nature biotechnology

Article

https://doi.org/10.1038/s41587-024-02394-x

Rapid generation of long, chemically modified pegRNAs for prime editing

Received: 31 January 2024

Accepted: 21 August 2024

Published online: 30 September 2024

Check for updates

Xinlin Lei[1,2,5], Anhui Huang[1,2,5], Didi Chen[1,2,5], Xuebin Wang[1,2,5], Ruijin Ji[1,2], Jinlin Wang[1,2], Yizhou Zhang[1,2], Yuming Zhang[1,2], Shuhan Lu[1,2], Kun Zhang[1,2], Qiubing Chen[1,2], Ying Zhang[1,2,3] & Hao Yin[1,2,4]

The editing efficiencies of prime editing (PE) using ribonucleoprotein (RNP) and RNA delivery are not optimal due to the challenges in solid-phase synthesis of long PE guide RNA (pegRNA) (>125 nt). Here, we develop an efficient, rapid and cost-effective method for generating chemically modified pegRNA (125–145 nt) and engineered pegRNA (epegRNA) (170–190 nt). We use an optimized splint ligation approach and achieve approximately 90% production efficiency for these RNAs, referred to as L-pegRNA and L-epegRNA. L-epegRNA demonstrates enhanced editing efficiencies across various cell lines and human primary cells with improvements of up to more than tenfold when using RNP delivery and several hundredfold with RNA delivery of PE, compared to epegRNA produced by in vitro transcription. L-epegRNA-mediated RNP delivery also outperforms plasmid-encoded PE in most comparisons. Our study provides a solution to obtaining high-quality pegRNA and epegRNA with desired chemical modifications, paving the way for the use of PE in therapeutics and various other fields.

CRISPR-based genome editing systems introduce targeted alternation of the genome[1]. CRISPR–Cas nucleases cause double-stranded breaks to facilitate small insertions or deletions (indels) or targeted insertion via nonhomologous end joining or homology-directed repair, respectively[2,3]. Applying catalytically impaired Cas nucleases together with various deaminases effectively convert C-to-T or A-to-G, referred to as base editing[4]. Fusion of Cas9 nickase with a reverse transcriptase created prime editors to generate small modifications around the nicked site without donor DNA[5]. The prime editing (PE) guide RNA (pegRNA) contains a reverse-transcription template (RTT) and primer binding sequence (PBS) at the 3′ end of single-guide RNA (sgRNA), resulting in at least approximately 125–145 nucleotides (nt) in length[5]. Unlike sgRNA, which is protected by the Cas9 protein, the 3′ extension of pegRNA, including RTT and PBS, is susceptible to degradation by ribonucleases within the cell. The structured RNA motifs have been incorporated at the 3′ end of pegRNA to enhance its stability and prevent the degradation of PBS and RTT sequences[6–8]. The evopreQ$_1$ motif has been included into pegRNA to generate the engineered pegRNA (epegRNA), and it has been shown to improve the editing efficiencies of PE by 3–4-fold in several cell types tested[6].

Chemical modifications of sgRNA can improve its stability against degradation and greatly enhance its performance in cells and in vivo[9–12]. Chemically modified sgRNA has been extensively used in research and therapeutic applications. The first Food and Drug Administration approved CRISPR-based therapy, CASGEVY, delivered chemically modified sgRNA and Cas9 protein as ribonucleoprotein (RNP) into hematopoietic stem and progenitor cells (HSPC), as an ex vivo therapy for β-thalassemia and sickle cell disease[13,14]. In vivo CRISPR-based therapy has entered late-stage clinical development to treat transthyretin (ATTR) amyloidosis with cardiomyopathy using

[1]Departments of Urology and Laboratory Medicine, Frontier Science Center for Immunology and Metabolism, Medical Research Institute, Zhongnan Hospital of Wuhan University, Wuhan University, Wuhan, China. [2]State Key Laboratory of Virology, TaiKang Center for Life and Medical Sciences, TaiKang Medical School, Wuhan University, Wuhan, China. [3]Department of Rheumatology and Immunology, Zhongnan Hospital of Wuhan University, Wuhan University, Wuhan, China. [4]RNA Institute, Wuhan University, Wuhan, China. [5]These authors contributed equally: Xinlin Lei, Anhui Huang, Didi Chen, Xuebin Wang. e-mail: haoyin@whu.edu.cn

RNA delivery[15]. Chemically modified sgRNA and Cas9 messenger RNA (mRNA), which are encapsulated into lipid nanoparticles, have been shown to effectively edit the *ATTR* gene in vivo in patients[16]. Chemical modifications have been applied to pegRNA for RNP and mRNA delivery of PE[6,17–22]. However, solid-phase synthesis needs to be stretched to produce chemically modified pegRNA and epegRNA, thus increasing costs and the quality of chemically synthesized RNA with such a length is not satisfactory[6,19]. While RNP and RNA delivery of Cas9 and/or sgRNA usually surpass the plasmids delivery in terms of editing efficiency, the editing efficiency of PE via RNP and RNA delivery is lower than expected[6,9–11,17,18,20–22]. We reasoned that the relatively low editing efficiency via RNP and RNA delivery of PE is in part due to the low quality of chemically synthesized pegRNA or epegRNA.

In vitro RNA synthesis relies on enzymatic-based method such as in vitro transcription (IVT) or chemical synthesis via solid-phase phosphoramidite chemistry[23,24]. IVT uses RNA polymerase such as T7 and several others to generate an RNA sequence up to 30 kb (refs. 24–26). Modified nucleotides including pseudouridine (Ψ), N1-methylpseudouridine (m1Ψ) and 5-methylcytidine are introduced randomly or completely into mRNA to attenuate immunogenicity[27–30]. However, standard IVT cannot generate site-specific modifications, and it cannot tolerate modified nucleotides such as 2′OH modifications of ribose for enhancing resistance to ribonucleases[9–12,31]. Moreover, IVT by T7 enzyme usually produces sequences containing heterogeneous 5′ and/or 3′ end products[26,32]. A hybrid solid–liquid phase transcription approach was combined with automated robotic platform to generate RNA with position-selective modification[33]. Recently, a biocatalytic method was developed to generate oligonucleotides by combining polymerases and endonucleases in one pot[34]. However, these two methods are suitable for generating high-quality modified short oligonucleotides in large quantities rather than producing RNA longer than 100 nt (refs. 33,34). Solid-phase chemical synthesis enables production of desired RNA sequences with position-selective modifications to incorporate modified residues at the base and sugar phosphate backbone to enhance stability[9,23,24,35–37]. The linear chemical synthesis of oligonucleotides uses nucleoside phosphoramidite as building blocks, and the synthesis relies on repeated rounds of several chemical reactions in order, extending by 1 nt each round[23,24,35–37]. Although the efficiency of individual elongation cycle is high, overall yield of synthesis drops sharply with longer sequences. Therefore, current solid-phase synthesis generally limits RNA fragment length to approximately 100 nt, and incorporation of modified nucleotides is a challenge for long RNA synthesis[24,35–39]. Despite that stretching the chemical synthesis for generation of RNA sequences between 100 and 200 nt is possible, the yield of production steeply declines and the frequent failure of synthesis occurs[24,35–39]. Truncated sequences arise from incomplete coupling reactions, and they could block solid support[40]. All these factors contribute to product impurities, some of which are difficult to remove by purification[40]. Hence, the synthesis of pegRNAs, ranging from 125 to 145 nt for introducing point mutations and small indels through solid-phase synthesis, is economically demanding and it would be very challenging to chemically synthesize epegRNAs, which are 170–190 nt (refs. 6,24,35–37,39).

RNA ligation enables creation of relatively long RNA molecules using small pieces[41,42]. RNA ligases facilitate this process by joining RNA fragments with 5′-phosphate and 3′-hydroxyl ends, and some ligases can also link RNA fragments with 2′,3′-cyclic phosphate and 5′-hydroxyl ends[42–44]. Splint ligation uses a splint DNA to hybridize the ends of RNA fragments and direct ligase activity in joining the 3′ and 5′ ends[35,38]. The efficiency of splint ligation varies but often remains low, and the low yield of products from RNA ligation restricted its broad use[31,42,43,45]. Here, we optimized various parameters of splint ligation and applied it to effectively generate chemically modified pegRNAs and epegRNAs with high yield and purity, referred to as L-pegRNA and L-epegRNA. The potency of L-epegRNA has been examined in different cell lines and human primary cells. Compared to epegRNA produced by IVT, L-epegRNA demonstrated up to several hundredfold improvements of PE's editing efficiencies by RNA delivery. L-epegRNA also greatly improves the editing efficiency of PE in RNP format, enabling RNP delivery of PE to surpass plasmid-encoded PE in editing efficiency for most comparisons examined. We also generated chemically modified epegRNAs for relatively large insertions, while the 210 and 234 nt RNAs are beyond the current limit of solid-phase synthesis. These epegRNAs enable approximately 60% efficiencies of 17 bp insertion and about 20 or 25% efficiencies of 40 bp insertion, via either RNP or RNA delivery, respectively. L-epegRNA dramatically boosts the efficiency of PE via RNP and RNA delivery, and the reduced cost and efficient ligation process will facilitate L-epegRNA for broad usage including therapeutic applications.

Results

Optimization of the RNA ligation process

We established a method to produce sgRNA and pegRNA with defined chemical modifications using RNA ligation and optimized the ligation process. We selected Bacteriophage T4 RNA ligase 2 (T4 Rnl2) for RNA ligation due to its preference for joining gaps between adjacent 5′ phosphates and 3′ hydroxyl groups in double-stranded structures[46]. To examine the feasibility of our method, we first ligated the sgRNA, which has a simpler structure compared to pegRNA. The 102 nt sgRNA was divided into two parts: the 20 nt spacer, serving as the acceptor RNA, and the 82 nt scaffold region as the donor RNA (Extended Data Fig. 1a). The 20 nt acceptor RNA was chemically synthesized with a hydroxyl group at the 3′ end, while the 82 nt scaffold RNA was in vitro transcribed to allow for cost-effective large-scale preparation. To ensure a single phosphate modification at the 5′ end of the donor RNA, guanosine monophosphate was introduced during IVT. The presence of a splint DNA facilitated RNA ligation (Extended Data Fig. 1a). We assessed the ligation products using PCR with reverse transcription (RT–PCR) and Sanger sequencing. The results indicated that this enzymatic ligation successfully produced a full-length sgRNA (Extended Data Fig. 1b).

To improve the efficiency of ligation, we optimized various parameters, including the ligation temperature, the length of splint DNA, the ratio of splint DNA to RNA and the dose of ligase. Our results indicated that ligation at 37 °C was more effective than at the previously described 25 °C (Extended Data Fig. 1c)[47]. A longer splint DNA length improved the ligation efficiency (Extended Data Fig. 1d). The optimal molar ratio of RNA to splint DNA was 1:1, and a higher ratio of splint DNA adversely affected ligation (Extended Data Fig. 1e). A moderate dose of T4 Rnl2 was sufficient for ligation, as further increasing the enzyme quantity did not enhance the ligation efficiency (Extended Data Fig. 1f). We also synthesized 5′ end-modified acceptor RNAs (with three nucleosides harboring 2′-*O*-methyl modifications and three phosphorothioate linkages) for sgRNA ligation. Effective ligation was achieved, indicating that modifications away from the ligation site did not affect the enzymatic reaction (Extended Data Fig. 1g). Finally, we used ligated sgRNA for in vitro cleavage of substrate DNA. Both 5′ end-modified and unmodified ligated sgRNA effectively cleaved the substrate DNA (Extended Data Fig. 1h). In summary, with optimized conditions for RNA ligation, a chemically defined and functional sgRNA can be efficiently obtained.

Ligation of pegRNA

The pegRNA includes additional RTT, PBS sequences, and an optional protection sequence at the 3′ end of the sgRNA[6,7,48]. Based on the optimized ligation process described above, we synthesized two relatively short RNAs. We synthesized a 32 nt acceptor RNA with three nucleosides harboring 2′-*O*-methyl modifications and phosphorothioate linkages at the 5′ end, along with a donor RNA containing a 5′ phosphate generated through IVT (Fig. 1a). The ligation results demonstrated the successful assembly of a full-length pegRNA through three cycles

https://doi.org/10.1038/s41587-024-02394-x

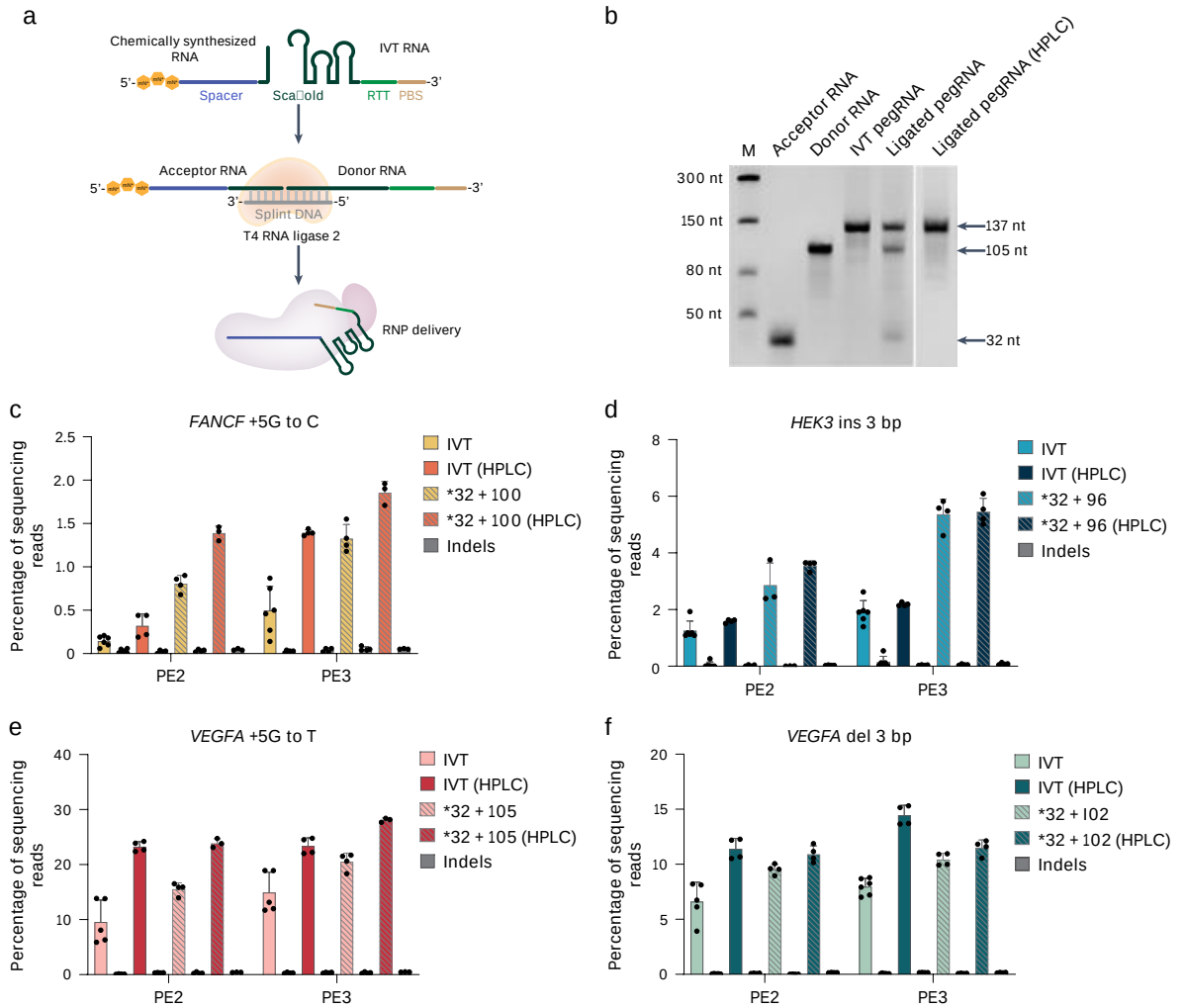


Fig. 1 | Ligated pegRNA mediates efficient PE. a, Overview of the initial design scheme for ligation of pegRNA. The acceptor RNA is a 32 nt 5′ end-modified RNA, and it includes the spacer sequence and part of the sgRNA scaffold sequence. The donor RNA was produced by IVT, and it includes the rest of scaffold sequence and RTT-PBS. The acceptor RNA has a hydroxyl group at the 3′ end, and the donor RNA has a phosphate group at the 5′ end. T4 Rnl2 ligates facilitates the joining of RNA molecules in the presence of splint DNA. **b**, Ligation of pegRNA (for +5G to T mutation at *VEGFA* locus). From left to right, lane 1, marker; lane 2, chemically synthesized 32 nt acceptor RNA; lane 3, 105 nt donor RNA by IVT; lane 4, full-length pegRNA by IVT as a control; lane 5, ligated 137 nt pegRNA and lane 6, HPLC-purified 137 nt ligated pegRNA. The samples were run in 6% denaturing urea-PAGE. **c**–**f**, The efficiencies of RNP-mediated PE in HEK293T cells were determined by deep sequencing. Four different pegRNAs were used as +5G to C mutation at *FANCF* (**c**), 3 bp insertion at *HEK3* (**d**), +5G to T mutation at *VEGFA* (**e**) and 3 bp deletion at *VEGFA* loci (**f**). 'IVT' indicates full-length pegRNA generated by IVT (PE2, $n = 6, 6, 4, 4$; PE3, $n = 6, 6, 5, 6$ from left to right side); 'IVT (HPLC)' indicates full-length pegRNA generated by IVT with HPLC purification (PE2, $n = 4$; PE3, $n = 4$); 32 + 100/96/105/102 indicates ligated pegRNA with 5′ end-modified (PE2, $n = 4, 3, 4, 4$ from left to right side; PE3, $n = 4$); 32 + 100/96/105/102 (HPLC) indicated ligated pegRNA with HPLC purification (PE2, $n = 3, 4, 3, 4$; PE3, $n = 3, 4, 3, 4$ from left to right side). For each electroporation, 140 pmol of PE protein, 186 pmol of pegRNA and 62 pmol of nicking sgRNA were used. Modification indicates three nucleosides harboring 2′-*O*-methyl modification and with three phosphorothioate linkages. Data and error bars represent the mean and standard deviation of three or more independent biological replicates. The *n* values for PE2 and PE3 in **c**–**f** are indicated alongside each sample type.

of annealing and ligation over a total of 2 hours at 37 °C (Fig. 1b and Methods). To evaluate the accuracy of the ligation site, we performed RT–PCR and deep sequencing on ligated pegRNA, using IVT pegRNA as the control. The results of deep sequencing indicated that the ligated pegRNA exhibited precise ligation sites without any additional changes (Extended Data Fig. 2a).

Subsequently, we used the same method to ligate different pegRNAs to introduce editing events at various loci in human embryonic kidney 293 (HEK293T) cells. Our results revealed that ligated pegRNA with chemical modifications at the 5′ end exhibited 1.4- to 5.6-fold higher editing efficiency than pegRNA produced by IVT for PE2, and 1.3- to 2.8-fold higher for PE3 via RNP delivery (Fig. 1c–f). Our initial donor RNA was produced by IVT to minimize synthesis costs (Fig. 1b–f). However, due to the essential nature of guanosine triphosphate (GTP) as a substrate in IVT, a considerable portion of donor RNAs lacked the 5′ single phosphate, rendering them incapable of being ligated by T4 Rnl2. As a result, unligated RNA persisted in the ligation products (Fig. 1b). High-performance liquid chromatography (HPLC) analysis indicated that approximately 30% of the unpurified RNA consisted of unligated RNA (Extended Data Fig. 2b,c). Subsequently, we performed HPLC purification on ligated pegRNA and IVT pegRNA, resulting in substantially improved product purity (Extended Data Fig. 2d,e). HPLC purification generally increased the editing efficiencies of IVT pegRNA- and ligated pegRNA-mediated RNP delivery compared to their unpurified controls, with the exception of the *HEK3* site (Fig. 1c–f). HPLC-purified ligated pegRNAs showed higher editing efficiencies than the IVT pegRNA counterparts for two out of four sites, suggesting that 5′ end modification of pegRNA alone is not sufficient for greatly improving the stability of pegRNA, as the 3′ end of pegRNA is unprotected by the Cas effector. Nevertheless, these data demonstrated

that pegRNA generated by the ligation approach can facilitate RNP delivery-mediated precision editing.

Stabilizing the 3′ end of pegRNA to enhance efficiency

The 3′ extension of pegRNA, when left unshielded, is exposed in the cellular environment, making it more susceptible to degradation by cellular nucleases. A truncated 3′ end pegRNA would lose its PBS and RTT but retain the ability to bind Cas9, severely impeding the efficiency of PE[6,7]. The evopreQ$_1$ motif has been used to protect the 3′ extension of pegRNA, generating 'epegRNA', which leads to increased stability and enhanced PE efficiency[6]. We hypothesized that stabilizing the 3′ end of pegRNA and epegRNA via ligation could improve the efficiency of PE. Therefore, we designed three different ligation strategies to protect the 3′ end (Fig. 2a–c). The first strategy used 32 nt synthetic 5′-modified acceptor RNA and a chemically synthesized donor RNA with 3′ end modification (referred to as *pegRNA*, indicating chemical modifications at both ends) (Fig. 2a). Full-length pegRNAs were chemically synthesized as controls (referred to as synthetic pegRNA, or briefly as S-pegRNA, with 3 nt chemical modifications at both ends). S-pegRNA and *pegRNA* are identical in both sequence and modification.

The second strategy involved ligating synthetic 5′-modified acceptor RNA with evopreQ$_1$-containing donor RNA generated by IVT (referred to as *epegRNA, with the asterisk * indicating chemical modifications at the 5′ end) (Fig. 2b). To simultaneously add chemical modifications at both ends and evopreQ1 to the ligated pegRNA, we split the acceptor and donor RNA into comparable lengths (referred to as *epegRNA*) (Fig. 2c). The acceptor RNA is 91 nt long and contains most of the sgRNA sequence, while the donor RNA ranges from 80 to 95 nt (Fig. 2c). For both *pegRNA* and *epegRNA*, three nucleotides at the 5′ end of the acceptor RNA and the 3′ end of the donor RNA are modified (Fig. 2a,c). We observed that when the donor RNA was chemically synthesized, the efficiency of the ligation reaction was higher compared to the reaction where the donor RNA was generated by IVT. This resulted in approximately 90% efficiency for *pegRNA* and *epegRNA* ligation, partly due to the predominant 5′ phosphate in chemically synthesized donor RNA (Extended Data Fig. 3a).

Ligated pegRNA contains by-products from incomplete ligation. To ensure a rigorous comparison with S-pegRNA, we performed HPLC purification on the three designed pegRNAs (Extended Data Fig. 3b). HPLC-purified S-pegRNA and *pegRNA* exhibited similar editing efficiencies in PE2 and PE3 via RNP delivery, indicating that the ligation approach does not compromise the activity of the resulting RNA (Fig. 2d–f). Our results revealed that *epegRNA (no chemical modification at 3′ end, with only evopreQ1 motif) exhibited 1.5- to 1.8-fold higher editing efficiency than S-pegRNA for PE2, and 1.4- to 1.6-fold higher for PE3 via RNP delivery, suggesting that the evopreQ1 motif alone may provide more effective protection than chemical modification alone for the 3′ end of pegRNA (Fig. 2d–f).

Among all the pegRNAs examined, *epegRNA*, which combined the evopreQ1 motif and chemical modification (Fig. 2c), demonstrated the highest editing efficiencies for both PE2 and PE3 via RNP delivery. While S-pegRNA exhibited average editing efficiencies of 23.5% for PE2 and 31.0% for PE3 across three different pegRNAs, *epegRNA* showed significantly higher average editing efficiencies of 59.0% for PE2 and 66.3% for PE3 (Fig. 2d–f). Notably, *epegRNA* achieved up to 77.3% for PE2 and 83.4% for PE3 with RNP delivery of PE, indicating a synergistic effect of chemical modification and the RNA motif in enhancing RNA stability (Fig. 2c–f).

We named *epegRNA* L-epegRNA, which is around 170–190 nt in length for PE-mediated point mutations, small deletions and small insertions. It is worth noting that generating epegRNA within this length range is extremely difficult using solid-phase synthesis[6,19,24,35,37,39]. Moreover, due to the low quality of RNA products, solid-phase synthesized epegRNA did not substantially improve editing efficiencies compared to chemically synthesized pegRNA[6,19]. When we split epegRNA into two RNA sequences of similar size (70–100 nt), each sequence can be synthesized at high quality using solid-phase synthesis.

HPLC purification is not widely adopted in biomedical research laboratories, raising concerns about the generalizability of the ligation method. However, L-epegRNA without HPLC purification still exhibited much higher editing efficiencies via RNP delivery than chemically modified pegRNA, achieving 43.4 to 69.2% for PE2 and 42.2 to 72.3% for PE3 (Extended Data Fig. 3c–e). This indicates the potency of L-epegRNA even without HPLC purification. To demonstrate the generalizability of the ligation method, all subsequent experiments, unless otherwise specified, used L-epegRNAs that were not HPLC purified. We compared L-epegRNA with epegRNA generated by IVT. For the six pegRNA sequences examined, L-epegRNA demonstrated average editing efficiencies of 39.9% for PE2 and 53.1% for PE3 in HEK293T cells via RNP delivery, which is 1.9 to 11.4 times higher than the epegRNA generated by IVT (Fig. 2g).

Splint DNA is required to ligate pegRNA, and DNase I is used to digest the splint DNA after ligation. However, incomplete digestion of the splint DNA could result in immune-stimulatory effects. To address this, we delivered L-pegRNA without HPLC purification via electroporation into THP1 cells, using S-pegRNA as the control. The signature genes in the cGAS/STING and TLR9 pathways were not activated after treatment with L-pegRNA or S-pegRNA (Extended Data Fig. 4a). Additionally, PAGE analysis of L-epegRNA after RNase treatment showed no residual nucleic acids, indicating complete digestion of the splint DNA (Extended Data Fig. 4b).

Impact of RTT and PBS length on efficiency via RNP delivery

The editing efficiencies of PE can be regulated by the lengths of RTT and PBS[5,49,50]. Furthermore, it has been suggested that the optimal PBS length varies depending on the delivery method, with differences noted between RNP, mRNA and plasmid deliveries[21]. In our investigation of optimal PBS lengths with varying RTT lengths via RNP delivery, we found that the optimal PBS lengths varied with different RTT lengths (Extended Data Fig. 5a–h). Notably, PBS lengths of 10 or 13 nucleotides generally produced the best results.

Optimization of ligated pegRNA-mediated RNP delivery

Due to the limitations of pegRNA synthesis, previous studies on RNP delivery of PE have been constrained from systematic optimization[20,21]. Here, we optimized various conditions that may affect RNP delivery of PE using ligated pegRNA. To ensure nonsaturating editing, experiments were conducted using *epegRNA. We optimized the total dosages, the ratio of PE protein to pegRNA and the dosage of nicking sgRNA in HEK293T and K562 cell lines using the RNP delivery system. We found that using 70 to 140 pmol of RNP, a 1:1 to 1:2 ratio of PE protein to pegRNA and 30 to 60 pmol of nick sgRNA was sufficient for efficient editing (Extended Data Fig. 6a–i).

It has been suggested that the RNase H domain could be dispensable for PE[51,52], and the PEmax version often exhibits superior performance compared to classic PE[17]. Therefore, we expressed and purified four versions of PE: PE, PE$^{\Delta RH}$, PEmax and PEmax$^{\Delta RH}$ (Extended Data Fig. 7a,b). We found that the removal of the RNase H domain increased the yield of protein in the *Escherichia coli* expression system (Extended Data Fig. 7c). Subsequently, we examined the editing efficiency of these PE proteins. The removal of the RNase H domain showed a trend of improved editing efficiency, with PEmax$^{\Delta RH}$ displaying slightly better efficiency than PE$^{\Delta RH}$ for PE2 editing (Extended Data Fig. 7d,e). Consequently, we used PEmax$^{\Delta RH}$ for most of the studies below, except in cases where a long RTT sequence formed secondary structures[53].

Suppression of the mismatch repair pathway via overexpression of a dominant-negative mismatch repair protein (MLH1dn) has been shown to improve the efficiencies of PE in several cell lines, including K562 cells[17]. Therefore, we expressed and purified the MLH1dn protein to examine its effect on RNP delivery of PE. The addition of purified

https://doi.org/10.1038/s41587-024-02394-x

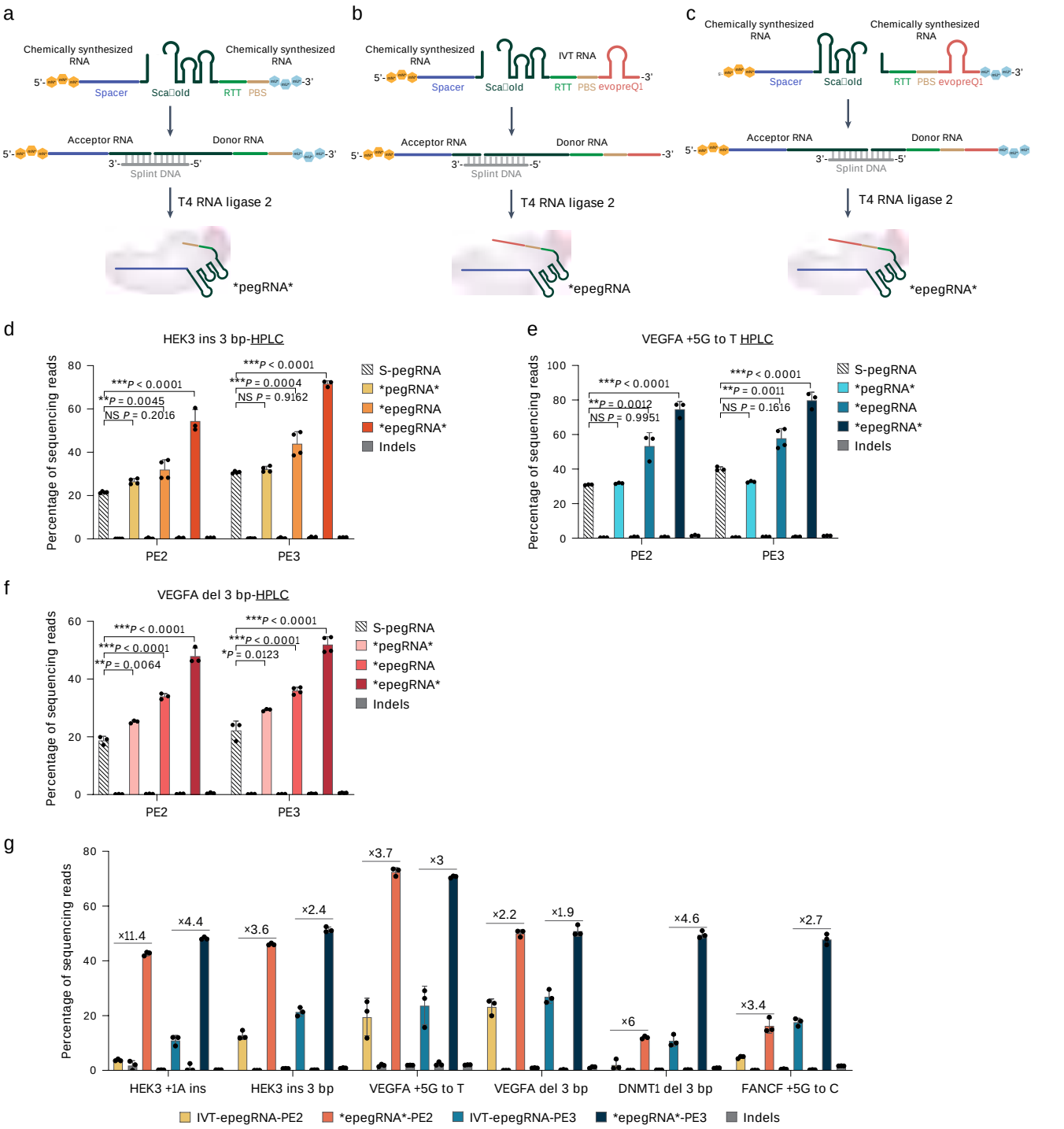


Fig. 2 | Comparison of different ligation strategies for pegRNA. a–**c**, Overview of design schemes to generate pegRNA with both 5′ and 3′ ends modifications (**a**), epegRNA with only 5′ end modification (**b**) and epegRNA with both 5′ and 3′ ends modifications (**c**). **d**–**f**, The efficiencies of RNP-mediated PE in HEK293T cells were determined by deep sequencing for 3 bp insertion at *HEK3* locus (**d**), *VEGFA* locus +5G to T mutation (**e**) and *VEGFA* locus 3 bp deletion (**f**). S-pegRNA: full-length pegRNAs were synthesized by solid-phase synthesis, with 3 nt chemical modifications at both ends (PE2, $n = 4, 3, 3$; PE3, $n = 4, 3, 3$ from left to right side). *pegRNA*, ligated pegRNA by a 32 nt synthetic acceptor RNA with 5′ end modifications and a synthetic donor RNA with 3′ end modifications but without evopreQ$_1$ (PE2, $n = 4, 3, 3$; PE3, $n = 4, 3, 3$ from left to right side). *epegRNA, ligated epegRNA by a 32 nt synthetic acceptor RNA with 5′ modifications and IVT-generated donor RNA containing evopreQ$_1$ (PE2, $n = 4, 3, 3$ from left to right side; PE3, $n = 4$). *epegRNA*, ligated epegRNA by a 91 nt synthetic acceptor RNA with 5′ end modifications and synthetic donor RNA with 3′ end modifications at evopreQ$_1$ sequence (PE2, $n = 3$; PE3, $n = 3, 3, 4$ from left to right side). All these RNAs were HPLC purified. For each electroporation, 140 pmol of PE protein, 186 pmol of pegRNA and 62 pmol of nicking sgRNA were used. **g**, Comparison of RNP-mediated PE efficiencies of L-epegRNA (*epegRNA*) and IVT-generated epegRNA. For each electroporation, 70 pmol of PE protein, 140 pmol of pegRNA and 60 pmol of nicking sgRNA were used ($n = 3$). Modification indicates three nucleosides harboring 2′-*O*-methyl modification and with three phosphorothioate linkages. Data and error bars represent the mean and standard deviation of three or more independent biological replicates. The *n* values for PE2 and PE3 in **d**–**g** are indicated alongside each sample type. Data analysis used one-way ANOVA with Tukey's multiple comparisons test; NS, not significant; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

https://doi.org/10.1038/s41587-024-02394-x

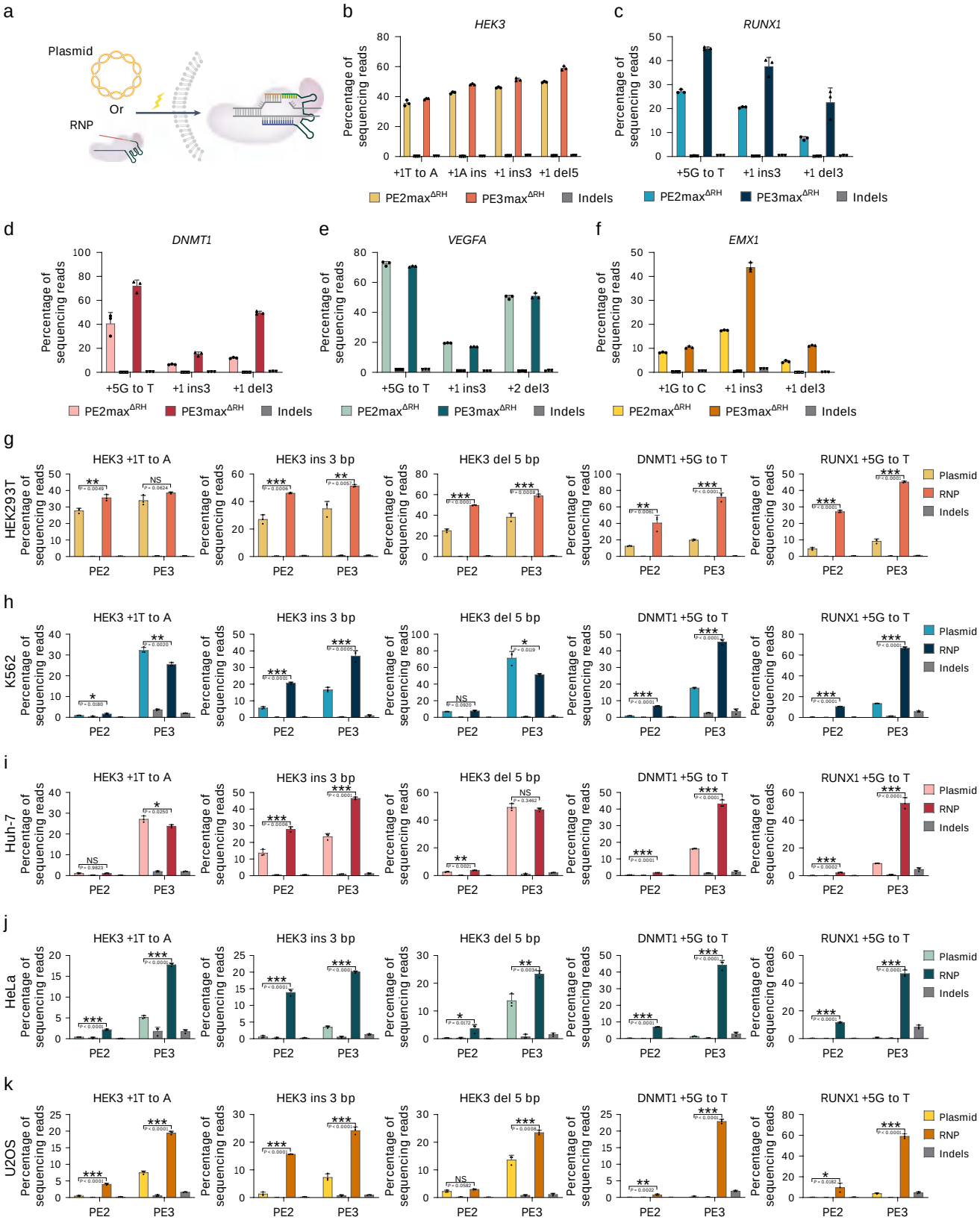


Fig. 3 | Comparison of PE efficiencies via plasmid and L-epegRNA-mediated RNP delivery. a, Schematic representation of L-epegRNA-mediated RNP and plasmid delivery for PE. Plasmid and RNP were introduced into cells via electroporation. **b**–**f**, The efficiencies of PE for five endogenous loci of *HEK3* (**b**), *RUNX1* (**c**), *DNMT1* (**d**), *VEGFA* (**e**), *EMX1* (**f**) loci in HEK293T cells using L-epegRNA-mediated RNP delivery. **g**–**k**, Comparisons of PE efficiencies using plasmid and L-epegRNA-mediated RNP delivery in HEK293T (**g**), K562 (**h**), Huh-7 (**i**), HeLa (**j**) and U2OS cells (**k**). For each sample of RNP delivery, 70 pmol of PEmax$^{\Delta RH}$ protein, 140 pmol of L-epegRNA and 60 pmol of nicking sgRNA were used. Data and error bars represent the mean and standard deviation from at least three independent biological replicates. Data were analyzed by two-tailed unpaired Student's *t*-test; NS indicates no significance; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

https://doi.org/10.1038/s41587-024-02394-x

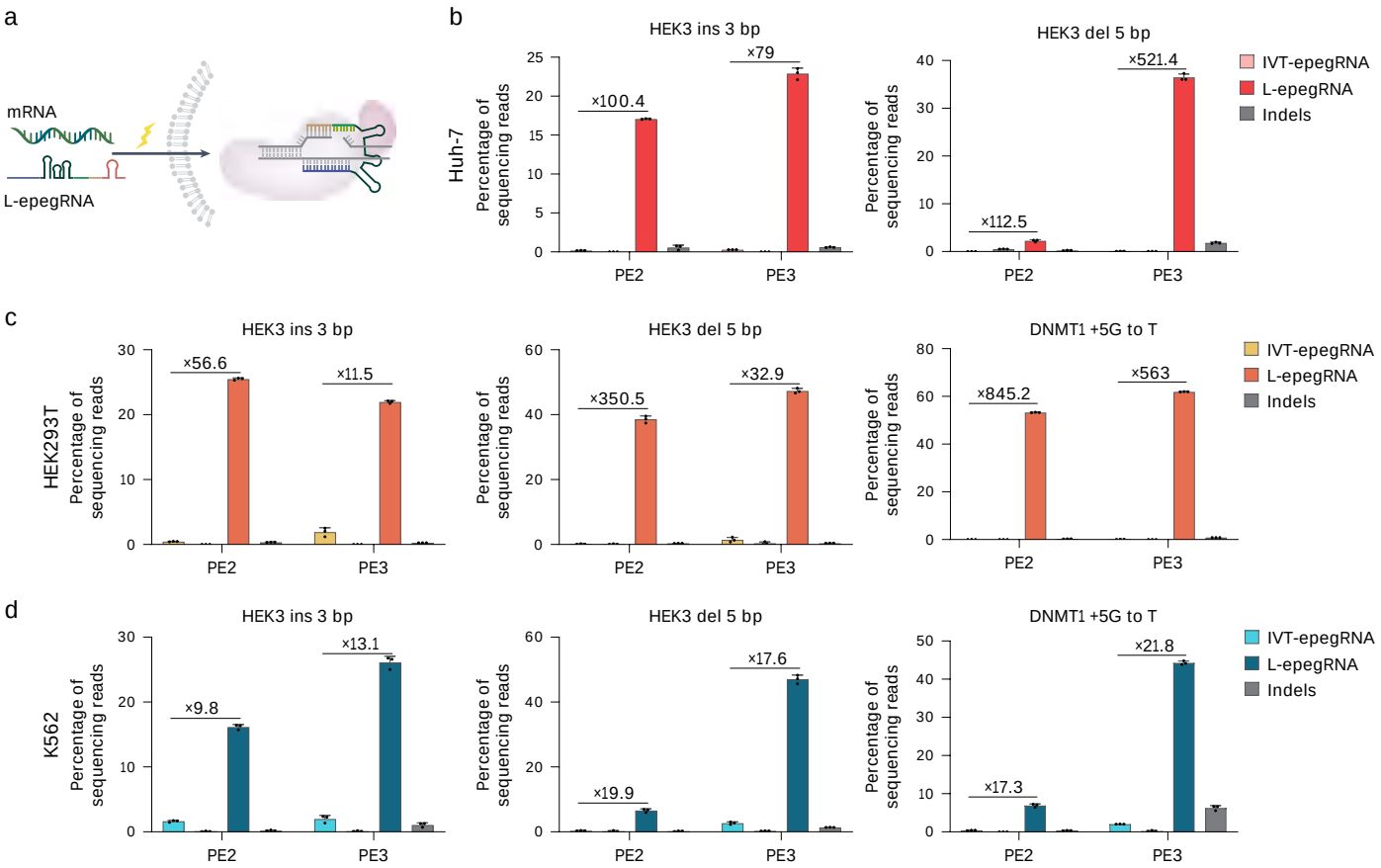


Fig. 4 | PE efficiencies by L-epegRNA-mediated RNA delivery. a, Illustration of RNA delivery for PE. L-pegRNA and mRNA encoding PEmax were cotransfected into cells through electroporation. **b**–**d**, Comparison of PE efficiencies by L-epegRNA or IVT-generated epegRNA when they were cotransfected with mRNA encoding PEmax in Huh-7 (**b**), HEK293T (**c**) and K562 cells (**d**). For each sample, 2 μg of PEmax mRNA, 180 pmol of epegRNA and 60 pmol of nicking sgRNA were used. Data and error bars represent the mean and standard deviation from three independent biological replicates.

MLH1dn protein at various doses (up to 140 pmol) for PE2 and PE3 did not increase the editing efficiencies (Extended Data Fig. 8a,b). As controls, when delivered as plasmids, PE4max (PE2max with overexpression of MLH1dn) showed superior editing compared to PE2max for both pegRNAs examined, and PE5max exhibited improved efficiency compared to PE3max in one of the two pegRNAs examined (Extended Data Fig. 8c,d).

We examined the editing efficiencies using L-epegRNA under the optimized RNP delivery conditions described above. Using PEmax$^{\Delta RH}$ protein, we examined three types of change (point mutation, small insertion and small deletion) at five endogenous loci. L-epegRNA efficiently edited all loci with PE2 and PE3, exceeding 70% efficiency for two loci (Fig. 3a–f).

L-epegRNA-mediated RNP delivery is superior to plasmid

For the same pegRNA sequences, editing efficiencies via RNP delivery of PE are usually lower than those achieved with plasmid delivery in previous studies[6,17–22]. We compared the performance of PE via RNP delivery or plasmid delivery for five different pegRNA sequences in five different cell lines, including HEK293T, K562, Huh-7, HeLa and U2OS cells. L-epegRNA-mediated RNP delivery exhibited significantly higher editing efficiencies for most comparisons (22 out of 25 for PE2, 20 out of 25 for PE3) (Fig. 3g–k). Notably, while plasmid-delivered PE exhibited extremely low or undetectable editing in several comparisons, L-epegRNA-mediated RNP delivery achieved excellent editing efficiencies. For example, at the *RUNX1* locus (+5G to T) in HeLa cells, plasmid delivery of PE2 and PE3 yielded average efficiencies of 0.14 and 0.53%, respectively. In contrast, L-epegRNA-mediated RNP delivery resulted in average efficiencies of 11.6 and 47.9% for PE2 and PE3, respectively (Fig. 3g). A similar phenomenon was observed at the *DNMT1* site (+5G to T) in U2OS cells, where plasmid delivery of PE3 yielded an editing efficiency of 0.17%. In comparison, L-epegRNA-mediated RNP delivery resulted in 22.9% editing (Fig. 3k). These data suggest that L-epegRNA-mediated PE can be effective in loci where plasmid delivery of PE fails to generate efficient editing.

L-epegRNA facilitates efficient editing via RNA delivery

In addition to RNP delivery, RNA delivery of PE can also be used for therapeutics, research and other applications. We explored the use of L-epegRNA in the RNA delivery of PE platforms in three cell lines (Fig. 4a). In comparison to epegRNA generated by IVT, L-epegRNA exhibited 56.6–845.2-fold, 9.8–19.9-fold and 100.4–112.5-fold improvements for PE2 in HEK293T, K562 and Huh-7 cells, respectively (Fig. 4b–d). Similarly, L-epegRNA exhibited 13.6–563.0-fold, 13.1–21.8-fold and 79.0–521.4-fold improvements compared to IVT-epegRNA for PE3 in these cells, respectively (Fig. 4b–d).

L-epegRNA achieves efficient PE in primary cells

Therapeutic applications of PE in primary cells often require the use of synthetic pegRNA for RNP or RNA delivery[17,18,20,21,54,55]. We used L-epegRNA for both RNP and RNA delivery of PE in human primary T cells and human CD34$^+$ HSPCs. L-epegRNA generated efficient editing in both cell types via RNP and RNA delivery. The editing efficiency generated by RNA delivery appeared to be higher than that of RNP delivery, ranging from 4.7 to 34.6% for RNP and 14.7 to 71.4% for RNA delivery, respectively (Fig. 5a–d).

https://doi.org/10.1038/s41587-024-02394-x

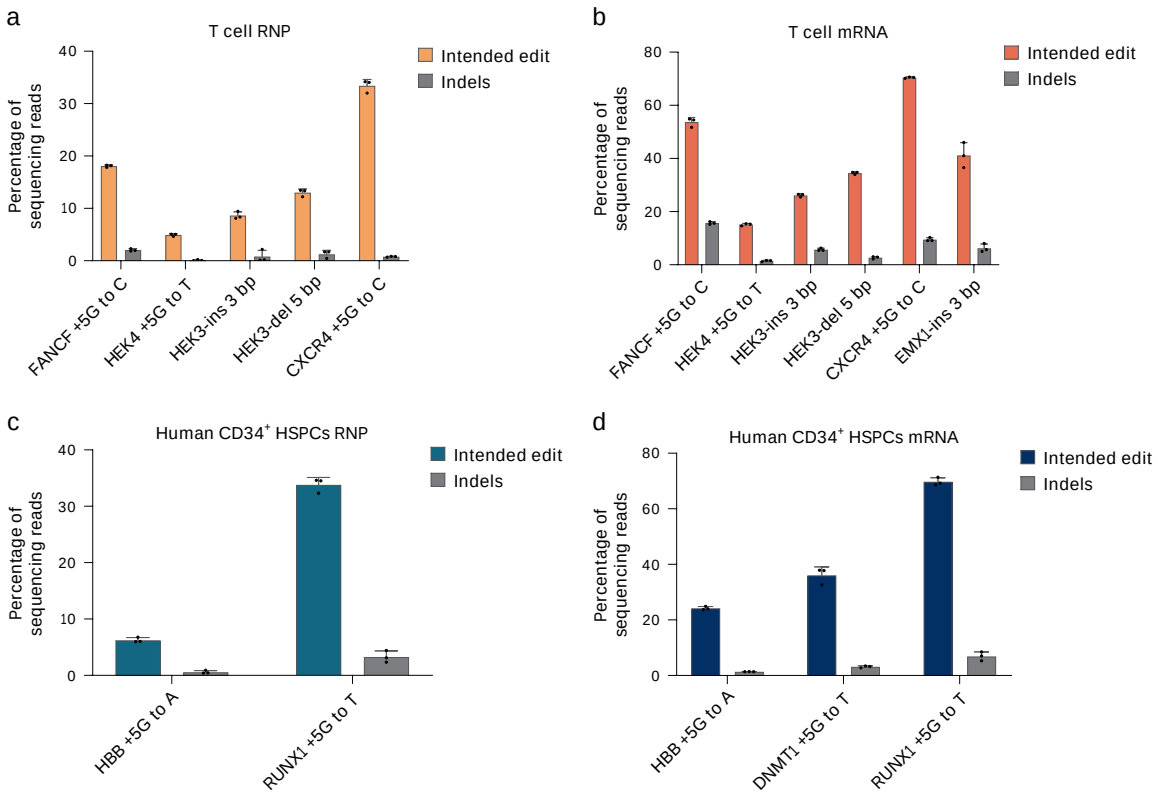


Fig. 5 | PE efficiencies using L-epegRNA in human primary T cells and hematopoietic stem cells. a–d, PE efficiencies of L-epegRNA-mediated RNP in human T cells (**a**) and CD34$^+$ HSPCs (**c**), and RNA delivery in human T cells (**b**) and CD34$^+$ HSPCs (**d**) were determined by deep sequencing. For RNP delivery, 70 pmol of PEmax$^{\Delta RH}$ protein, 140 pmol of L-epegRNA and 60 pmol of nicking sgRNA were used for each sample. For RNA delivery, 2 µg of mRNA encoding PEmax was cotransfected with 180 pmol of L-epegRNA and 60 pmol of nicking sgRNA. Data and error bars represent the mean and standard deviation from three independent biological replicates.

L-epegRNA achieves relatively large insertions

The upper limit of chemical synthesis for RNA is approximately 200 nt, and the feasibility of synthesis depends on the complexity of the sequence[24,35–37,39]. While epegRNA reached a size of 170 to 190 nt for point mutations or small insertions and/or deletions, the size of epegRNA for longer insertions exceeds 200 nt. We aimed to ligate an epegRNA for the insertion of 17 bp at the *HEK3* locus, resulting in an epegRNA length of 210 nt. We designed two ligation strategies: a two-segment ligation (105 + 105 nt) and a three-segment ligation (76 + 54 + 80 nt), respectively (Fig. 6a). Our data indicate that the ligation efficiency of the three-segment was higher than that of the two-segment approach, likely due to the reduced accuracy of synthesizing the 5′ monophosphate of donor DNA longer than 100 nt or the complexity of RNA structure in the two-segment connection (Fig. 6b). As the 210 nt epegRNA cannot be directly synthesized by solid-phase synthesis, we excluded the evopreQ1 motif and chemically synthesized a 166 nt S-pegRNA as a control (Extended Data Fig. 9a). It is worth noting that RNA of this length is very difficult to synthesize by solid-phase, and this RNA was obtained only after multiple failed attempts. HPLC-purified S-pegRNA demonstrated an efficiency of 5.1% for 17 bp insertion via RNP delivery and 10.3% via RNA delivery (Fig. 6c,d). IVT-generated epegRNA induced up to 8.7% of 17 bp insertion, and HPLC purification did not substantially improve the efficiency of this IVT-generated epegRNA for 17 bp insertion (Fig. 6c,d). In contrast, L-epegRNA generated by three-segment ligation without HPLC purification achieved efficiencies of 42.1 and 49.4% for 17 bp insertion via RNP and RNA delivery, respectively (Fig. 6c,d). HPLC purification further boosted the rate of 17 bp insertion by L-epegRNA to 55.8 and 59.7% for RNP and RNA delivery, respectively (Fig. 6c,d). HPLC-purified L-epegRNA exhibited more than tenfold higher editing efficiency for 17 bp insertion than S-pegRNA for PE3 via RNP delivery, and about sixfold higher via RNA delivery (Fig. 6c,d). This further demonstrates the synergistic effect of chemical modification and the RNA motif in improving the activities of pegRNA.

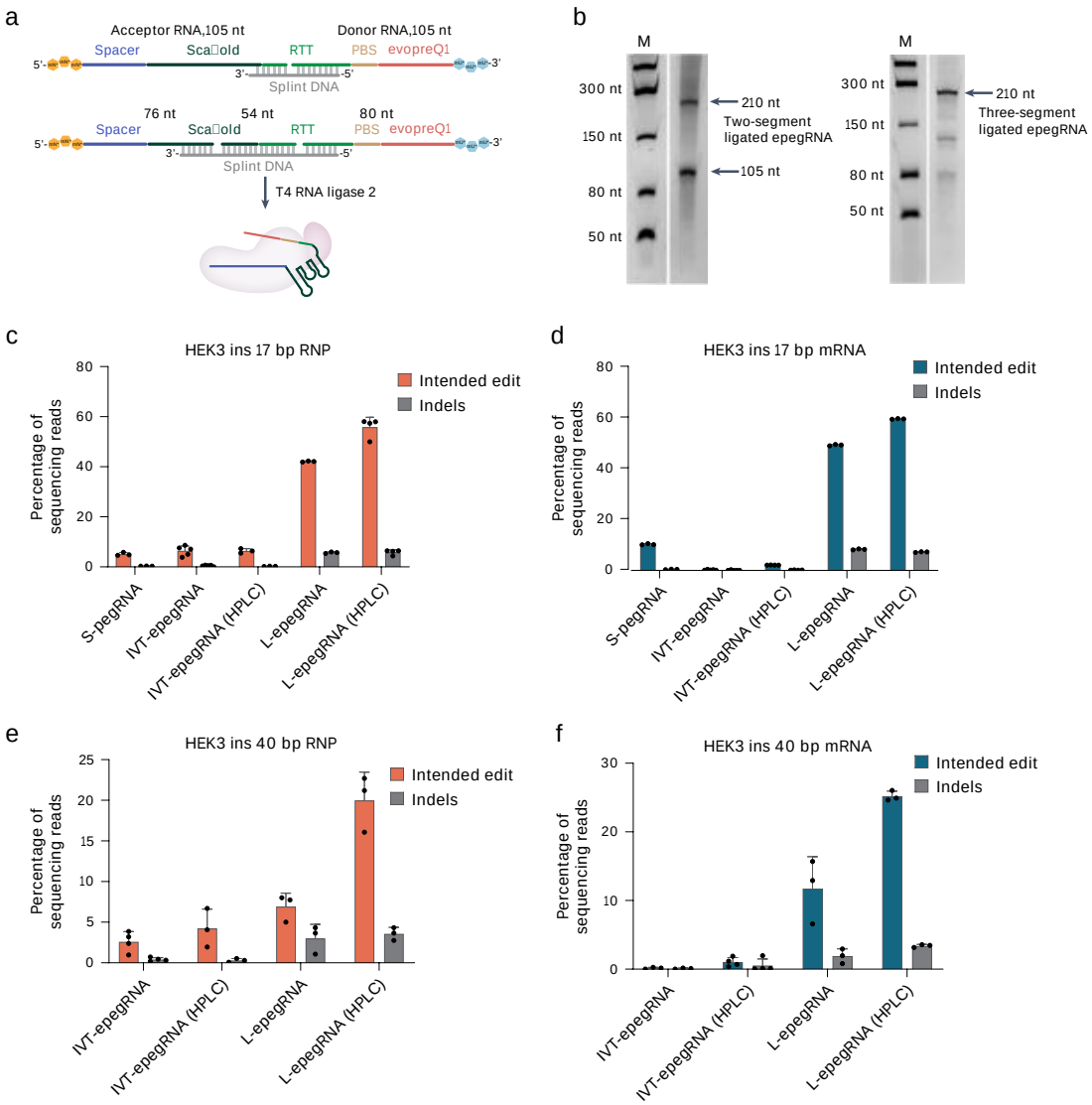
Next, we tested the ability of L-epegRNA to insert a 40 bp sequence, which is at the upper limit for classic PE. For this length of insertion, pegRNA cannot be synthesized by solid-phase synthesis. We obtained a 234 nt (91 + 59 + 84 nt) L-epegRNA for inserting a 40 bp loxP sequence at the *HEK3* site through three-fragment ligation (Extended Data Fig. 9b). The efficiencies by HPLC-purified L-epegRNA for 40 bp loxP insertion were 20.0 and 25.2% using RNP and RNA delivery of PE, respectively (Fig. 6e,f). In contrast, HPLC-purified IVT-epegRNA induced only 4.2 and 1.0% efficiency for the loxP insertion (Fig. 6e,f).

The synthesis of epegRNA by solid-phase chemistry is extremely challenging, with feasibility depending on sequence complexity, and the cost is usually not affordable (Fig. 6g)[6,19,35,39]. In contrast, the synthesis of L-epegRNA is straightforward, ensuring the purity of L-epegRNA, and the total cost is usually affordable (Fig. 6g). We compared the cost and production cycle between the four S-pegRNAs used above and the corresponding L-pegRNAs. The cost of L-pegRNA is lower than S-pegRNA, and its production cycle is faster (Extended Data Fig. 9c). Additionally, we calculated the cost and synthesis cycle required for L-pegRNAs and L-epegRNAs of different lengths at varying synthesis scales, both with and without HPLC purification (Extended Data Figs. 9d and 10). These low-cost, short-production cycle and high-potency L-epegRNAs enable broad applications.

Discussion

In this study, we optimized the conditions for RNA splint ligation and expanded its application to generate chemically modified sgRNA and pegRNA (Fig. 1b and Extended Data Fig. 1). The ligated pegRNA with

https://doi.org/10.1038/s41587-024-02394-x



g

Categories of chemically modified RNA	Feasibility of synthesis	Purity	Cost	Performance
pegRNA by solid-phase synthesis	Difficult	Low	High	Mediocre
epegRNA by solid-phase synthesis	Very difficult[a]	Low	Very high	Mediocre
L-epegRNA	Easy	High	Low	High

[a]Feasibility depends on sequence; synthesis is not feasible for epegRNA to insert a long sequence

Fig. 6 | Multi-fragments assembled L-epegRNA mediates efficient large insertions. a, Overview of epegRNA ligation design for insertion of 17 bp at the *HEK3* locus. The top shows the two-fragment ligation strategy using chemically synthesized 105 nt acceptor RNA and 105 nt donor RNA, to be ligated by splint DNA. The bottom shows the three-fragment ligation using chemically synthesized 76, 54 and 80 nt RNA ligated by the 84 nt splint DNA. **b**, The left shows urea-PAGE of two-fragment ligation products; the right shows urea-PAGE of three-fragment ligation products. M, marker. **c**–**f**, Comparison of PE efficiencies using different pegRNAs for insertion of 17 bp (**c**,**d**) or 40 bp (**e**,**f**) at the *HEK3* site in HEK293T cells. PE was delivered in RNP (**c**,**e**) or RNA format (**d**,**f**), respectively. For each sample, 70 pmol of PEmax protein, 140 pmol of epegRNA and 60 pmol of nicking sgRNA were used for RNP delivery, and 2 μg of mRNA encoding PEmax was cotransfected with 180 pmol of epegRNA and 60 pmol of nicking sgRNA for RNA delivery. For **c**,**d**, S-pegRNA indicates full-length pegRNAs that were chemically synthesized with 3 nt chemical modifications at both ends. For **c**–**f**, IVT-epegRNA indicates epegRNA produced by IVT; IVT-epegRNA (HPLC) indicates epegRNA produced by IVT with HPLC purification; L-epegRNA indicates epegRNA generated through three-fragment ligation. L-epegRNA (HPLC) indicates L-epegRNA with HPLC purification. Data and error bars represent the mean and standard deviation from at least three independent biological replicates. **g**, Comparison of solid-phase synthesized pegRNA and epegRNA with L-epegRNA.

the best design (L-epegRNA) demonstrated superior editing activity via RNP and RNA delivery of PE, showing potency in both cell lines and human primary cells, thus promoting therapeutic applications using the PE system (Figs. 3–5). For the same loci, our optimized RNP delivery efficiencies are 2.8–3.6 times higher in HEK293T cells and 2.4–5.8 times higher in human primary T cells than previously reported data (Figs. 3b and 5a)[17,20].

RNP delivery of Cas9 often exhibits superior editing efficiency compared to plasmid delivery of Cas9 (ref. 56). However, plasmid transfection usually achieves much higher efficiencies than RNP delivery of PE (refs. 5,17,20,21). With the enhanced activity of L-epegRNA and optimization of various parameters for RNP delivery of PE, we demonstrated that RNP delivery of PE can be superior to plasmid delivery for most comparisons examined (Fig. 3g–k). We analyzed pegRNA sequences for plasmids with poor editing inefficiencies, and found no T-rich regions. Among all pegRNA sequences compared with plasmids, there were no regions with more than three consecutive Ts. Therefore, the sites with poor editing efficiencies by plasmids in our study are unlikely to be due to premature termination resulting from high T content. Moreover, we speculate that our method may further enhance efficiency in some T-rich pegRNA sequences in comparison with plasmid delivery. L-epegRNA can be applied for ex vivo therapy and in vivo delivery in RNA format by encapsulating it into nonviral vectors such as lipid nanoparticles[57–59]. Further investigation into its in vivo performance is feasible, leveraging the high efficiencies and reasonable cost of L-epegRNA.

We recommend using PEmax$^{\Delta RH}$ for RNP delivery of PE in most cases due to its ease of production and superior efficacy (Extended Data Fig. 7). However, PE proteins lacking the RNase H domain do not perform well when the RTT region contains secondary structures[53]. Therefore, we recommend using PEmax for RNP delivery in such cases (Fig. 6). The addition of MLH1dn protein or its encoding mRNA (referred to as PE4 and PE5) did not enhance editing efficiency in RNP or RNA delivery of PE (Extended Data Fig. 8a,b), consistent with a recent report[18]. PE4 and PE5 have been demonstrated to have better performance than PE2 and PE3, respectively, via plasmid delivery, which we successfully replicated in our study as a positive control[17]. It is possible that MLH1dn needs to be persistently expressed from plasmid to be functional. Further investigation is warranted to understand the efficiency gaps of PE4/5 across different delivery formats.

We used T4 RNA Ligase 2, which exhibits a preference for double-stranded ligation but can also catalyze the joining of single-stranded RNA molecules. Therefore, reduced ligation efficiency can be observed in the absence of splint DNA (Extended Data Fig. 1a). However, the presence of splint DNA can sequester the 3′ end of the acceptor RNA and the 5′ end of the donor RNA, thereby preventing self-circularization[35,42]. Additionally, the 2′ OH modification of ribose used in our study sterically minimizes the ligase reaction at the 3′ OH (ref. 35). For three-segment ligation, the middle fragment is fully complementary to the splint DNA, preventing its self-circularization (Fig. 6a,b). The efficiencies of both two- and three-segment ligation are already high, so HPLC purification only provides a small improvement in PE efficiency in many cases (Figs. 1b–f and 6c,d). We used more than 25 different splint DNA sequences for various sites in our study. We noticed that the ligation efficiencies generally maintained high, illustrating the versatility of splint ligation. Since the skillsets for HPLC purification are not common in many biomedical research laboratories, L-epegRNA can bypass the costly chromatographic purification steps for research purposes.

It is extremely challenging to chemically synthesize RNA molecules of approximately 200 nt or beyond[24,35–37,39]. By using a three-segment ligation approach, we successfully generated 210 and 234 nt epegRNAs with chemical modifications, achieving high-efficiency 17 and 40 bp insertions via both RNP and RNA delivery (Fig. 6c–f). These results underscore the potential of L-epegRNA in facilitating the insertion of longer sequences using PE. L-epegRNA can be applied to insert a landing pad for integrases and recombinases, enabling the targeted insertion of large fragments for research and therapeutic applications (Fig. 6e,f)[14,49,50,60]. Furthermore, using the three-segment ligation approach, it is possible to incorporate additional RNA structures into chemically modified epegRNA. For example, MS2 can be integrated into L-pegRNA to recruit effector proteins to the PE targeting site. This optimized splint ligation protocol can be applied to synthesize high-quality long RNA or nucleic acids with chemical modifications. For instance, it can be used to generate stable sgRNA with MS2/PP7 motifs for effector recruitment and chemically modified oligonucleotides to bind endogenous adenosine deaminases acting on RNA enzymes[61–64].

In conclusion, our study demonstrates that optimized RNA ligation can be applied to generate chemically modified epegRNA for efficient PE, ranging from point mutations to 40 bp insertions in human cells. L-epegRNA and the optimized RNP delivery of PE facilitate efficient PE, surpassing the commonly used PE plasmid method. Our findings advance the clinical translation of the PE system and other nucleic acid therapeutics that require site-specific modifications of synthetic RNA.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-024-02394-x.

References

1. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
2. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR–Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
3. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR–Cas9. *Science* **346**, 1258096 (2014).
4. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
5. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
6. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2021).
7. Zhang, G. et al. Enhancement of prime editing via xrRNA motif-joined pegRNA. *Nat. Commun.* **13**, 1856 (2022).
8. Liu, Y. et al. Enhancing prime editing by Csy4-mediated processing of pegRNA. *Cell Research* **31**, 1134–1136 (2021).
9. Yin, H. et al. Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.* **35**, 1179–1187 (2017).
10. Hendel, A. et al. Chemically modified guide RNAs enhance CRISPR–Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).
11. Ryan, D. E. et al. Improving CRISPR–Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res.* **46**, 792–803 (2018).
12. Chen, Q., Zhang, Y. & Yin, H. Recent advances in chemical modifications of guide RNA, mRNA and donor template for CRISPR-mediated genome editing. *Adv. Drug Deliv. Rev.* **168**, 246–258 (2021).
13. Meisel, R. CRISPR–Cas9 gene editing for sickle cell disease and β-thalassemia. *N. Engl. J. Med.* **384**, e91 (2021).
14. Wang, J. et al. Efficient targeted insertion of large DNA fragments without DNA donors. *Nat. Methods* **19**, 331–340 (2022).

15. Solomon, A. CRISPR–Cas9 in vivo gene editing for transthyretin amyloidosis. *N. Engl. J. Med.* **385**, 1721–1722 (2021).
16. Finn, J. D. et al. A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing. *Cell Rep.* **22**, 2227–2235 (2018).
17. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652.e5629 (2021).
18. Everette, K. A. et al. Ex vivo prime editing of patient haematopoietic stem cells rescues sickle-cell disease phenotypes after engraftment in mice. *Nat. Biomed. Eng.* **7**, 616–628 (2023).
19. Liu, B. et al. Targeted genome editing with a DNA-dependent DNA polymerase and exogenous DNA-containing templates. *Nat. Biotechnol.* **42**, 1039–1045 (2024).
20. Petri, K. et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **40**, 189–193 (2021).
21. Ponnienselvan, K. et al. Reducing the inherent auto-inhibitory interaction within the pegRNA enhances prime editing efficiency. *Nucleic Acids Res.* **51**, 6966–6980 (2023).
22. Zhang, W. et al. Enhancing CRISPR prime editing by reducing misfolded pegRNA interactions. *eLife* https://doi.org/10.7554/eLife.90948.2 (2024).
23. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862 (1981).
24. Ryczek, M., Pluta, M., Blaszczyk, L. & Kiliszek, A. Overview of methods for large-scale RNA synthesis. *Appl. Sci.* **12**, 1543 (2022).
25. Wang, G. et al. mRNA produced by VSW-3 RNAP has high-level translation efficiency with low inflammatory stimulation. *Cell Insight* **1**, 100056 (2022).
26. Dousis, A., Ravichandran, K., Hobert, E. M., Moore, M. J. & Rabideau, A. E. An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. *Nat. Biotechnol.* **41**, 560–568 (2023).
27. Karikó, K. et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol. Ther.* **16**, 1833–1840 (2008).
28. Kormann, M. S. et al. Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. *Nat. Biotechnol.* **29**, 154–157 (2011).
29. Warren, L. et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618–630 (2010).
30. Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D. mRNA vaccines - a new era in vaccinology. *Nat. Rev. Drug Discov.* **17**, 261–279 (2018).
31. Hertler, J. et al. Synthesis of point-modified mRNA. *Nucleic Acids Res.* **50**, e115 (2022).
32. Kao, C., Zheng, M. & Rüdisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268–1272 (1999).
33. Liu, Y. et al. Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* **522**, 368–372 (2015).
34. Moody, E. R., Obexer, R., Nickl, F., Spiess, R. & Lovelock, S. L. An enzyme cascade enables production of therapeutic oligonucleotides in a single operation. *Science* **380**, 1150–1154 (2023).
35. Stark, M. R. & Rader, S. D. Efficient splinted ligation of synthetic RNA using RNA ligase. *Methods Mol. Biol.* **1126**, 137–149 (2014).
36. Bartosik, K., Debiec, K., Czarnecka, A., Sochacka, E. & Leszczynska, G. Synthesis of nucleobase-modified RNA oligonucleotides by post-synthetic approach. *Molecules* **25**, 3344 (2020).
37. Flamme, M., McKenzie, L. K., Sarac, I. & Hollenstein, M. Chemical methods for the modification of RNA. *Methods* **161**, 64–82 (2019).
38. Kurschat, W. C., Müller, J., Wombacher, R. & Helm, M. Optimizing splinted ligation of highly structured small RNAs. *RNA* **11**, 1909–1914 (2005).
39. Wang, S., Chen, D., Gao, L. & Liu, Y. Short oligonucleotides facilitate co-transcriptional labeling of RNA at specific positions. *J. Am. Chem. Soc.* **144**, 5494–5502 (2022).
40. Van Giesen, K. J. D., Thompson, M. J., Meng, Q. & Lovelock, S. L. Biocatalytic synthesis of antiviral nucleosides, cyclic dinucleotides, and oligonucleotide therapies. *JACS Au* **3**, 13–24 (2023).
41. Hengesbach, M. et al. RNA intramolecular dynamics by single-molecule FRET. *Curr. Protoc. Nucleic Acid Chem.* **Ch. 11**, Unit 11.12 (2008).
42. Stark, M. R., Pleiss, J. A., Deras, M., Scaringe, S. A. & Rader, S. D. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* **12**, 2014–2019 (2006).
43. Nandakumar, J. & Shuman, S. How an RNA ligase discriminates RNA versus DNA damage. *Mol. Cell* **16**, 211–221 (2004).
44. Chakravarty, A. K. & Shuman, S. The sequential 2′,3′-cyclic phosphodiesterase and 3′-phosphate/5′-OH ligation steps of the RtcB RNA splicing pathway are GTP-dependent. *Nucleic Acids Res.* **40**, 8558–8567 (2012).
45. Gamper, H. et al. Enzymatic synthesis of RNA standards for mapping and quantifying RNA modifications in sequencing analysis. *Methods Enzymol.* **692**, 127–153 (2023).
46. Bullard, D. R. & Bowater, R. P. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem. J.* **398**, 135–144 (2006).
47. Viollet, S., Fuchs, R. T., Munafo, D. B., Zhuang, F. & Robb, G. B. T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. *BMC Biotechnol.* **11**, 72 (2011).
48. Li, X. et al. Enhancing prime editing efficiency by modified pegRNA with RNA G-quadruplexes. *J. Mol. Cell. Biol.* **14**, mjac022 (2022).
49. Anzalone, A. V. et al. Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2021).
50. Yarnall, M. T. N. et al. Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat. Biotechnol.* **41**, 500–512 (2022).
51. Zheng, C. et al. A flexible split prime editor using truncated reverse transcriptase improves dual-AAV delivery in mouse liver. *Mol. Ther.* **30**, 1343–1351 (2022).
52. Gao, Z. et al. A truncated reverse transcriptase enhances prime editing by split AAV vectors. *Mol. Ther.* **30**, 2942–2951 (2022).
53. Doman, J. L. et al. Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* **186**, 3983–4002.e3926 (2023).
54. An, J. et al. Enhancement of the viability of T cells electroporated with DNA via osmotic dampening of the DNA-sensing cGAS–STING pathway. *Nat. Biomed. Eng.* **8**, 149–164 (2023).
55. Han, W. et al. Base editing of the HBG promoter induces potent fetal hemoglobin expression with no detectable off-target mutations in human HSCs. *Cell Stem Cell* **30**, 1624–1639.e1628 (2023).
56. Qiu, H. Y., Ji, R. J. & Zhang, Y. Current advances of CRISPR-Cas technology in cell therapy. *Cell Insight.* **1**, 100067 (2022).

57. Yin, H. et al. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* **32**, 551–553 (2014).
58. Yin, H. et al. Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat. Biotechnol.* **34**, 328–333 (2016).
59. Yin, H. et al. Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* **15**, 541–555 (2014).
60. Sun, C. et al. Precise integration of large DNA sequences in plant genomes using PrimeRoot editors. *Nat. Biotechnol.* **42**, 316–327 (2023).
61. Katrekar, D. et al. In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. *Nat. Methods* **16**, 239–242 (2019).
62. Khosravi, H. M. & Jantsch, M. F. Site-directed RNA editing: recent advances and open challenges. *RNA Biol.* **18**, 41–50 (2021).
63. Fu, Y. et al. CRISPR-dCas9 and sgRNA scaffolds enable dual-colour live imaging of satellite sequences and repeat-enriched individual loci. *Nat. Commun.* **7**, 11707 (2016).
64. Zalatan, J. G. et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

Methods

Mammalian cell culture

Human HEK293T, Huh-7, HeLa, U2OS and K562 cells were procured from ATCC. HEK293T and Huh-7 cells were cultured in DMEM (Thermo Fisher). HeLa and U2OS cells were maintained in DMEM/F12 (1:1) medium (HyClone), while K562 cells were cultured in RPMI-1640 medium with L-glutamine (Gibco). Each culture medium was supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin (P–S). All cell types were incubated, maintained and cultured at 37 °C with 5% CO_2, and were routinely tested to ensure the absence of mycoplasma contamination.

Plasmid construction

Plasmids for the mammalian expression of prime editors and other proteins were cloned using the pEASY-Basic Seamless Cloning and Assembly Kit (TransGen Biotech, cat. no. CU201). Plasmids designed for expressing pegRNAs were constructed using a pGCL acceptor plasmid[14]. The pegRNA sequences were obtained by PCR using Phanta Max Super-Fidelity DNA Polymerase (Vazyme, cat. no. P505). The sequences of pegRNA constructs are provided in Supplementary Table 1. The U6 promoter-driven nicking guide RNA (gRNA) mammalian expression plasmid was created using the gRNA cloning vector (Addgene no. 41824). Nicking gRNA cloning was executed by phosphorylation of oligonucleotides corresponding to spacer sequences with T4 polynucleotide kinase (New England Biolabs, cat. no. M0201), followed by annealing and ligation into BbsI-digested Gcl vector. Plasmids designed for expressing the prime editor protein were derived from pCMV-PEmax (Addgene no. 174820) and pCMV-PEmax-P2A-hMLH1dn (Addgene no. 174828).

To obtain the PE2 protein, we constructed a pET28a-PE2-His vector that underwent *E. coli* codon optimization. The SpCas9 (H840A) sequence was derived from pET28a-Cas9-His (Addgene no. 98158) and subjected to point mutations[65,66]. The nuclear localization signal, linker and M-MLV reverse-transcription sequence were synthesized by GenScript and codon-optimized for *E. coli*[5]. For the PEmax protein expression vector (pET28a-PEmax-His), mutations were introduced based on the pET28a-PE2-His vector, and the nuclear localization signal and linker were replaced as reported previously[17,67,68]. The protein expression vectors for $PE2^{\Delta RH}$ (pET28a-PE2 ΔRH-His) and $PEmax^{\Delta RH}$ (pET28a-PEmax ΔRH-His) were generated by excluding the RNase H domain from M-MLV ORFs[51,52]. To express the MLH1dn protein, we generated an *E. coli* codon-optimized MLH1dn sequence (GenScript) and cloned it into the pET28a-His expression vector[17].

Purification of PE proteins and T4 RNA Ligase 2

Rosetta (DE3) competent cells (WEIDI, cat. no. EC1010) and BL21 (DE3) competent cells (WEIDI, cat. no. EC1002) were transformed with pET28a-PE2 or pET30C-gp24.1, respectively, following the manufacturer's instructions. A 5 ml overnight culture grown from a single colony in Luria-Bertani medium with 50 μg ml^{-1} kanamycin was transferred into 0.5 l of the same medium and cultured at 37 °C. Once the optical density at 600 nm of the culture reached 0.7–0.8, the expressions of PE and T4 RNA Ligase 2 were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (Sangon Biotech, cat. no. A1004870005) for 18 h at 18 °C and for 3–6 h at 37 °C, respectively. Cells were collected and lysed in a lysis buffer using a high-pressure homogenizer (ATS, cat. no. AH1500). The supernatant was collected after centrifugation and then filtered with 0.45 μm filters. Protein purification was carried out through affinity purification followed by size exclusion chromatography. In brief, the clarified lysate was loaded onto a HisTrap HP column (Cytiva) in the NGC Quest 10 Chromatography System (BioRad). The column was prebalanced in lysis buffer. Protein was eluted in buffer B1 using a gradient program. Different elution fractions were collected and then verified by SDS–PAGE to identify the target protein. The resulting protein was then loaded onto a HiLoad 16/60 Superdex 200 pg column (Cytiva) in buffer B2. The eluted protein was concentrated using centrifugal filters (Millipore) and stored in buffer B2 at −80 °C with 10% glycerol[69–71]. The other four proteins, $PE^{\Delta RH}$, PEmax, $PEmax^{\Delta RH}$ and MLH1dn, were also purified using the method described above.

The lysis buffer for purifying PE proteins and MLH1dn contains 20 mM Tris-HCl and 500 mM NaCl, pH 7.4. Buffer B1 consists of 20 mM Tris-HCl, 500 mM NaCl and 500 mM imidazole, pH 7.4 and buffer B2 consists of 20 mM Tris-HCl and 200 mM NaCl, pH 7.4. The lysis buffer for purifying T4 RNA Ligase 2 contains 50 mM Tris-HCl, 250 mM NaCl and 10%(w/v) sucrose, pH 7.5. Buffer B1 consists of 50 mM Tris-HCl, 250 mM NaCl, 10%(v/v) glycerol and 500 mM imidazole, pH 8.0, and buffer B2 consists of 10 mM Tris-HCl, 0.1 mM EDTA, 0.1 mM dithiothreitol (DTT), 10% glycerol and 90 mM NaCl, pH 7.5.

IVT of pegRNA

IVT was performed using T7 RNA polymerase in a reaction mixture containing 5× T7 RNA Polymerase Reaction Buffer (0.2 M Tris-HCl pH 8.0, 0.125 M NaCl, 40 mM $MgCl_2$, 10 mM Spermidine-$(HCl)_3$), 50 mM DTT, pyrophosphatase and 0.5–1 μg DNA in a total volume of 50 μl. The nucleoside triphosphates (GTP, ATP, CTP, UTP) were supplemented to a final concentration of 2 mM (Vazyme Biotech Co., Ltd, cat. nos. DD4108-PA, DD4106-PA, DD4107-PA, DD4105-PA)[72]. For transcripts requiring a 5′-phosphate for subsequent ligation, the final concentration of GTP was adjusted to 2 mM and guanosine monophosphate (Sigma, cat. no. G8377) was added to a final concentration of 16 mM. The DNA template was then enzymatically digested by adding 4 U of RQ1 RNase-Free DNase (Promega, cat. no. M6101) at 37 °C for 15 min. The resulting products were purified using the Monarch RNA Cleanup Kit (New England Biolabs, cat. no. T2040). The pegRNAs obtained by IVT were treated with CIP enzyme (New England Biolabs, cat. no. M0525S) before being used for cell transfection.

Nucleic acids synthesis

Chemically modified short 32 nt RNA, nicking sgRNA, S-pegRNA and unmodified splint DNA fragments were synthesized by the GenScript Corporation. Other chemically modified RNAs were synthesized by General Bio. The sequences are provided in Supplementary Tables 1–3.

RNA ligation

For RNA ligation, 200 pmol of acceptor RNA, 200 pmol of donor RNA and equimolar amounts of splint DNA, together with 5× annealing buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl), were added for a total volume of 30 μl. This mixture was incubated for 3 min at 70 °C and then gradually cooled to room temperature at a rate of 0.1 °C s^{-1}. Subsequently, ligation was performed by adding 4 μl of 10× T4 RNA Ligase 2 buffer (500 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 10 mM DTT, 4 mM ATP), 3 μl of T4 RNA Ligase 2 (1 μg μl^{-1}) and 3 μl of nuclease-free water, followed by incubation at 37 °C for 1 h. The annealing and ligation process was repeated twice, with the addition of 1.5 μl of T4 RNA ligase in the second and third ligation rounds, followed by incubation at 37 °C for 30 min. The DNA splint was digested with RNase-Free DNase (Promega, M6101) for 30 min at 37 °C. The ligation products were purified using the Monarch RNA Cleanup Kit (New England Biolabs, cat. no. T2040), which has a reduced recovery rate for RNAs shorter than 25 nt.

HPLC purification of ligated pegRNA

HPLC purification of the ligated pegRNA was performed using the Agilent 1260 Infinity II system with a column packed with nonporous polystyrene-divinylbenzene copolymer microspheres (Agilent, cat. no. PL1512-5802). HPLC was conducted with 0.1 M triethylammonium acetate (TEAA, pH 7.0) (Sigma, cat. no. 625718) buffer and HPLC-grade acetonitrile (Sigma, cat. no. 34851) as the mobile phase. The initial equilibrium condition consisted of 12% acetonitrile and 88% TEAA. Subsequently, the RNA sample was loaded onto the column and subjected to a linear gradient elution with acetonitrile, transitioning from 13 to

18% acetonitrile over 20 min at a flow rate of 0.4 ml min^{-1}. The gradient was adjusted based on the RNA length and modifications, with longer RNA strands requiring a higher proportion of acetonitrile for effective gradient elution. Fine-tuning the gradient change rate of acetonitrile and the flow rate can improve the resolution of HPLC. An example of HPLC purification of L-epegRNA is presented in Extended Data Fig. 10 for the users' reference. The HPLC purifications of ligated RNA were performed in our laboratory.

IVT of PE mRNA

The IVT of PE2 and PEmax mRNA followed a previously established protocol using VSW3 RNA polymerase[25,73]. In brief, the VSW3 promoter sequence (5′-TTAATTGGGCCACCTATA-3′) was introduced into PE constructs to serve as the template for PCR, and the reverse primer introduced a Poly A tail to the 3′ untranslated region of the resulting PCR product[6,74]. The IVT reaction mixture (10 µl) comprised 40 mM Tris-HCl (pH 8.0), 16 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine, 35 ng $µl^{-1}$ of the purified PCR product as a template, 1.5 U $µl^{-1}$ of RNase inhibitor, 0.2 µM inorganic pyrophosphatase, 4 mM of each of the nucleoside triphosphates and 0.15 µM of VSW3 RNA polymerase. The reaction was conducted at 25 °C for 12–16 h (ref. 25). RNase-Free DNase (Promega, cat. no. M6101) was used to eliminate the DNA template, and the resulting mRNA was precipitated using lithium chloride. The mRNA was enzymatically capped using the Vaccinia Capping System (New England Biolabs, cat. no. M2080S) and further purified using the Monarch RNA Cleanup Kit (New England Biolabs, cat. no. T2040).

Cell transfection

All electroporation procedures were conducted using the Lonza 4D Nucleofector system in B1mix buffer following established protocols[75]. Each procedure comprised 2×10^5 cells in 20 µl. RNPs were preincubated at room temperature for 15 min before electroporation. The electroporation programs used were as follows: EO-115 for HEK293T, EO-138 for Huh-7 and K562, DS-137 for U2OS and DN-130 for HeLa cells. The initial RNP electroporation applied 140 pmol of PE protein, 186 pmol of pegRNA and 62 pmol of nicking sgRNA. The optimized RNP electroporation used 70 pmol of PEmax protein, 140 pmol of L-epegRNA and 60 pmol of nicking sgRNA. The mRNA electroporation included 2 µg of PEmax mRNA, 180 pmol of L-epegRNA and 60 pmol of nicking sgRNA. For plasmid electroporation, 800 ng of PE plasmid, 200 ng of pegRNA plasmid and 83 ng of nicking sgRNA plasmid were used for HEK293T, Huh-7 and K562, and the doses of plasmids were doubled for U2OS and HeLa cells, followed previously established procedures[5,17].

Isolation and electroporation of primary human T cells and $CD34^+$ human HSPCs

Peripheral blood mononuclear cells (PBMCs) were purchased from Milestone Biotechnologies. $CD3^+$ T cells were isolated from PBMCs using the EasySep Human T Cell Isolation Kit (STEMCELL) and cultured in X-Vivo15 medium (Lonza) supplemented with 5% fetal bovine serum (Gibco), human IL-2 (50 ng ml^{-1}, Peprotech), human IL-7 (10 ng ml^{-1}, Peprotech) and 1% P–S. To activate T cells before electroporation, CD3/CD28 Dynabeads (Thermo Fisher, cat. no. 11131D) were added to the culture at a 1:3 ratio and cultured for 3 days. After removing the beads, the cells were allowed to rest for 5–7 h before electroporation, which was carried out using the EO-138 program in B1mix buffer[54,75]. The electroporation details are the same as described above for RNP and mRNA delivery. Three days after electroporation, cells were collected by centrifugation, and genomes were isolated.

Cryopreserved $CD34^+$ HSPCs were purchased from Milestone Biotechnologies and thawed according to the manufacturer's instructions. $CD34^+$ HSPCs were cultured in StemSpan SFEM (StemCell Technologies) supplemented with human stem cell factor, human thrombopoietin and human Flt3 ligand (100 ng ml^{-1} for each, Peprotech), with 1% P–S. $CD34^+$ human HSPCs were electroporated with RNA or RNP 24 h post-thaw using the Lonza 4D Nucleofector system with the CM-137 program in B1mix buffer[55,75]. Each electroporation procedure consisted of 20 µl containing 2×10^5 $CD34^+$ human HSPCs. The doses of RNA or RNP followed the details described above. Genomes were collected 3 days after electroporation.

Genomic DNA extraction

The collected cells were centrifuged at 300*g* for 5 min, and the pellet was resuspended in 50 µl lysis buffer (10 mM Tris-HCl pH 8.0, 0.05% SDS) and 1 µl of proteinase K (Thermo Fisher, cat. no. EO0491). The mixture was then incubated for 2 h at 37 °C, followed by an additional 30 min at 85 °C to inactivate the proteinase K. The editing loci were subsequently PCR amplified and prepared for amplicon sequencing (Illumina) (Supplementary Tables 4 and 5).

Next-generation sequencing analysis

PE target sites were amplified and sequenced using the Illumina sequencing platform (Novaseq 6000). Genomic DNA samples were amplified with primers specific for Illumina adapters using Phanta Max Super-Fidelity DNA Polymerase (Vazyme, cat. no. P505). The first round of PCR reactions was carried out with the following parameters: 95 °C for 3 min, followed by 23 PCR cycles and a final 72 °C extension for 5 min. In the second round of PCR reactions, 13 cycles were performed. The primers contained Illumina adapters and a 7- and/or 8-bp index. The PCR products were purified before sequencing on the Illumina sequencing platform. Raw paired-end reads were merged using the fastp software to generate full-length reads[76]. Reads with a mean quality score <30 or adapter contamination were discarded using Trimmomatic[77]. Alignment of reads to a reference sequence was performed using CRISPResso2 (ref. 78). Based on the deep sequencing results, target editing was classified into two categories, intended edits and indels, which included undesired mutations[5,17].

Deep sequencing of ligated pegRNA

IVT pegRNA and ligated pegRNA were reverse transcribed into complementary DNA (cDNA) using specific reverse-transcription primers. PCR amplification was performed on the cDNA using primers containing Illumina adapters and a 7 or 8 bp index. The resulting products were then subjected to next-generation sequencing. Bioinformatic analysis was conducted as described above. The primer sequences are provided in Supplementary Table 4.

Statistical analysis

GraphPad Prism v.9 software was applied to analyze the data. Two-tailed Student's *t*-tests were used to compare differences between two groups, and one-way analysis of variance (ANOVA) with Tukey's multiple comparisons test was performed for comparisons between multiple groups. We independently conducted the experiments related to the gel images shown in the paper three times, and obtained similar results.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

High-throughput sequencing data have been deposited in the National Center for Biotechnology Information Sequence Read Archive database under accession PRJNA1067838 (ref. 79). Source data are provided with this paper.

References

65. Liang, Z. et al. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.* **8**, 14261 (2017).

66. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
67. Liu, P. et al. Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat. Commun.* **12**, 2121 (2021).
68. Park, S.-J. et al. Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor. *Genome Biol.* **22**, 170 (2021).
69. Lu, S. et al. Fast and sensitive detection of SARS-CoV-2 RNA using suboptimal protospacer adjacent motifs for Cas12a. *Nat. Biomed. Eng.* **6**, 286–297 (2022).
70. Wu, J. et al. Characterization of a thermostable Cas12a ortholog. *Cell Insight.* **2**, 100126 (2023).
71. Zhang, H. X. et al. Cas12a-based one-pot SNP detection with high accuracy. *Cell Insight.* **2**, 100080 (2023).
72. Shi, Y.-J. et al. DNA topology regulates PAM-Cas9 interaction and DNA unwinding to enable near-PAMless cleavage by thermophilic Cas9. *Mol. Cell* **82**, 4160–4175.e4166 (2022).
73. Xia, H. et al. Psychrophilic phage VSW-3 RNA polymerase reduces both terminal and full-length dsRNA byproducts in in vitro transcription. *RNA Biol.* **19**, 1130–1142 (2022).
74. Yi, J. et al. Co-delivery of Cas9 mRNA and guide RNAs edits hepatitis B virus episomal and integration DNA in mouse and tree shrew models. *Antiviral Res.* **215**, 105618 (2023).
75. Zhang, C.-P. et al. Efficient non-viral delivery of macromolecules in human primary hematopoietic stem cells and lymphocytes. *J. Mol. Cell. Biol.* **15**, mjad018 (2023).
76. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
77. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
78. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
79. Lei, X. et al. Rapid generation of long, chemically modified pegRNAs for prime editing. *NCBI Bioproject* www.ncbi.nlm.nih.gov/bioproject/1067838 (2024).

Acknowledgements

This work is kindly supported by National Science and Technology Major Project (grant no. 2023ZD0500600), National Key R&D Program of China (grant nos. 2019YFA0802801 and 2018YFA0801401 to H.Y. and 2022YFF1002801 to Ying Zhang), the Ministry of Agriculture and Rural Affairs of China, Key R&D Program of Hubei Province (grant nos. 2022BCA089 to H.Y. and 2022ACA005 to Ying Zhang), the National Natural Science Foundation of China (grant nos. 31871345 and 32071442 to H.Y. and 31972936 to Ying Zhang), the Fundamental Research Funds for the Central Universities (grant nos. 2042022dx0003 and 2042022kf1190), and the startup funding from Wuhan University (to H.Y. and Ying Zhang). We thank the members of core facility of Medical Research Institute at Wuhan University for their technical support.

Author contributions

H.Y. conceived, designed and managed the project. X.L., A.H., D.C. and X.W. performed most of the experiments with the help of R.J., J.W., Yuming Zhang, S.L., K.Z. and Q.C. Ying Zhang edited the manuscript. Yizhou Zhang performed bioinformatic analysis. X.L. and H.Y. analyzed the data. X.L. and H.Y. wrote the paper with inputs from all authors.

Competing interests

H.Y., Ying Zhang, X.L. and X.W. have filed a patent application on ligation of pegRNA through Wuhan University (application number PCT/CN2024/078744). The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41587-024-02394-x.

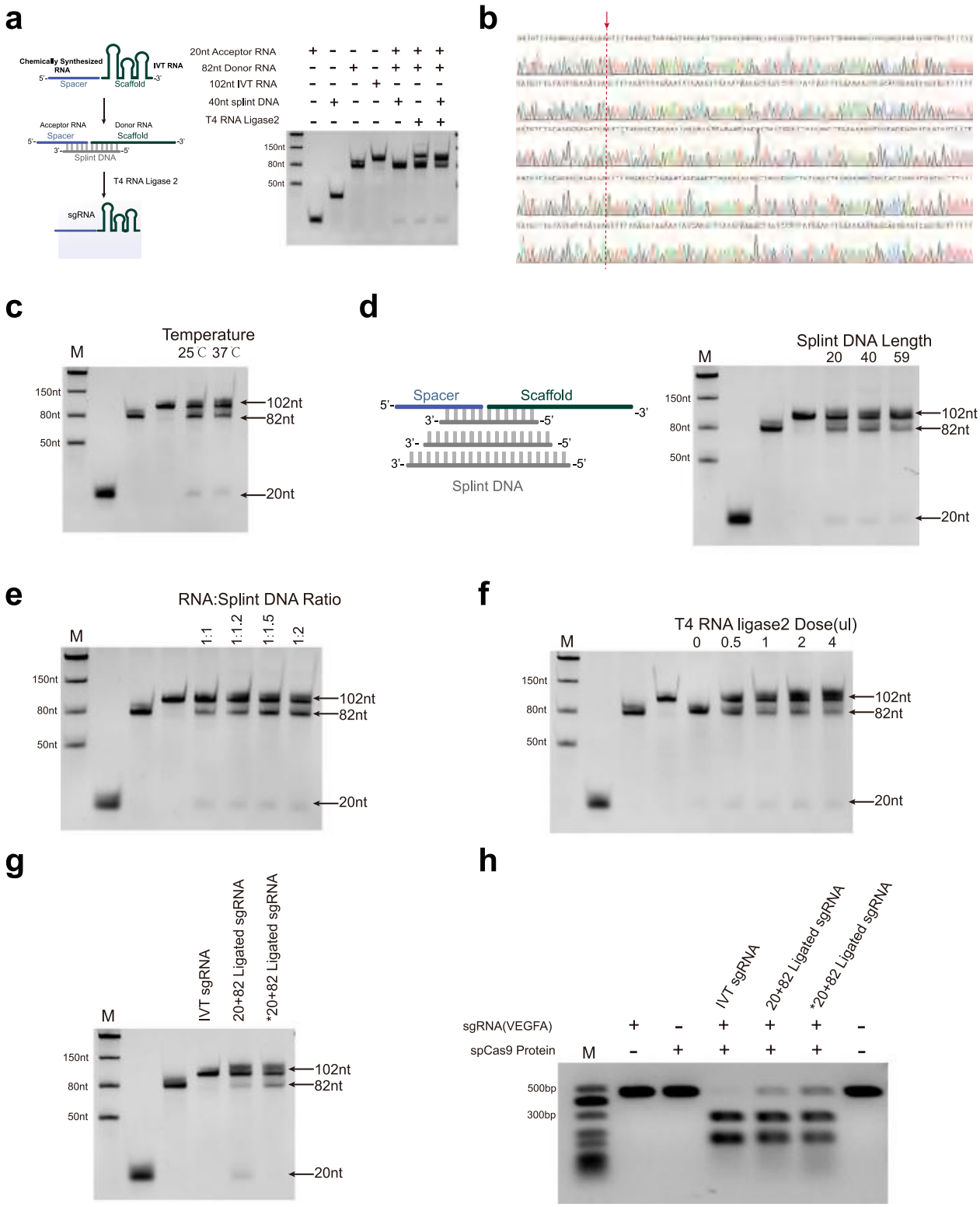
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-024-02394-x.

Correspondence and requests for materials should be addressed to Hao Yin.

Peer review information *Nature Biotechnology* thanks Chang Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

https://doi.org/10.1038/s41587-024-02394-x

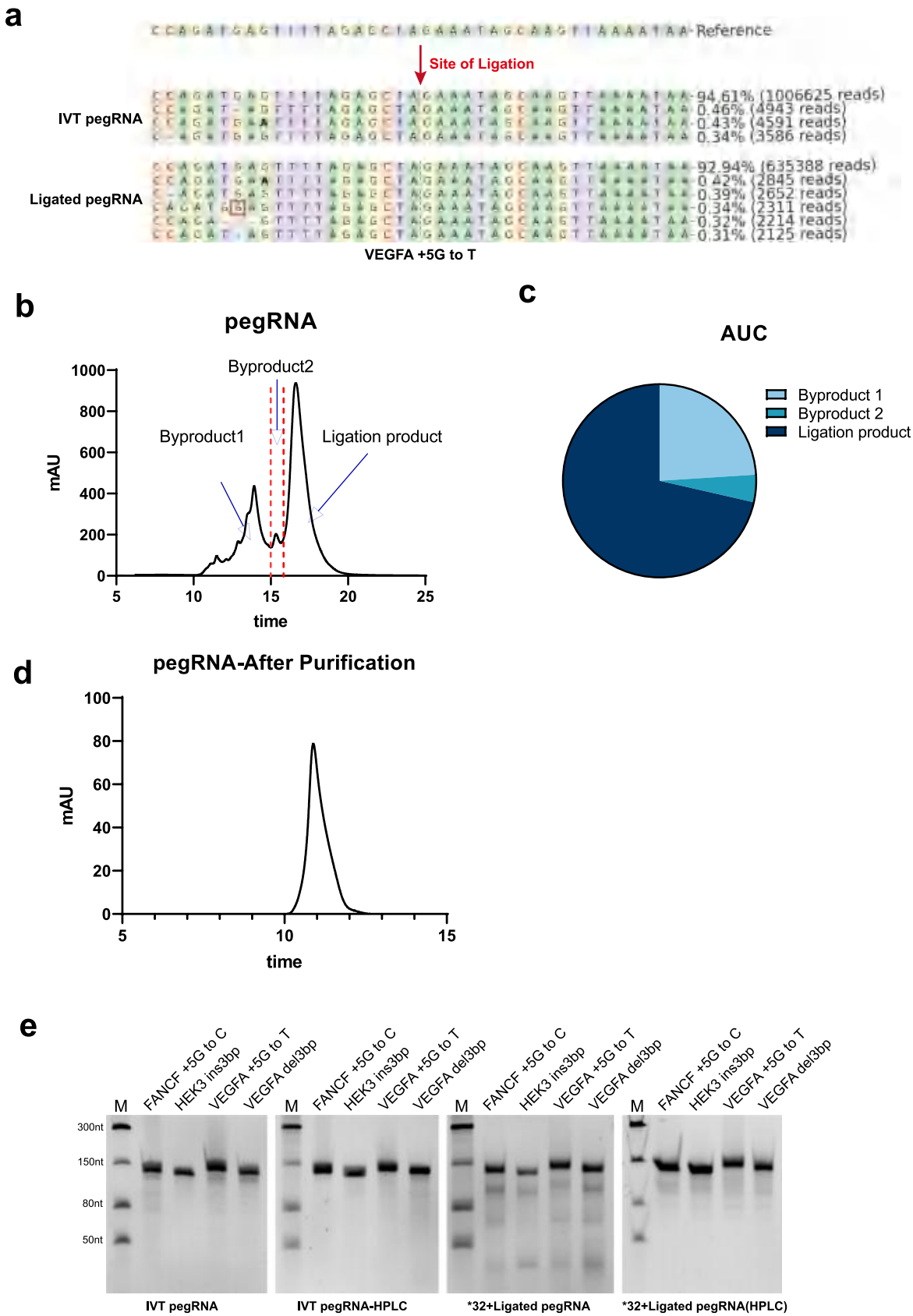


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41587-024-02394-x

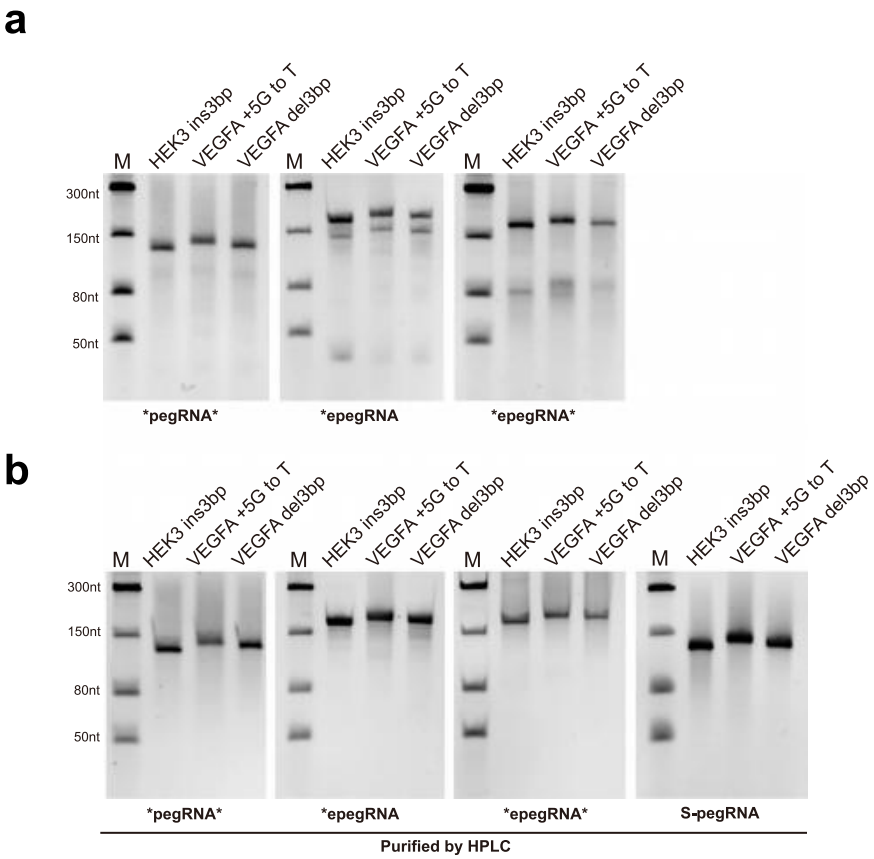
Extended Data Fig. 1 | Optimization of RNA ligation enables assembly of sgRNA. a. (Left) Overview of design scheme for sgRNA ligation. Acceptor RNA is 20 nt spacer sequence by chemical synthesis, and donor RNA is 82 nt scaffold sequence generated by IVT. (Right) Urea-PAGE of ligated sgRNA (for *VEGFA* locus). The 102nt IVT RNA depicted in the figure serves as a control band, to indicate the position of bands for successful ligation. The components labeled with '+' were present at equal concentrations across all bands. **b**. The sequence of ligated sgRNA was determined by Sanger sequencing. The ligated sgRNA was reverse transcribed, PCR amplified, and then TA cloned for Sanger sequencing. Five clones were sequenced, and the arrow indicates the ligation site. **c**. Urea-PAGE analysis of ligated sgRNA. A mixture of 100 pmol acceptor RNA, 100 pmol donor RNA, and 100 pmol splint DNA (40 nt) were annealed and ligated with 0.5 µl T4 RNA Ligase 2. The ligation reactions were performed at 25 °C or 37 °C, respectively. **d**. (Left) Overview of design scheme for sgRNA ligation using various lengths of splint DNA (20, 40, or 59 nt). (Right) Urea-PAGE analysis of ligation products. The reactions were performed at 37 °C following the conditions described above. **e**. Urea-PAGE analysis of sgRNA ligation products with different doses of splint DNA. **f**. Urea-PAGE of sgRNA ligation products with different doses of T4 RNA Ligase 2. **g**. The 20+82 ligated RNA: sgRNA was ligated using 20 nt synthetic acceptor RNA and 82 nt IVT-generated donor RNA; the *20+82 Ligated RNA: sgRNA was ligated using 20 nt synthetic acceptor RNA with 5′ modification and 82nt IVT donor RNA. **h**. In vitro cleavage of ligated sgRNA in TAE agarose gel. The molar ratio of SpCas9 protein and sgRNA was 1:1. The RNP was incubated at room temperature for 10 minutes, followed by the addition of the DNA template and incubation at 37 °C for 1 hour.

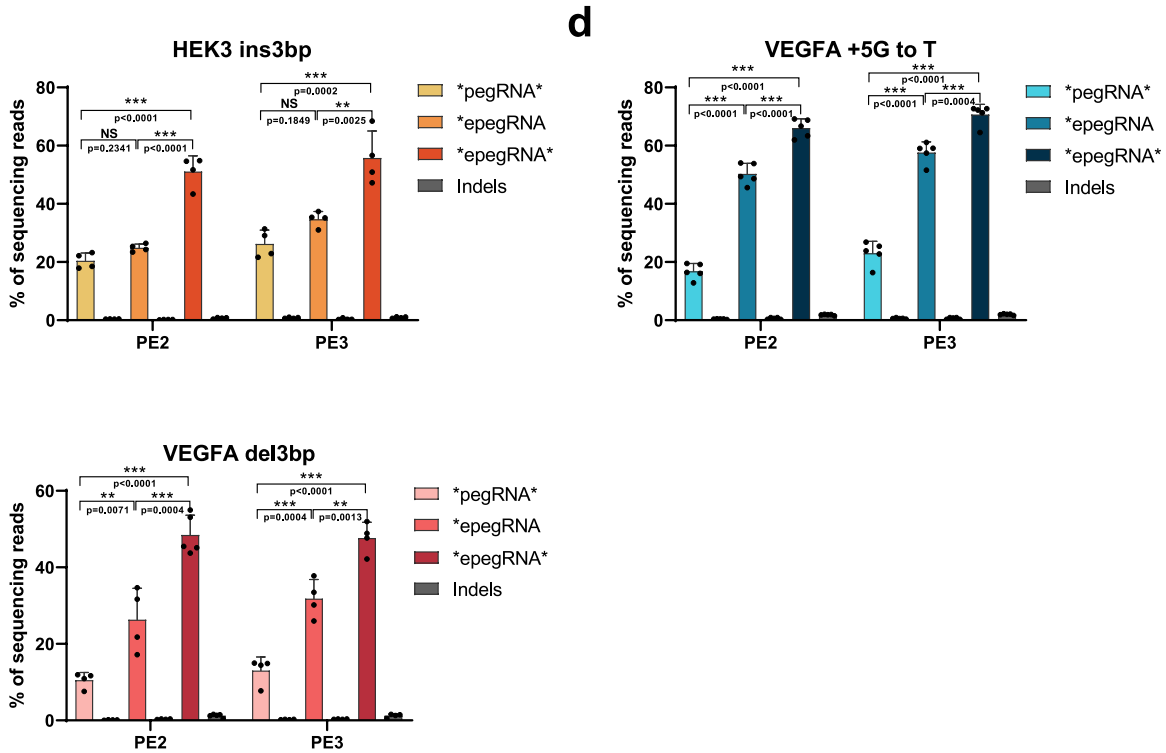
https://doi.org/10.1038/s41587-024-02394-x



Extended Data Fig. 2 | Assessment of ligation accuracy and HPLC purification of ligated pegRNA. a. RT-PCR and deep sequencing on ligated pegRNA (*32+105) and IVT pegRNA (+5 G to T Mutation in *VEGFA* locus). **b**. HPLC purification of ligated pegRNA (+5 G to T Mutation at the *VEGFA* locus). **c**. Analysis of pegRNA purity using area under curve (AUC) of each peak. **d**. Detection of purity by HPLC analysis. mAU, milli-absorbance unit; time, the execution time of the program for (**b**) and (**d**). **e**. Urea-PAGE (6%) analysis of pegRNAs. 'IVT pegRNA' indicates full length pegRNA generated by IVT; 'IVT pegRNA-HPLC' indicates full length pegRNA generated by IVT with HPLC purification; '*32+ Ligated pegRNA' indicates ligated pegRNA with 5′ end modified; '*32+Ligated pegRNA (HPLC)' indicated ligated pegRNA with 5′ end modified that was HPLC purified.

https://doi.org/10.1038/s41587-024-02394-x



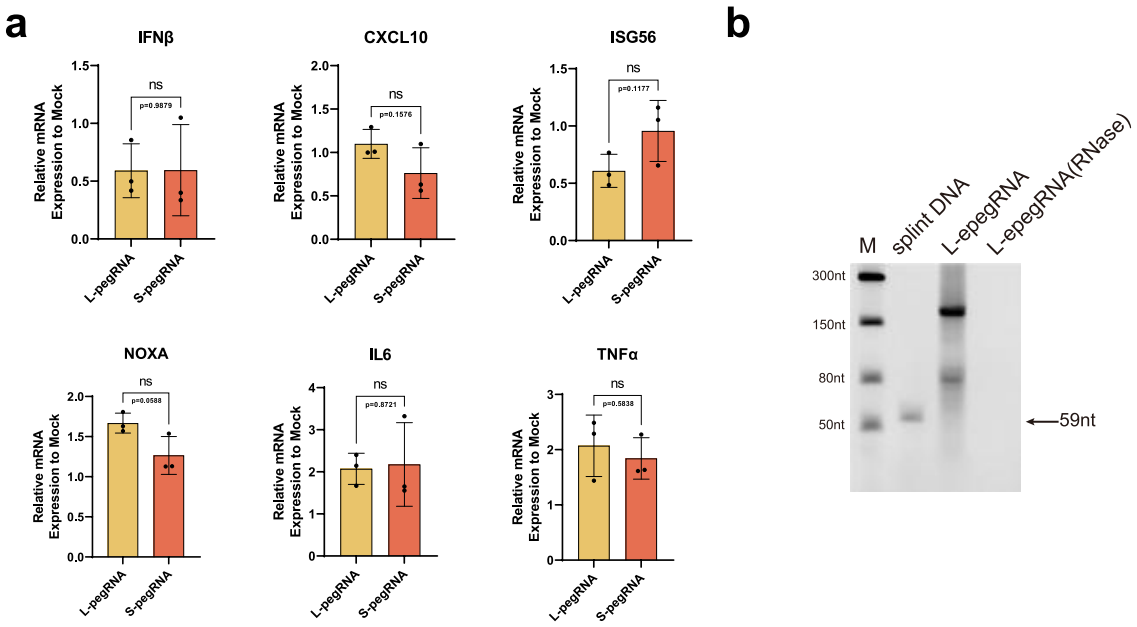


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41587-024-02394-x

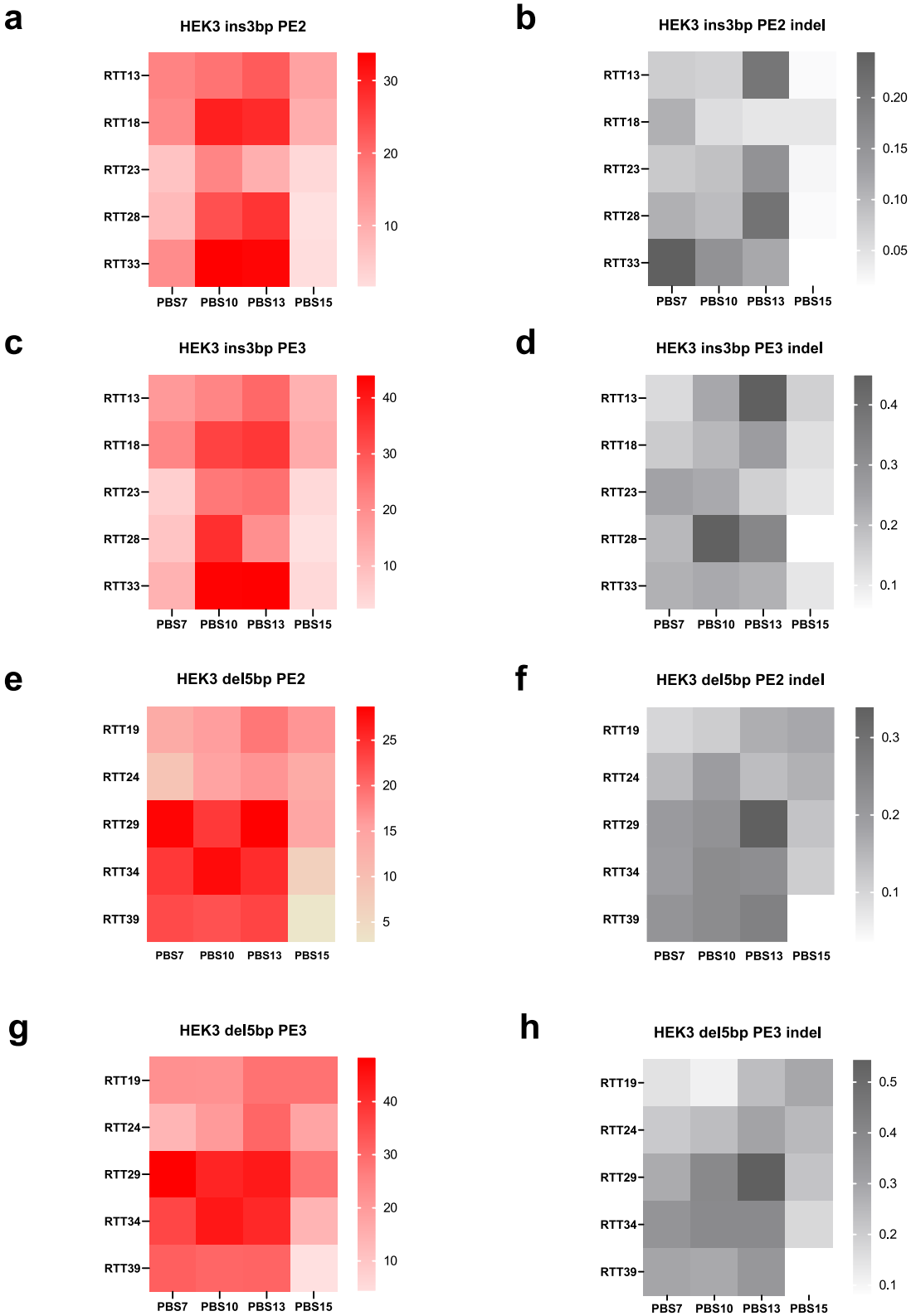
Extended Data Fig. 3 | PAGE analysis and comparison of non-HPLC purified pegRNA for editing. a. Urea-PAGE (6%) analysis of ligation products before HPLC purification. **b**. Urea-PAGE (6%) analysis of HPLC-purified ligated pegRNA and S-pegRNA. S-pegRNA is full-length pegRNA that was solid-phase synthesized with 3 nt chemical modifications at both ends. **c**–**e**. The efficiencies of RNP-mediated prime editing in HEK293T cells were determined by deep sequencing for 3 bp insertion at *HEK3* locus (**c**) (n = 4), +5 G to T mutation (**d**) (n = 5), and 3 bp deletion (**e**) at *VEGFA* locus (PE2, n = 4, 4, 5 from left to right side; PE3, n = 4). For each electroporation, 140 pmol PE protein, 186 pmol pegRNA, and 62 pmol nicking sgRNA were used. Data and error bars represent the mean and standard deviation of three or more independent biological replicates. The n values for PE2 and PE3 in figures c-e are indicated alongside each sample type. Data analysis used One-way ANOVA with Tukey's multiple comparisons test; NS, not significant; *p < 0.05; **p < 0.01; ***p < 0.001. *pegRNA*: ligated pegRNA by a 32 nt synthetic acceptor RNA with 5′ end modifications and synthetic donor RNA with 3′ end modifications. *epegRNA: ligated epegRNA by a 32 nt synthetic acceptor RNA with 5′ modifications and IVT-generated donor RNA containing evopreQ1. *epegRNA*: ligated epegRNA by a 91 nt synthetic acceptor RNA with 5′ end modifications and synthetic donor RNA with 3′ end modifications at evopreQ1 sequence.

https://doi.org/10.1038/s41587-024-02394-x



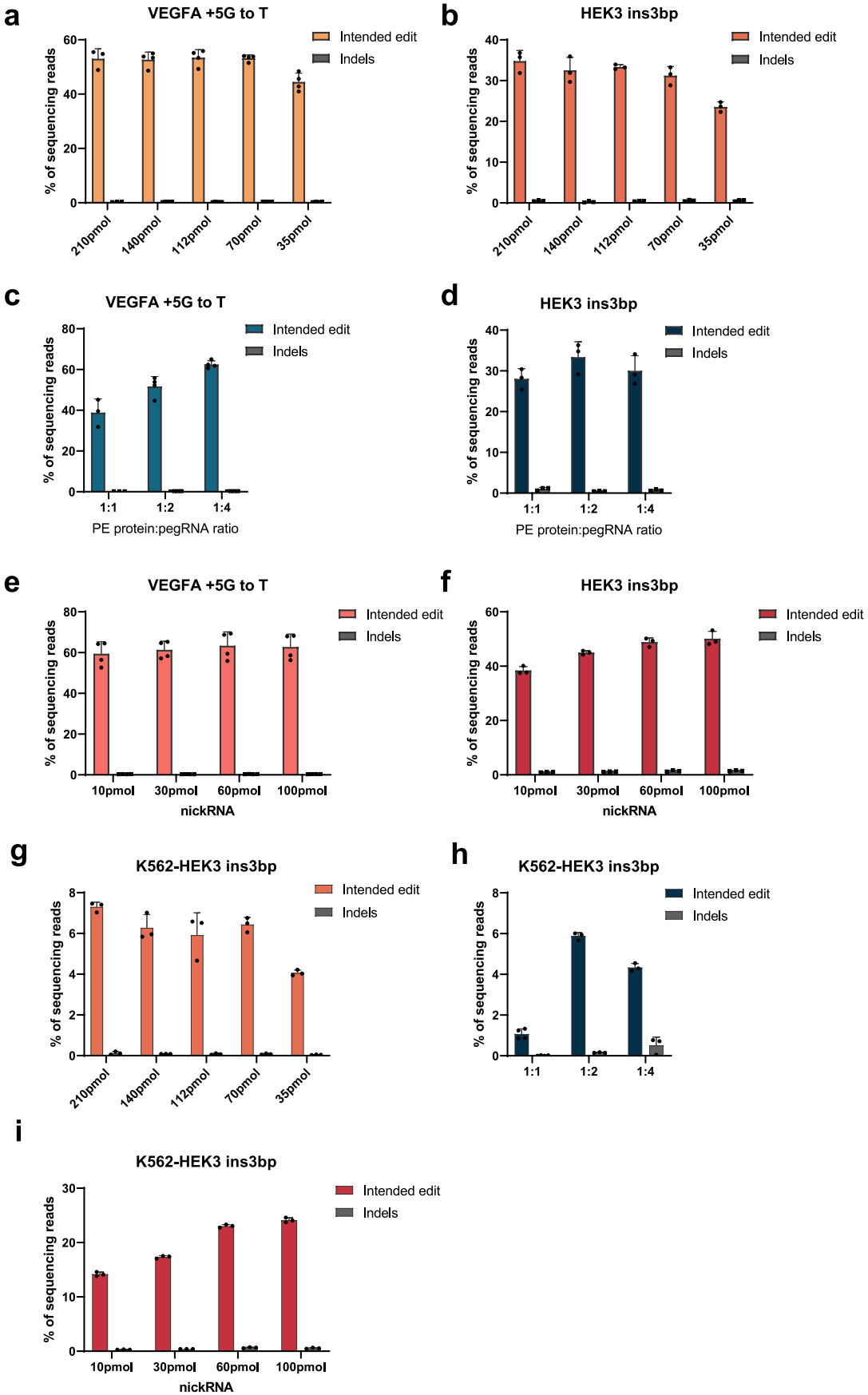
Extended Data Fig. 4 | Toxicity assessment of ligated pegRNA. a. A dose of 180 pmol of L-pegRNA (without HPLC purification) and S-pegRNA were delivered via electroporation into 5×10^5 THP1 cells. The expression of signature genes at 4 hours after electroporation were determined by RT-qPCR. Data and error bars represent the mean and standard deviation of three independent biological replicates. Data were analyzed by two-tailed unpaired Student's t-test; NS indicates no significance. **b**. Urea-PAGE (6%) analysis of L-epegRNA with or without RNase treatment. Lane 1: marker; Lane 2: 59 nt splint DNA as a reference band; Lane 3: L-epegRNA (for 1 bp insertion at *HEK3* site); Lane 4: the same L-epegRNA treated with RNase.

https://doi.org/10.1038/s41587-024-02394-x



Extended Data Fig. 5 | Effects of RTT and PBS lengths on RNP editing efficiency. a–**h**. Editing efficiencies by PE2 (**a**, **e**), PE3 (**c**, **g**) and respective indels (**b**, **d**, **f**, **h**) for *epegRNAs targeting *HEK3* site across various RTT and PBS lengths. The pegRNAs are for 3 bp insertion (**a**-**d**) and 5 bp deletion (**e**-**h**), respectively. Each electroporation used 140 pmol PE protein, 186 pmol pegRNA, and 62 pmol nicking sgRNA. Data and error bars represent the mean and standard deviation from at least two technical replicates.

https://doi.org/10.1038/s41587-024-02394-x

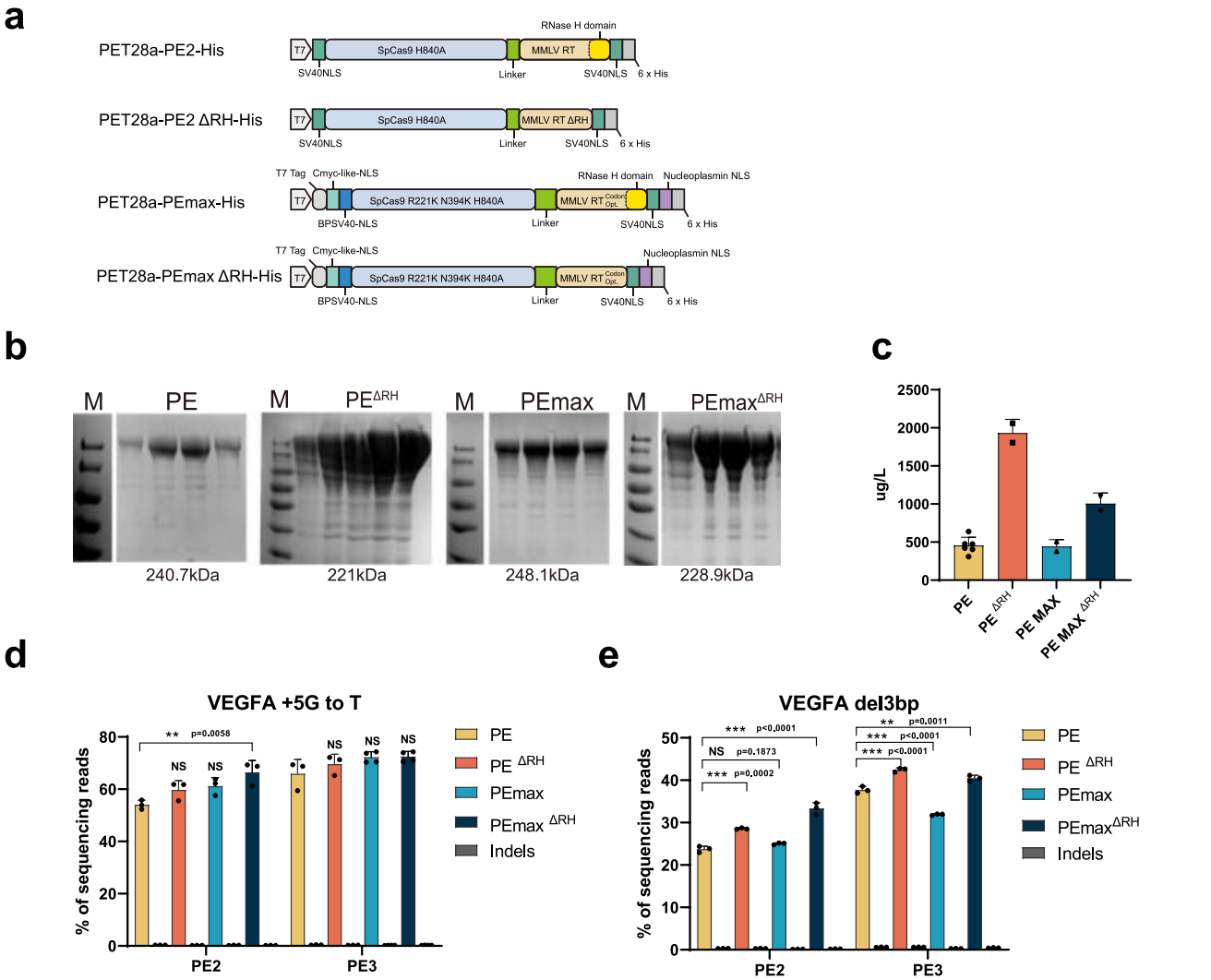


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41587-024-02394-x

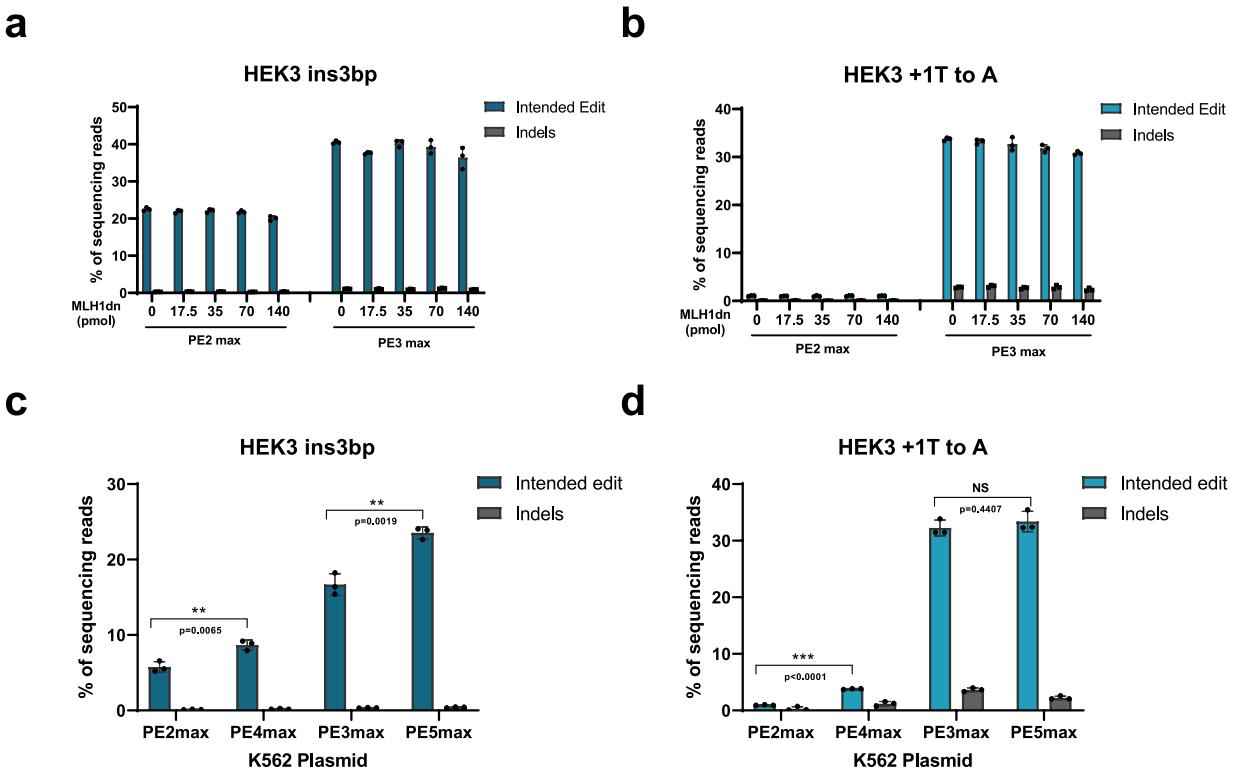
Extended Data Fig. 6 | Dose optimization of RNP delivery. a, b. Doses of pegRNA and protein were adjusted in equal proportions. The abscissa represents the protein dose. **c, d**. Optimization of PE protein and pegRNA ratios. For each sample, 70 pmol PE protein, and 70, 140, or 280 pmol pegRNA were used. **e, f**. Optimization of nickRNA dosages for PE3 system. For each sample, 70 pmol PE protein, 140 pmol pegRNA, and 10, 30, 60, or 100 pmol nickRNA were used. Dose optimizations were for editing at the *VEGFA* (**a**, **c**, **e**) and *HEK3* loci (**b**, **d**, **f**), respectively. **g**–**i**. Optimizations for editing at the *HEK3* locus in K562 cells via RNP delivery, including total dosages (**g**), PE protein to pegRNA ratios (**h**), and nicking sgRNA dosages (**i**). For **a**-**i**, *epegRNAs were used. Data and error bars represent the mean and standard deviation of three independent biological replicates. For each sample, 2×10^5 cells were used for electroporation.

https://doi.org/10.1038/s41587-024-02394-x



Extended Data Fig. 7 | Production of different PE proteins and their prime editing efficiencies via RNP. a. Illustration of the protein expression vectors. **b**. SDS-PAGE analysis of PE proteins after NI column purification: ΔRH refers to the RT enzyme of PE lacking RNase H domain. The arrow points to the target protein, M, marker. **c**. Yield of PE proteins after purification. μg/L: protein yield purified from 1 L bacterial solution. Data and error bars represent the mean and standard deviation from at least two independent biological replicates (n = 6,2,2,2 from left to right side). **d, e**. Prime editing efficiencies mediated by different PE proteins in HEK293T cells were determined by deep sequencing for +5 G to T mutation (**d**) and deletion of 3 bp (**e**) at *VEGFA* locus. For each sample, 70 pmol PE protein, 140 pmol L-epegRNA, and 60 pmol nicking sgRNA were used. Data and error bars represent the mean and standard deviation of three biological replicates. Data analysis used One-way ANOVA; NS, not significant; *p < 0.05; **p < 0.01; ***p < 0.001.

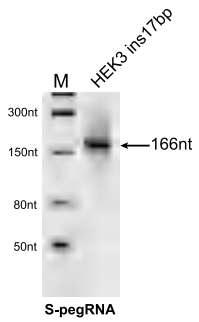
https://doi.org/10.1038/s41587-024-02394-x



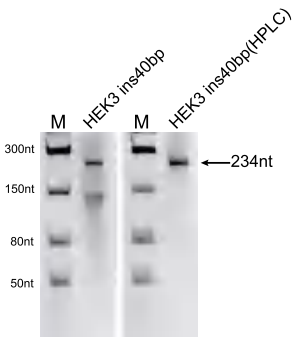
Extended Data Fig. 8 | The editing efficiencies of PE4max and PE5max via RNP in K562 cells. a, b. Prime editing efficiencies of PE4max and PE5max via RNP delivery for insertion of 3 bp (**a**) and +1 T to A mutation (**b**) at the *HEK3* locus in K562 cells. For each sample, purified MLH1dn of 0 pmol, 17.5 pmol, 35 pmol, 70 pmol and 140 pmol were used, with 70 pmol PEmax $^{\Delta RH}$ protein, 140 pmol L-epegRNA, and 60 pmol nicking sgRNA. **c**-**d**. Prime editing efficiencies of PE4max and PE5max via plasmid for insertion of 3 bp (**c**) and +1 T to A mutation (**d**) at the *HEK3* locus in K562 cells. For each sample, 800 ng PE expression plasmid, 200 ng pegRNA plasmid, and 83 ng nickRNA plasmid were used. Data and error bars represent the mean and standard deviation of three independent biological replicates. Data analysis used One-way ANOVA; NS, not significant; *p < 0.05; **p < 0.01; ***p < 0.001.

https://doi.org/10.1038/s41587-024-02394-x

a



b



c

Name	Length	Cost (CNY)	Synthesis Time (days)
L-pegRNA(HEK3 ins3bp)	128nt	4000	8
S-pegRNA(HEK3 ins3bp)		7200	17
L-pegRNA(VEGFA +5G to T)	137nt	4000	8
S-pegRNA(VEGFA +5G to T)		11000	24
L-pegRNA(VEGFA del3bp)	134nt	4000	8
S-pegRNA(VEGFA del3bp)		10500	24
L-pegRNA(HEK3 ins17bp)	166nt[a]	7400	8
S-pegRNA(HEK3 ins17bp)		11000	32

[a] This price does not guarantee successful synthesis of the RNA sequence.

d

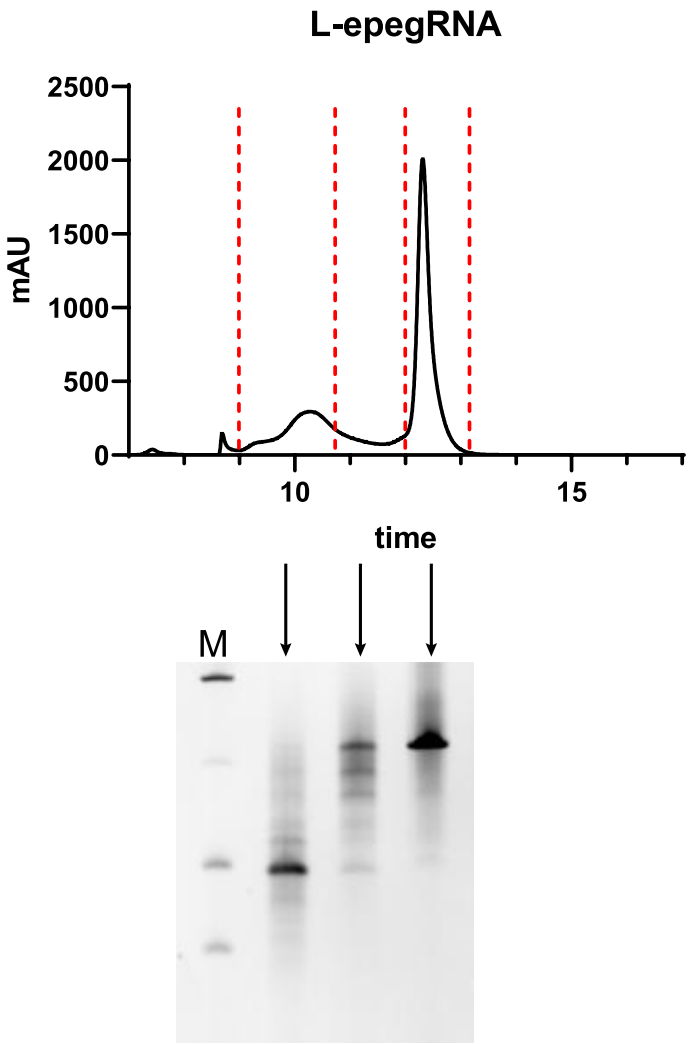
Name	Length	Total Quantity	Unpurified		Purified by HPLC	
			Cost(CNY) [a]	Synthesis Time (days)	Cost(CNY)[a]	Synthesis Time (days)
L-pegRNA	X≤140nt	5nmol(~220ug)	2000	7	4000	8
		20nmol(~900ug)	5900	7	11800	8
		50nmol(~2.2mg)	10500	8	21000	10
L-epegRNA	140nt<X≤200 nt	5nmol(~320ug)	3700	7	7400	8
		20nmol(~1.3mg)	11000	7	22000	8
		50nmol(~3.2mg)	19000	8	38000	10
L-epegRNA	X>200nt	5nmol(~400ug)	5600	7	19000	8
		20nmol(~1.6mg)	16500	7	55000	8
		50nmol(~4mg)	29500	8	99000	10

[a] This price is the list price and varies with different discounts offered by different biotech companies. The current exchange rate is 1 USD = 7.2 CNY.

For a standard 5 nmol L-epegRNA (140 nt<X≤200 nt), the total cost includes 1800 CNY for each RNA fragment, making a total of 3600 CNY per sample. The cost of splint DNA, DNase I and HPLC process is about 100 CNY per sample. We expressed and purified T4 RNA ligase 2 in a large amount. Using the commercially available T4 RNA ligase 2 would cost additional 50-300 CNY per sample.

Extended Data Fig. 9 | Comparing cost and synthesis time of pegRNAs with different production methods. a. Urea-PAGE analysis of S-pegRNA for 17 bp insertion at the *HEK3* locus. **b.** Urea-PAGE analysis of three-fragment ligation products for 40 bp insertion at the *HEK3* locus (234 bp). M, marker. **c.** Comparative analysis of cost and production period between four S-pegRNAs and their corresponding L-pegRNAs. **d.** Cost and production period for L-pegRNAs and L-epegRNAs of varied lengths across different synthesis scales, with or without HPLC purification.

https://doi.org/10.1038/s41587-024-02394-x



Extended Data Fig. 10 | HPLC purification of L-epegRNA and the corresponding urea-PAGE. An example of L-epegRNA HPLC purification and collecting corresponding fractions, followed by examining each fraction via 6% urea-PAGE after purification.

nature portfolio

Corresponding author(s): Hao Yin

Last updated by author(s): Aug 16, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	By Illumina sequencing platform (Novaseq 6000)
Data analysis	Reads which pass the QC were mapped to the PCR sequences as references, and bwa was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data that support the findings of this study are available within this manuscript. The raw data of deep-sequencing are available from the Sequence Read Archive (SRA) of NCBI under accession number PRJNA1067838

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Experiments were done in biological duplicate or triplicate to ensure reproducibility.
Data exclusions	No data were excluded.
Replication	Experiments were repeated at least two times, and the value of standard deviations is within expected range.
Randomization	The results were confirmed by various DNA locus and pegRNAs.
Blinding	Blinding was not relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Human (HEK293T, K562, U2OS, Hela, Huh-7, THP1, primary human T cells and CD34+ human HSPCs)
Authentication	HEK293T (CRL-3216, ATCC), K562 (CCL-243, ATCC),U2OS(U-2OS, HTB-96, ATCC), Hela(CCL-2, ATCC), Huh-7 (PTA-4583, ATCC) and THP1(TIB-202, ATCC) were authenticated by the supplier. Primary human T cells and CD34+ human HSPCs were purchased from Milestone® Biotechnologies.
Mycoplasma contamination	No, routine tests for Mycoplasma contamination were performed
Commonly misidentified lines (See ICLAC register)	No