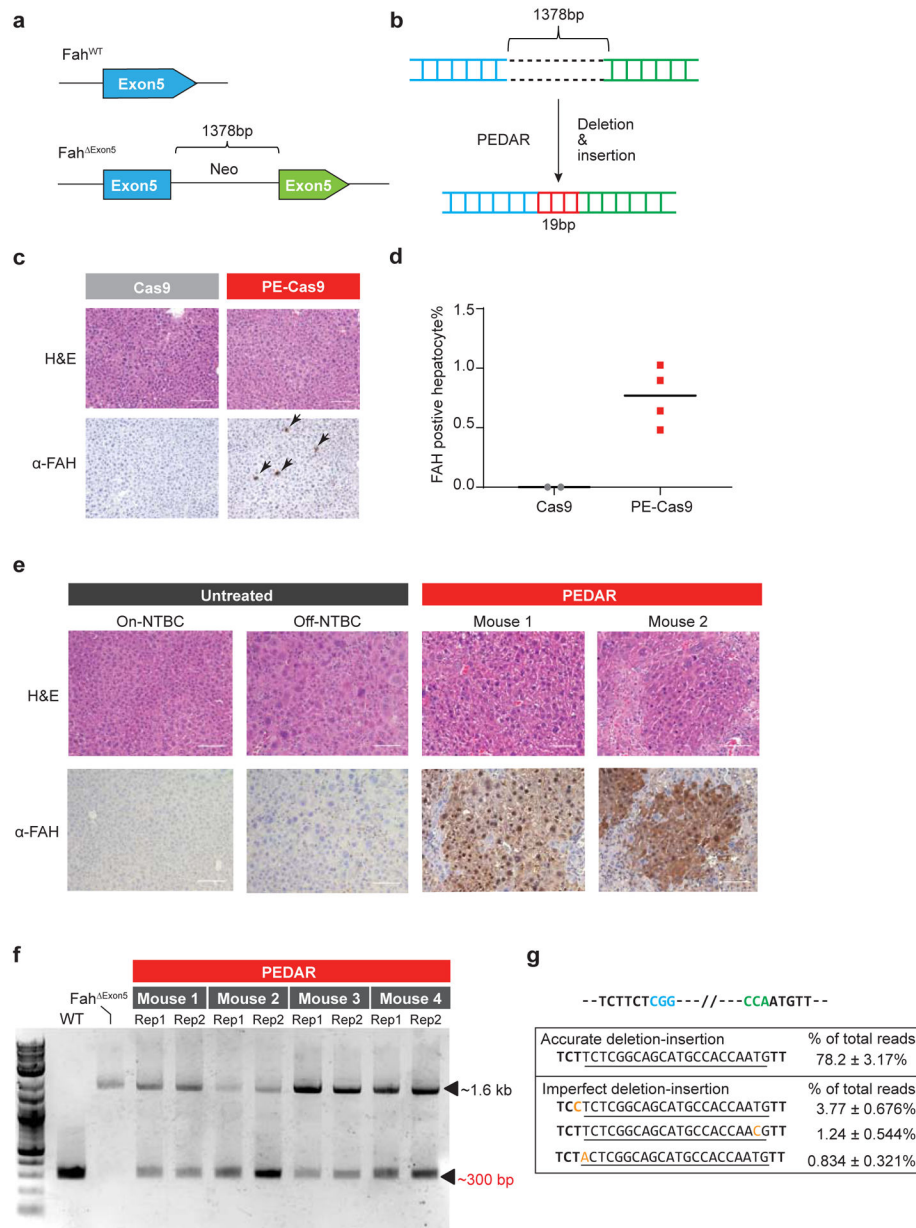
Fig 4. PEDAR corrects the pathogenic insertion in a Tyrosinemia I mouse model.
a, The Tyrosinemia I mouse model, referred to as Fah^Exon5, was derived by integrating a ~1.38-kb neo expression cassette at exon 5 of the *Fah* gene. **b,** Diagram showing the application of PEDAR to delete the ~1.38-kb insertion and concurrently repair the target region by inserting a 19-bp DNA fragment (marked in red). **c,** Immunohistochemistry staining and Hematoxylin and Eosin staining (H&E) of mouse liver sections seven days after injection of dual pegRNAs with Cas9 or PE-Cas9. FAH positive hepatocytes are pointed by arrows. Scale bar=100μm. **d,** Quantification of FAH expressing hepatocytes shown in (c). n=2 (Cas9-treated group), n=4 (PE-Cas9-treated group). **e,** Immunohistochemistry and H&E staining of mouse liver sections 40 days after injection of PE-Cas9 with dual pegRNAs. Mice (n=4) were kept off NTBC. Mouse 1 and 2 denote two representative mice from the

treatment group. The liver sections from untreated Fah ^{Exon5} kept on or off NTBC serve as negative controls. Scale bar=100μm. **f,** Amplification of exon 5 of *Fah* gene from mouse livers 40 days post injection of PE-Cas9 and paired pegRNAs. The corrected amplicon size is around ~300 bp, compared to a ~1.6-kb amplicon without deletion. Four mice in treated group and two liver lobes (denoted as Rep 1 and 2) per mouse were analyzed. WT: wild type C57BL/6J mouse. Fah ^{Exon5}: untreated Fah ^{Exon5} mouse. **g,** Accurate correction rate and the top-three imperfect editing events identified by deep sequencing. Two PAM sequences are in blue and green. The 22-bp intended insertion (19-bp deletion fragment plus a 3-bp unintentionally deleted sequence) is underlined. Mutated nucleotides in imperfect editing sequences are highlighted in yellow. Data represent mean ± SD (n = 8, two liver lobes/mouse; four mice in total).