nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-021-01133-w

Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing

Andrew V. Anzalone[1,2,3,4], Xin D. Gao[1,2,3,4], Christopher J. Podracky[1,2,3,4], Andrew T. Nelson[1,2,3], Luke W. Koblan[1,2,3], Aditya Raguram[1,2,3], Jonathan M. Levy[1,2,3], Jaron A. M. Mercer[1,2,3] and David R. Liu[1,2,3] ✉

The targeted deletion, replacement, integration or inversion of genomic sequences could be used to study or treat human genetic diseases, but existing methods typically require double-strand DNA breaks (DSBs) that lead to undesired consequences, including uncontrolled indel mixtures and chromosomal abnormalities. Here we describe twin prime editing (twinPE), a DSB-independent method that uses a prime editor protein and two prime editing guide RNAs (pegRNAs) for the programmable replacement or excision of DNA sequences at endogenous human genomic sites. The two pegRNAs template the synthesis of complementary DNA flaps on opposing strands of genomic DNA, which replace the endogenous DNA sequence between the prime-editor-induced nick sites. When combined with a site-specific serine recombinase, twinPE enabled targeted integration of gene-sized DNA plasmids (>5,000 bp) and targeted sequence inversions of 40 kb in human cells. TwinPE expands the capabilities of precision gene editing and might synergize with other tools for the correction or complementation of large or complex human pathogenic alleles.

Disease-associated human genetic variants range from single-base-pair substitutions to megabase duplications, deletions and rearrangements[1–3]. Gene editing approaches that can install, correct or complement these pathogenic variants in human cells have the potential to advance understanding of genetic disease and could also enable new therapeutics[4,5]. Several mammalian cell gene editing approaches based on CRISPR–Cas systems have been developed over the past decade[6], including nucleases[7–9], base editors[10,11] and prime editors[12], each with the potential to address a subset of known pathogenic sequence changes.

CRISPR–Cas nucleases, such as Cas9, can be used to disrupt genes by creating DSBs that lead to uncontrolled mixtures of indels. Additionally, paired Cas9 nuclease strategies can mediate targeted deletion of genomic DNA sequences ranging from ~50 to >100,000 base pairs in length[13]. By providing a linear donor DNA sequence, targeted insertion of new DNA sequences can be performed at single cut sites or between paired cut sites through end-joining or homology-directed repair (HDR) processes[14,15]. Although versatile, single-nuclease and paired-nuclease editing approaches have considerable drawbacks. DNA donor knock-in is accompanied by efficient indel byproducts[16], as HDR is generally inefficient compared to end-joining processes in most cell types[17,18]. The use of paired nucleases for targeted deletion generates multiple byproducts[13,19], and the precise location of the deletions is restricted by PAM availability. Moreover, DSBs at on-target or off-target sites can promote large deletions[20–22], chromosomal abnormalities[23,24] and chromothripsis[25]. The tendency of DSBs to generate complex mixtures of undesired byproducts and chromosomal changes[26–28] poses considerable challenges when applying nuclease-based editing for the manipulation of larger DNA sequences, especially in therapeutic settings.

Base editors can precisely install single-nucleotide mutations without requiring DSBs[6,10,11]. Prime editors can precisely install any of the 12 possible base substitutions as well as small insertions and deletions without requiring DSBs[12]. These DSB-free editing technologies cannot currently be used to directly install edits that alter thousands of base pairs. Prime editing has been shown to be capable of making precise insertions of up to ~40 bp and deletions of up to ~80 bp in human cells with high ratios of desired edits to byproducts[12]. Although prime editing offers the flexibility to replace one DNA sequence with another, PE2 and PE3 systems have not yet been able to mediate insertions or deletions the size of typical exons or gene coding sequences. These large DNA changes would presumably require long pegRNA reverse transcription templates and long-range DNA polymerization, eroding efficiency. In contrast, site-specific recombinases (SSRs) can excise, invert and integrate large DNA sequences in mammalian cells[29]. The longstanding challenge of reprogramming SSRs[30–32], however, has limited their use for precision gene editing applications.

Here we report the development of twinPE, which enables the deletion, substitution or insertion of larger DNA sequences at targeted endogenous genomic sites with high efficiencies in human cells. We applied twinPE to efficiently insert recombinase recognition sites at different genomic loci in a programmable manner, allowing for targeted integration and inversion of gene-sized DNA segments.

Results

TwinPE strategy. Prime editing uses a prime editor protein containing a catalytically impaired Cas9 nickase and a wild-type (PE1) or engineered (PE2) reverse transcriptase (RT) enzyme and a

[1]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [3]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA. [4]These authors contributed equally: Andrew V. Anzalone, Xin D. Gao, Christopher J. Podracky. ✉e-mail: drliu@fas.harvard.edu

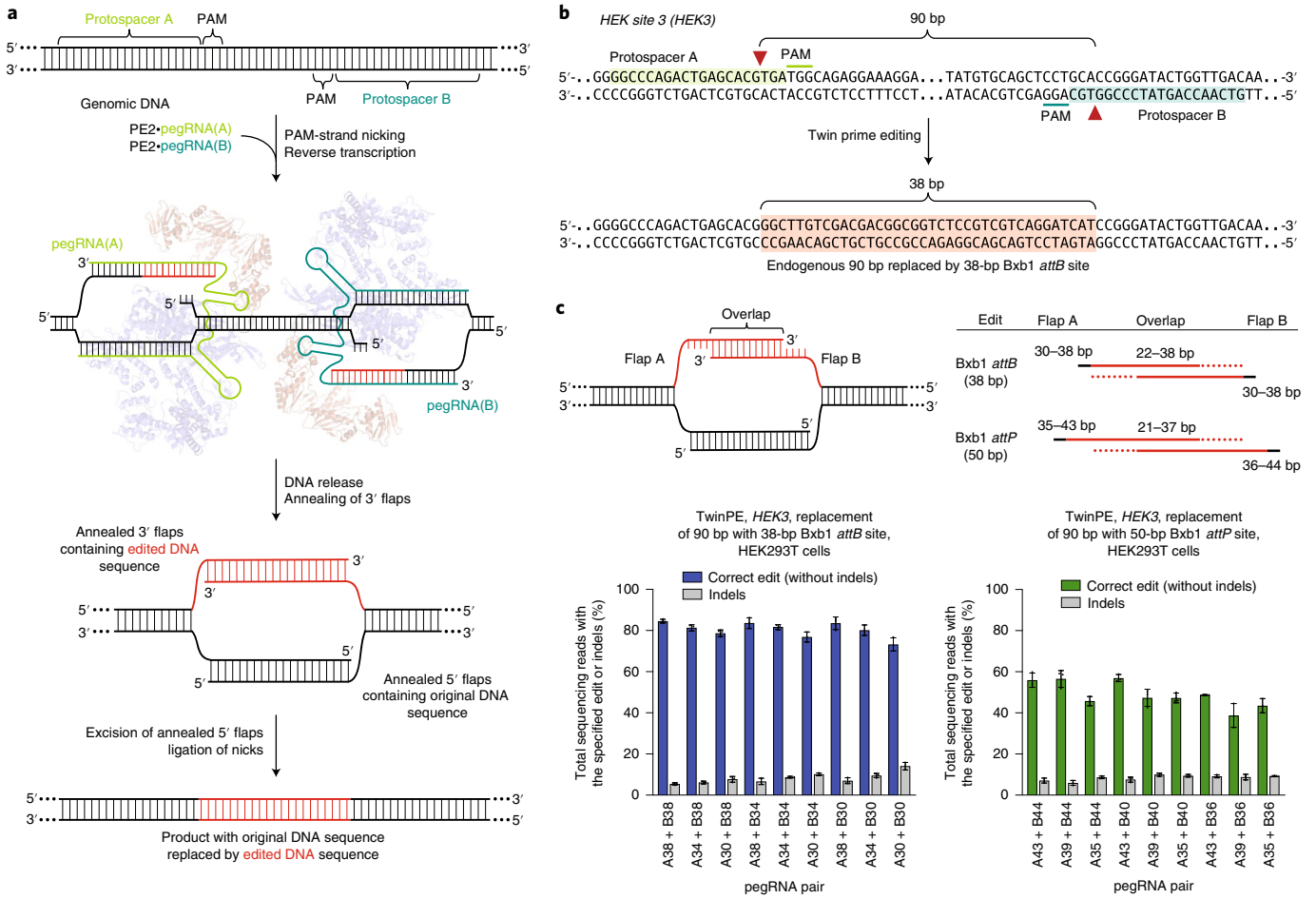


Fig. 1 | Overview of twinPE and twinPE-mediated sequence replacement. a, TwinPE systems target genomic DNA sequences that contain two protospacer sequences on opposite strands of DNA. PE2–pegRNA complexes target each protospacer, generate a single-stranded nick and reverse transcribe the pegRNA-encoded template containing the desired insertion sequence. After synthesis and release of the 3′ DNA flaps, a hypothetical intermediate exists possessing annealed 3′ flaps containing the edited DNA sequence and annealed 5′ flaps containing the original DNA sequence. Excision of the original DNA sequence contained in the 5′ flaps, followed by ligation of the 3′ flaps to the corresponding excision sites, generates the desired edited product. **b**, Example of twinPE-mediated replacement of a 90-bp sequence in *HEK3* with a 38-bp Bxb1 *attB* sequence. Red arrows indicate the position of pegRNA-induced nicks. **c**, Evaluation of twinPE in HEK293T cells for the installation of the 38-bp Bxb1 *attB* site as shown in **b** or the 50-bp Bxb1 *attP* site at *HEK3* using pegRNAs that template varying lengths of the insertion sequence. pegRNA names indicate spacer (A or B) and length of RT template. Values and error bars reflect the mean and s.d. of three independent biological replicates.

pegRNA that both specifies the target genomic site and encodes the desired edit[12]. Upon target site recognition, PE–pegRNA complexes nick the PAM-containing DNA strand and reverse transcribe the pegRNA's RT template into genomic DNA using the nicked strand as a primer. After reverse transcription, the newly synthesized 3′ flap containing the edited sequence invades the adjacent DNA to replace the redundant 5′ flap sequence. The opposing non-edited strand is then repaired using the edited DNA strand as a template. This proposed prime editing pathway presents at least two opportunities for cellular DNA repair to reject the desired edit and revert the DNA sequence to its original form: during 3′ flap annealing and ligation and during heteroduplex resolution.

We hypothesized that bypassing these steps in DNA repair might allow prime editing to occur with increased efficiency and enable new classes of targeted gene edits in mammalian cells. We envisioned a twinPE strategy that uses a pair of pegRNAs, each targeting a different DNA strand, that each template the synthesis of a 3′ flap that is complementary to the 3′ flap templated by the other pegRNA (Fig. 1a). We hypothesized that, if the newly synthesized DNA strands were highly dissimilar to the endogenous target site, the complementary 3′ flaps would preferentially hybridize with each other to create an intermediate containing annealed 3′ overhangs of new DNA sequence and annealed 5′ overhangs of original DNA sequence (Fig. 1a). As both edited strands are synthesized by prime editor complexes, there is no requirement for strand invasion of the target site or for the edit to be copied to the complementary DNA strand. Excision of the original DNA sequence (annealed 5′ overhangs) and ligation of the pair of nicks would result in the replacement of the endogenous sequence with the paired 3′ flap sequences (Fig. 1a). Owing to the flexibility in template design, the edit could, in principle, insert a new DNA sequence or delete or replace the original DNA sequence.

TwinPE-mediated large sequence replacement and deletion. To test the twinPE strategy, we initially targeted the HEK293T site 3 locus (hereafter referred to as *HEK3*) in HEK293T cells to replace 90 bp of endogenous sequence with a 38-bp Bxb1 integrase *attB* substrate sequence[33] (Fig. 1b). For each protospacer, three pegRNAs were designed with RT templates that contained 30, 34 or 38 nt of the 38-bp *attB* sequence (Fig. 1c). Pairwise combinations of these

pegRNAs are predicted to generate 3′ flaps with overlapping complementarity ranging from 22 bp to 38 bp (Fig. 1c). When both pegRNAs were transfected into HEK293T cells along with PE2, we observed highly efficient *attB* site insertion, with some combinations of pegRNAs yielding more than 80% conversion to the desired product (Fig. 1c). A similar strategy for the replacement of the 90-bp endogenous sequence with the 50-bp Bxb1 *attP* attachment sequence achieved editing efficiencies up to 58% (Fig. 1c). Notably, it was not necessary for either pegRNA to encode the full insertion sequence, because partially overlapping complementary flaps enabled full-length *attB* or *attP* sequence incorporation, although 3′ flaps with greater overlap led to slightly higher editing efficiencies and fewer indels (Fig. 1c).

Independent of our efforts, Choi et al.[34] and Lin et al.[35] reported elegant dual-pegRNA prime editing systems that can be used to perform precise large deletions in human cells or improve base substitution and small insertion or deletion edits in plant cells, respectively. For each of these strategies, the templated 3′ flap sequence is homologous to the target site DNA sequence to facilitate DNA repair for incorporation of the edit (Fig. 2a). In contrast, twinPE was designed to bypass the need for any homologous DNA sequence in the pegRNA RT template, offering more template sequence flexibility and the potential to make larger insertions with shorter RT templates. To test if twinPE could support the insertion of DNA sequences larger than the ≤44-bp insertions previously demonstrated[12] using PE2 or PE3, we compared the ability of twinPE and PE3 to generate a 108-bp *FKBP12* cDNA fragment insertion. Targeting the *HEK3* locus with PE3 in HEK293T cells, we achieved moderate editing efficiencies for the shorter 12-bp and 36-bp insertions (32% and 17%, respectively) but inefficient 108-bp insertion (0.80%) (Fig. 2b). In contrast, twinPE enabled 16% insertion efficiency for the 108-bp fragment with concomitant deletion of 90 bp of *HEK3* sequence, a 20-fold improvement. TwinPE was two- to fourfold more efficient than PE3 for inserting the 108-bp *FKBP12* cDNA fragment at *CCR5* and inserted 113-bp and 103-bp sequences containing pairs of Bxb1 recombinase sites (*attB*–27 bp spacer–*attP* or *attB*–27 bp spacer–*attB*) at this locus with ~10% efficiency (Extended Data Fig. 1). Collectively, these results demonstrate the ability of twinPE to mediate larger insertions than have been previously demonstrated with PE2 and PE3 systems.

One potential application of the twinPE strategy is the replacement of mutated exonic coding sequence with DNA that encodes wild-type sequence. Such an approach has the potential to correct any combination of mutations between twinPE-induced nick sites and raises the possibility of using a single pegRNA set to correct multiple pathogenic mutations within a stretch of DNA. To test this approach, we targeted *PAH*, which encodes phenylalanine hydroxylase. Mutations within *PAH* cause the genetic metabolic disorder phenylketonuria (PKU)[36]. We tested the ability of twinPE to recode portions of *PAH* exon 4 and exon 7, which commonly harbor mutations in patients with PKU, in HEK293T cells. By testing different flap overlap lengths using engineered pegRNAs (epegRNAs) containing a 3′ evoPreQ1 motif[37], we achieved the desired sequence recoding with efficiencies of 23% for a 46-bp recoding in exon 7, 27% for a 64-bp recoding in exon 7 and 9.4% for a 64-bp recoding in exon 4 (Fig. 2c and Extended Data Fig. 2). Additional exons in *PAH* could also be recoded, albeit with lower efficiency (Extended Data Fig. 2). These results demonstrate that twinPE can replace stretches of dozens of nucleotides in human cells with a single pair of pegRNAs.

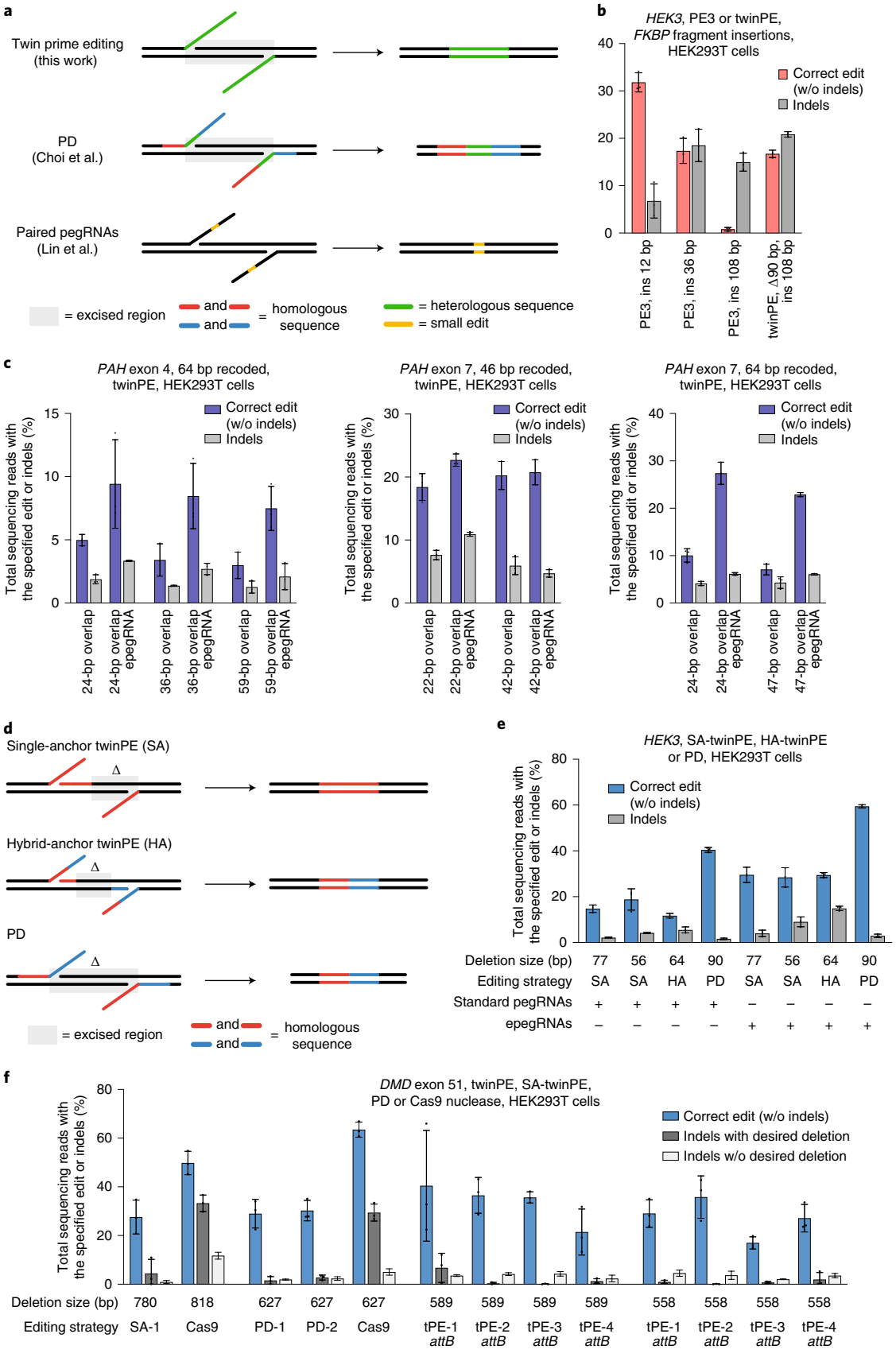
In addition to insertion and replacement of DNA sequences, twinPE might also mediate precise deletions more effectively and with greater flexibility than previously described methods. We compared three strategies using paired pegRNAs: a 'single-anchor' (SA)-twinPE strategy, which fixes the deletion junction at one of the two nick sites; a 'hybrid-anchor' (HA)-twinPE strategy, which allows flexible deletion junctions between the nick sites; and the PrimeDel (PD) strategy, recently reported by Choi et al.[34], which generates deletions between the nick sites with the option of inserting additional DNA sequence (Fig. 2d). These strategies differ in the relationship of the sequences of the two flaps and the positioning of the deletion with respect to the nick sites (Fig. 2d).

Targeting the *HEK3* site in HEK293T cells using the SA strategy, 13-nt complementary flaps were used to delete 77 bp adjacent to one of the pegRNA-induced nick sites with 15% efficiency and 2.1% indels (Fig. 2e; SA-Δ77), and 34-nt complementary flaps were used to precisely excise 56 bp of sequence with 19% efficiency and 4.2% indels (Fig. 2e; SA-Δ56). Using the HA strategy, we deleted 64 bp between the pegRNA-induced nick sites such that the product retains 13 bp of sequence 3′ of each nick with 12% efficiency and 5.5% indels (Fig. 2e; HA-Δ64). Finally, we tested the PD strategy for the deletion of 90 bp between the pegRNA-induced nick sites, which occurred with 40% efficiency and 1.6% indels (Fig. 2e; PD-Δ90). The higher efficiency of the PD strategy might arise from its ability to disrupt the PAM sequences on both strands, which likely increases deletion efficiency by limiting target re-engagement that can lead to indels or re-nicking of the prime-edited strand. Editing efficiencies were improved for all three approaches by 1.5-fold to 2.5-fold upon the addition of the evoPreQ1 motif to the 3′ end of the resulting epegRNAs[37] (Fig. 2e). Together, these data show that twinPE offers a strategy for performing targeted deletions with high flexibility and high product purity that does not rely on the availability of nuclease cut sites.

We applied twinPE-mediated deletions to target *DMD*. Pathogenic *DMD* alleles, which cause Duchenne muscular dystrophy, commonly contain large deletions in exonic regions that

Fig. 2 | Targeted sequence insertion, deletion and recoding with twinPE in human cells. a, Schematic diagram illustrating designs and edits generated for twinPE, PD[34] and paired pegRNAs[35]. Shaded gray boxes indicate regions where DNA is excised; green lines indicate the incorporation of heterologous DNA sequence; red and blue lines indicate regions of sequence homology (red to red and blue to blue); and yellow lines indicate regions with small edits. PD can introduce but does not require introduction of heterologous sequence (green). **b**, Insertion of *FKBP* coding sequence fragments with PE3 (12 bp, 36 bp and 108 bp) or twinPE (108 bp) at *HEK3* in HEK293T cells. **c**, Recoding of sequence within exons 4 and 7 in *PAH* in HEK293T cells using twinPE. A 64-bp target sequence in exon 4 was edited using 24 bp, 36 bp or 59 bp of overlapping flaps; a 46-bp target sequence in exon 7 was edited using 22 bp or 42 bp of overlapping flaps; or a 64-bp sequence in exon 7 was edited using 24 bp or 47 bp of overlapping flaps. Editing activity was compared using standard pegRNAs or epegRNAs containing 3′ evoPreQ1 motifs. **d**, Schematic diagram showing three distinct dual-flap deletion strategies that were investigated for carrying out targeted deletions. The SA-twinPE strategy allows for flexible deletion starting at an arbitrary position 3′ of one nick site and ending at the other nick site. The HA-twinPE strategy allows for flexible deletion of sequence at arbitrarily chosen positions between the two nick sites. The PD strategy of Choi et al.[34] tested here allows for deletion of the sequence starting at one nick site and ending at another nick site. Shaded gray boxes indicate regions where DNA is excised; red and blue lines indicate regions of sequence homology (red to red and blue to blue). **e**, Deletion of sequences at *HEK3* in HEK293T cells using the SA-twinPE, HA-twinPE or PD strategies targeting the same protospacer pair. Editing activity was compared using standard pegRNAs or epegRNAs containing 3′ evoPreQ1 motifs. **f**, Deletion of exon 51 sequence at the *DMD* locus in HEK293T cells using SA-twinPE, PD, paired Cas9 nuclease or twinPE-mediated *attB* sequence replacement. A UMI protocol was applied to remove PCR bias (Supplementary Note 1). Values and error bars in **b**–**f** reflect the mean and s.d. of three independent biological replicates.

result in frame-shifted transcripts[38]. Because production of nearly full-length dystrophin protein without replacement of the deleted exons can partially rescue protein function, deletion of another exon to restore the reading frame has been pursued as a therapeutic strategy[39]. We examined three twinPE deletion strategies along with a previously reported Cas9 nuclease deletion strategy for



excising exon 51 in *DMD*[39]. Using SA-twinPE deletion approaches, we observed 28% average efficiency for deletion of a 780-bp sequence containing *DMD* exon 51 (Fig. 2f). Exon 51 of *DMD* could also be excised using alternative spacer pairs that generate a 627-bp deletion with the PD strategy (30% average efficiency) or by using twinPE to replace a 589-bp sequence with a 38-bp Bxb1 *attB* sequence (40% average efficiency) (Fig. 2f). Although the paired single guide RNA (sgRNA) Cas9 nuclease strategy for deleting 818 bp achieved higher efficiency (averaging 50% precise deletion) compared to twinPE, the deletion was also accompanied by much higher indel levels of 45% (33% desired deletion with additional indels, plus 12% indels without the desired deletion) compared to twinPE, which averaged 5.1% total indels (Fig. 2f). These experiments show that twinPE and PD can generate large deletions at therapeutically relevant loci in human cells with far fewer indel byproducts compared to paired Cas9 nuclease strategies.

To examine potential off-target editing by twinPE, we amplified and sequenced four previously characterized off-target loci for one of the *HEK3* spacer sequences after treatment by eight different sets of twinPE reagents targeting *HEK3* (SA-Δ77, SA-Δ56, HA-Δ64 and PD-Δ90 with standard pegRNAs and epegRNAs). For each of the eight twinPE treatments, no detectable off-target edits or indels were detected at any of the four off-target sites beyond background levels in untreated control samples (Supplementary Table 5). The requirement for multiple hybridization events (pegRNA spacer to genome, pegRNA PBS to nicked genomic DNA primer and synthesized 3′ DNA flap to either the genome in standard PE or the second 3′ DNA flap in twinPE) might explain the low off-target editing of prime editing compared to Cas9 nuclease, as observed by our group and others[12,40,41].

Site-specific DNA integration at safe harbor loci with twinPE and Bxb1 integrase. Although twinPE can mediate large insertions more efficiently than PE3, the upper limit of sequence insertion that we observed with twinPE was ~100 bp. To integrate larger DNA cargo, we combined twinPE with a site-specific integrase (Fig. 3a). Researchers have previously used Bxb1 integrase to integrate large DNA segments into pre-installed *attB* and *attP* sites in mammalian cells[42]. To identify sites for cargo integration, we first tested twinPE-mediated insertion of Bxb1 *attB* and *attP* sequences at established safe harbor loci in HEK293T cells. We screened 15 spacer pairs with three PBS variants per pegRNA (135 pegRNA pairs total) for insertion of the 38-bp Bxb1 *attB* sequence at *CCR5* (Extended Data Fig. 3). Optimal pegRNAs for six of these spacer pairs achieved more than 50% editing efficiency with 3.9–5.4% indel byproducts (Fig. 3b). Likewise, 32 spacer pairs (288 pegRNA pairs total) targeting *AAVS1* were screened for insertion of the 50-bp Bxb1 *attP* sequence (Extended Data Fig. 4). Optimal pegRNAs for 17 of these spacer pairs achieved more than 50% efficiency with a median of 6.6% indels (Fig. 3c). Notably, twinPE outperformed PE3 for insertion of *attB* at *CCR5* in HEK293T cells (Extended Data Fig. 5). These results demonstrate that twinPE can be used to insert recombinase target sites at safe harbor loci in human cells with high efficiency.

Next, we examined if twinPE-inserted Bxb1 attachment sequences could serve as substrates for the integration of DNA plasmids. We used twinPE to generate a clonal HEK293T cell line with a homozygous *attB* insertion at *CCR5*. Transfection of this cell line with a plasmid expressing codon-optimized Bxb1 and a 5.6-kb *attP*-containing DNA donor plasmid resulted in 12–17% knock-in of the donor plasmid at *CCR5*, as measured by droplet digital polymerse chain reaction (ddPCR) quantification of an amplicon spanning the donor–genome junction and comparison to a reference amplicon in *ACTB* (Extended Data Fig. 6). This knock-in efficiency is consistent with previously reported Bxb1-mediated plasmid integration efficiencies in mammalian cells[43].

We next explored whether the components for twinPE-mediated attachment site insertion and Bxb1-mediated recombination could achieve donor integration in a single transfection step. Our initial transfection of HEK293T cells with plasmids encoding PE2, both pegRNAs, Bxb1 and a 5.6-kb donor plasmid resulted in 1.4–6.8% knock-in efficiency as measured by ddPCR (Fig. 3d). The anticipated junction sequences containing *attL* and *attR* recombination products were confirmed by amplicon sequencing, with very high product purities (84.8–99.9%, 99.3% median) (Extended Data Fig. 6).

To optimize single-transfection targeted donor integration, we investigated the effects of genomic attachment site *(attB* versus *attP*), attachment site central dinucleotide (wild-type GT versus mutant GA)[43] and the length of 3′ flap overlapping complementarity on knock-in efficiency. We found that knock-in was more efficient when twinPE was used to insert *attB* rather than *attP*, especially when pegRNAs encoded full-length *attB* or *attP* (3.3% versus 0.5% targeted integration; Fig. 3e). We also observed that the wild-type GT central dinucleotide supported higher knock-in efficiencies than GA (Fig. 3e). Notably, attachment sites with different central dinucleotides are orthogonal to one another[43], which allowed multiplexed knock-in of an *attB*-GT donor at an *attP*-GT site in *AAVS1* and GA donors at GA attachment sites in *CCR5* (Extended Data Fig. 6). Reducing 3′ flap overlap improved single-transfection targeted plasmid integration efficiency from 3.3 ± 0.2% with 38 bp of overlap to 5.5 ± 0.6% with 20 bp of overlap when inserting *attB* and from 0.55 ± 0.1% with 50 bp of overlap to 4.1 ± 0.8% with 30 bp of overlap when inserting *attP* (Fig. 3e). Similar twinPE efficiencies were observed for *attB* insertion with long (38-bp) or short (20-bp) flap overlaps. Recombination between *attP*-containing donor DNA and pegRNA expression plasmid was observed when the pegRNAs had 38 bp of overlap but was not detected with pegRNA plasmids encoding 20 bp of flap overlap (Extended Data Fig. 6). The enhanced knock-in efficiency with partially overlapping flaps might arise from reducing or eliminating unproductive recombination between donor and pegRNA expression plasmid DNA.

We next applied plasmid integration via twinPE and Bxb1 recombination at *ALB*, which has been investigated clinically for therapeutic transgene expression in hepatocytes[44,45]. Given high albumin expression in the liver, therapeutic transgene integration at *ALB* in a small percentage of hepatocytes could produce clinically relevant protein levels for many loss-of-function diseases[44]. For example, increasing circulating factor IX levels to just 1% of normal is therapeutic for the treatment of hemophilia[45]. We devised a strategy in which twinPE targets intron 1 of *ALB* and Bxb1 integrates a donor bearing a splice acceptor, followed by the cDNA for a protein of interest. After integration, splicing of the secretion signal in *ALB* exon 1 to the therapeutic transgene enables its expression and secretion.

We screened pegRNA pairs for the twinPE-mediated insertion of *attB* within intron 1 of *ALB* (Fig. 3f) and identified a spacer combination that achieved 43% correct insertion of the *attB* sequence. Single-transfection integration of a mock donor plasmid (a promoter-less copy of EGFP and PuroR-T2A-BFP under control of the EF1α promoter; 5.6-kb total) proceeded with 1.3% efficiency in HEK293T cells (Fig. 3g). In Huh7 cells, we observed 34% correct insertion of the *attB* sequence (Fig. 3f) and achieved single-transfection targeted knock-in efficiency of 2.6%. Transfection of Huh7 cells with plasmids encoding Bxb1, PE2, pegRNAs for *attB* insertion in *ALB* intron 1 and an *attP*-containing donor bearing a splice acceptor and the cDNA for human factor IX (hFIX) exons 2–8 led to detectable levels of hFIX in conditioned media (Extended Data Fig. 7). No hFIX was detected when the pegRNAs targeted *CCR5* instead of *ALB*. Collectively, these results establish a method for the integration of gene-sized DNA sequences at targeted genomic loci in previously unmodified human cells without double-strand breaks or HDR.

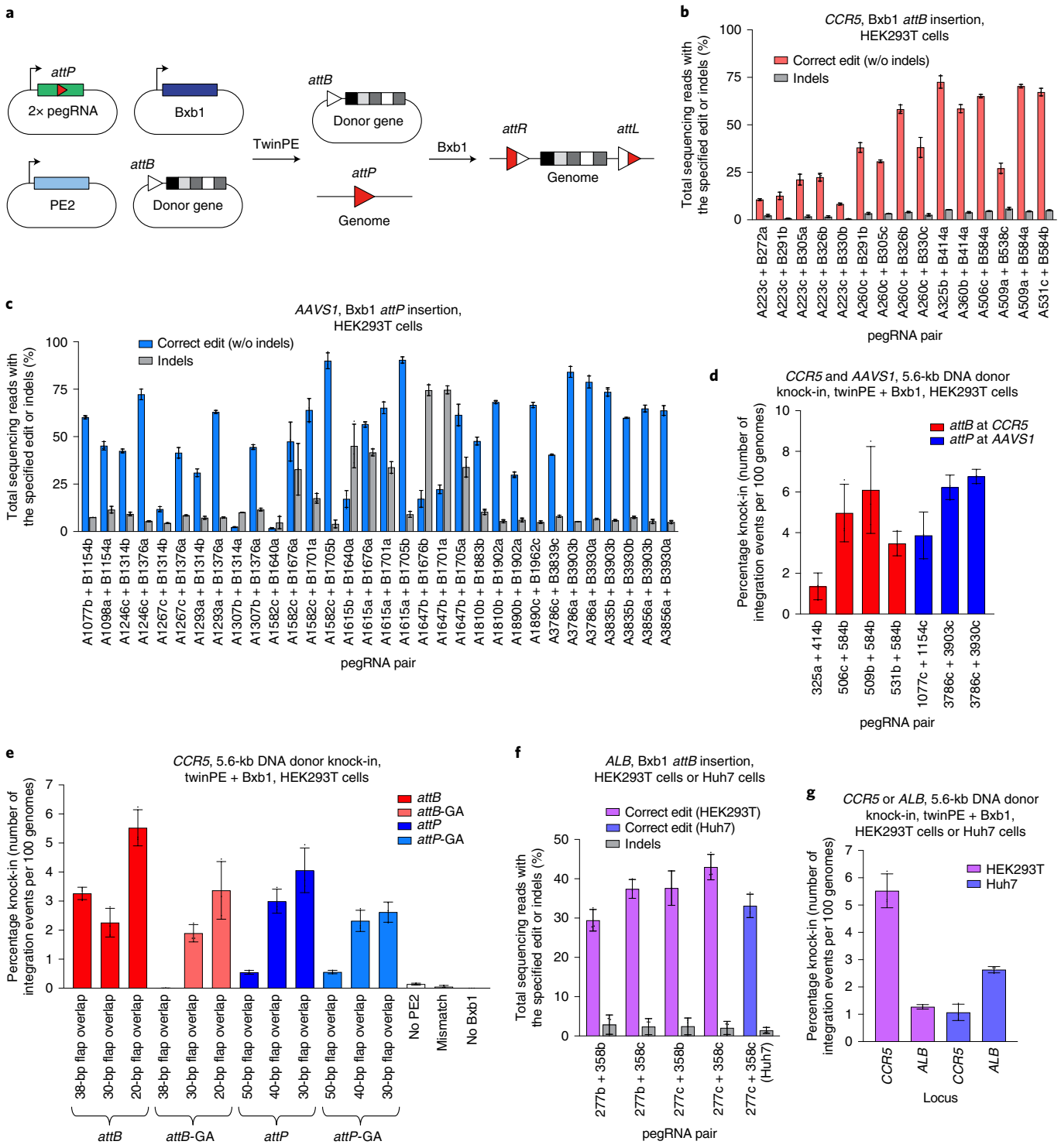


Fig. 3 | Site-specific genomic integration of DNA cargo with twinPE and Bxb1 recombinase in human cells. a, Schematic diagram of twinPE and Bxb1 recombinase-mediated site-specific genomic integration of DNA cargo. **b**, Screening of twinPE pegRNA pairs for insertion of the Bxb1 *attB* sequence at the *CCR5* locus in HEK293T cells. **c**, Screening of twinPE pegRNA pairs for installation of the Bxb1 *attP* sequence at the *AAVS1* locus in HEK293T cells. **d**, Single transfection knock-in of 5.6-kb DNA donors using twinPE pegRNA pairs targeting *CCR5* (red) or *AAVS1* (blue). The twinPE pegRNAs install *attB* at *CCR5* or *attP* at *AAVS1*. Bxb1 integrates a donor bearing the corresponding attachment site into the genomic attachment site. The number of integration events per 100 genomes is defined as the ratio of the target amplicon spanning the donor–genome junction to a reference amplicon in *ACTB*, as determined by ddPCR. **e**, Optimization of single-transfection integration at *CCR5* using the A531+B584 spacers for the twinPE pegRNA pair. Identity of the templated edit (*attB* or *attP*), identity of the central dinucleotide (wild-type GT or orthogonal mutant GA) and length of the overlap between flaps were varied to identify combinations that supported the highest integration efficiency. Percent knock-in quantified as in **d**. **f**, Pairs of pegRNAs were assessed for their ability to insert Bxb1 *attB* into the first intron of *ALB*. Protospacer sequences (277 and 358) are constant across the pegRNA pairs. The pegRNAs vary in their PBS lengths (variant B or C). The 277c/358c pair that performs best in HEK293T cells can also introduce the desired edit in Huh7 cells. **g**, Comparison of single transfection knock-in efficiencies at *CCR5* and *ALB* in HEK293T and Huh7 cell lines. Percent knock-in quantified as in **d**. Values and error bars reflect the mean and s.d. of three independent biological replicates.

To assess the possibility of undesired Bxb1-mediated donor integration or sequence modifications at off-target sites in the human genome, we selected five Bxb1 pseudo-sites for further characterization. Pseudo-sites were nominated based on the presence of a minimal Bxb1 recognition motif (ACNACNGNNNNNNCNGTNGT) identified by recombinase specificity profiling[46] and a GT central dinucleotide matching the donor plasmid used in our experiments[43]. Amplicon sequencing of all five nominated pseudo-sites in seven different samples treated with donor plasmid, twinPE reagents targeting *CCR5* or *AAVS1* and Bxb1 recombinase showed undetectable (<0.1%) or near-background levels of indels compared to untreated controls (Extended Data Fig. 7). To capture donor plasmid integration events at the nominated pseudo-sites, we attempted to amplify the expected integration junctions from treated genomic DNA samples. We observed no PCR products corresponding to off-target donor integration at any pseudo-site in any sample, whereas amplicons corresponding to on-target donor integration were readily observed (Extended Data Fig. 7). These results demonstrate that twinPE+Bxb1 can mediate precise donor integration at the intended target site without generating detectable levels of undesired donor integration or sequence alteration at predicted Bxb1 pseudo-sites in the human genome.

TwinPE and recombinase-mediated large inversions. Large structural variants are found in many human pathogenic alleles[1–3]. Given the high efficiency of recombinase attachment site insertions using twinPE, we reasoned that multiplexing the insertion of Bxb1 *attB* and *attP* could be used to correct complex genetic variants by deleting or inverting the intervening DNA sequence. We first tested whether twinPE and Bxb1 could revert an inverted H2B-EGFP coding sequence stably integrated into the HEK293T genome via lentiviral transduction (Extended Data Fig. 8). After transfection of the reporter cells with PE2, four pegRNAs and Bxb1, we observed up to 19% GFP-positive cells by flow cytometry, indicating successful inversion (Extended Data Fig. 8).

To test the twinPE recombinase inversion approach at a therapeutically relevant locus, we performed a 40-kb inversion between *IDS* and its pseudogene *IDS2*. Inversions between these sites have been observed in ~13% of patients with Hunter syndrome[47], and characterization of the breakpoints in pathogenic alleles revealed that the inversion often occurs within a recombination hotspot that is present in both *IDS* and *IDS2* (ref. [47]). We targeted regions flanking these recombination hotspots for the insertion of *attB* and *attP* sequences, which allow unidirectional inversion by Bxb1 to either generate or correct the pathogenic allele when present in opposing orientations (Fig. 4a). After screening pegRNAs, we found pegRNAs that are capable of inserting *attB* or *attP* sites at the left and right targeted sites with 70% and 74% efficiency, respectively, in HEK293T cells (Fig. 4b).

Next, we explored multiplexed twinPE insertion of *attB* and *attP* sites with Bxb1-mediated inversion of the 40-kb sequence in *IDS* and *IDS2*. We performed sequential DNA transfections with twinPE components, followed by Bxb1. An initial set of pegRNAs (set 1) was tested that installs a reverse-oriented *attP* sequence in intron 3 of *IDS2* and a forward-oriented *attB* sequence in intron 7 of *IDS*. A second set of pegRNAs (set 2) was used to install a forward-oriented *attP* sequence in *IDS2* and a reverse-oriented *attB* sequence in *IDS*. We observed 52% or 55% *attP* insertion and 28% or 31% *attB* insertion for multiplexed editing with pegRNA set 1 and set 2, respectively (Fig. 4c). When edited cells were transfected with Bxb1 recombinase, we observed a significant decrease in the percentage of amplified alleles containing *attP* and *attB* sequences compared to mock transfection controls ($P < 0.05$), suggesting that some pairs of sites had recombined (Fig. 4c). Amplicon sequencing of the anticipated inversion junctions revealed the presence of both *attL* and *attR* recombination products with product purity ≥80% at both junctions (Fig. 4d).

To circumvent unwanted recombination between pegRNA expression plasmids when combining twinPE and Bxb1-mediated inversion in a single step, we nucleofected HEK293T cells with PE2 mRNA, synthetic pegRNAs from set 1 and Bxb1 mRNA. Using amplicon sequencing, we captured the expected inverted allele junctions containing *attR* and *attL* sequences (Extended Data Fig. 9). To quantify inversion efficiency, we designed a reverse primer that binds to identical sequences in both the non-inverted and inverted alleles, allowing amplification of both alleles using the same primer pair (Extended Data Fig. 9). We observed 7.7–9.6% and 2.1–2.6% inversion efficiency for sequential and one-step methods, respectively (Fig. 4e). Collectively, these data suggest that combining twinPE with site-specific serine recombinases could be used to install or correct a common 39-kb inversion that causes Hunter syndrome and might eventually serve as a therapeutic strategy for correcting other large or complex pathogenic gene variants.

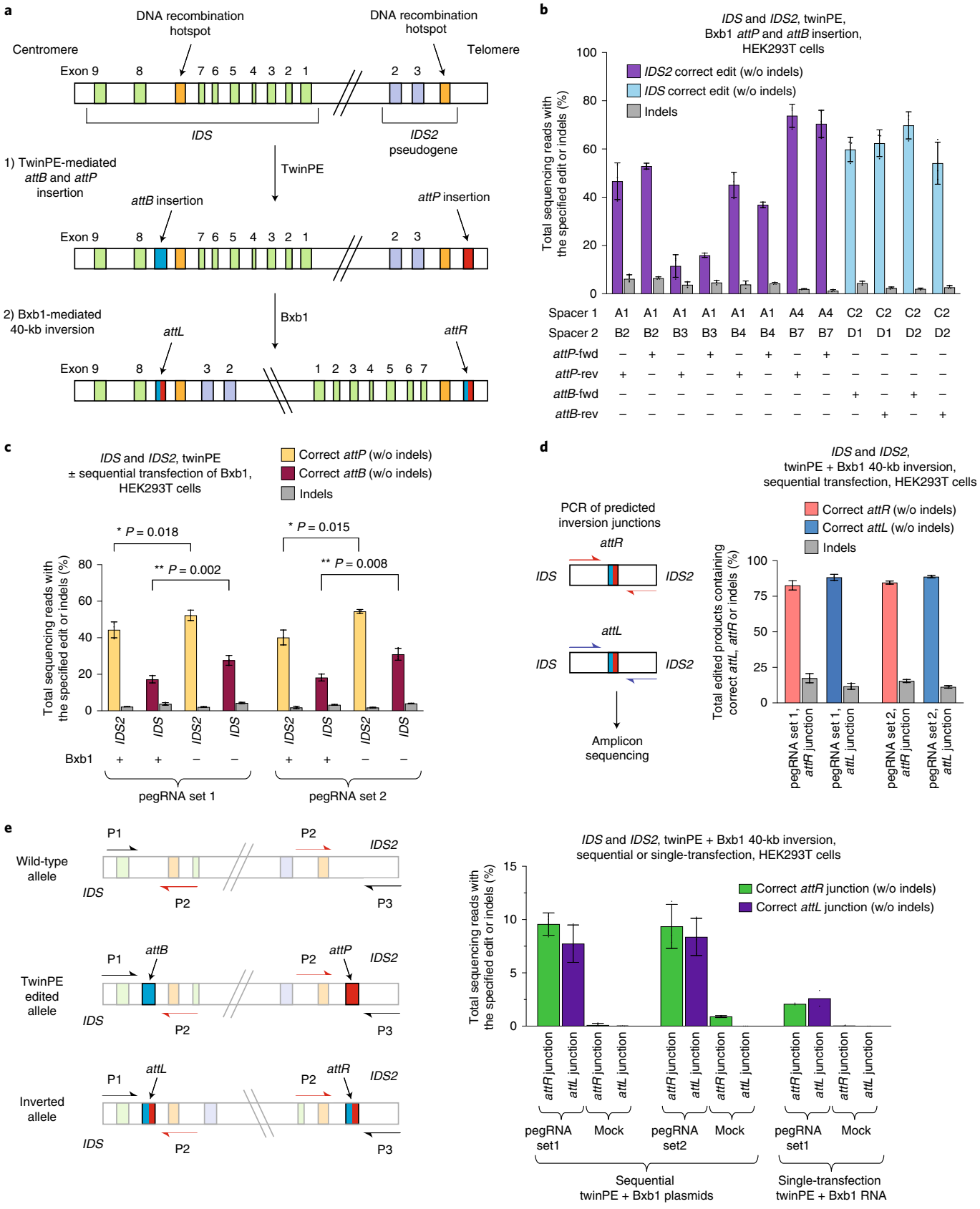
Discussion

We developed a twin prime editing approach that can be used to replace, insert or delete DNA sequences at targeted locations in the genomes of human cells. TwinPE can efficiently replace endogenous genomic DNA sequences with exogenous sequences containing Bxb1 *attB* or *attP* sites at a variety of genomic loci in HeLa, U2OS and K562 cells (Extended Data Fig. 10). An active RT is required for editing, as Cas9 nickase (H840A) and Cas9 nickase fused to an inactivated RT (K103L/R110S) both failed to generate the desired edits at all sites tested (Extended Data Fig. 10). We used twinPE to recode portions of exon sequences within *PAH*, which, in principle, could be used to correct multiple pathogenic variants using a pegRNA pair. Moreover, because twinPE does not require pegRNA RT templates to possess homology to the target site or encode the entire replacement sequence, twinPE pegRNAs can encode larger insertions for a given RT template length, including the insertion of a 108-bp sequence with 16% efficiency using twinPE, representing a 20-fold improvement over PE3. Although the design principles governing twinPE pegRNA performance are not yet fully understood (Supplementary Note 3), the use of epegRNAs[37] generally improved twinPE (Fig. 2c,e and Extended Data Fig. 9). Empirical testing of epegRNA pairs is important to optimize twinPE efficiency.

We demonstrated precise and flexible twinPE deletion strategies that make use of the programmability of 3′ flap sequences to fully specify deletion junctions, a limitation of nuclease-based

Fig. 4 | Site-specific large genomic sequence inversion with twinPE and Bxb1 recombinase in human cells. a, Schematic diagram of DNA recombination hotspots in *IDS* and *IDS2* that lead to pathogenic 39-kb inversions and the combined twinPE-Bxb1 strategy for installing or correcting the *IDS* inversion. **b**, Screen of pegRNA pairs at *IDS* and *IDS2* for insertion of *attP* or *attB* recombination sites. Values and error bars reflect the mean and s.d. of three independent biological replicates. **c**, DNA sequencing analysis of the *IDS* and *IDS2* loci after twinPE-mediated insertion of *attP* or *attB* sequences, with or without subsequent transfection with Bxb1 recombinase. *P* values were derived from a Student's two-tailed *t*-test. Values and error bars reflect the mean and s.d. of three independent biological replicates. **d**, 40,167-bp *IDS* inversion product purities at the anticipated inversion junctions after twinPE-mediated attachment site installation and sequential transfection with Bxb1 recombinase. Values and error bars reflect the mean and s.d. of three independent biological replicates. **e**, Analysis of inversion efficiency by amplicon sequencing at *IDS* and *IDS2* loci after sequential transfection or single-step transfection of twinPE editing components and Bxb1 recombinase. Values and error bars for sequential transfection reflect the mean and s.d. of three independent biological replicates. Values for single transfection reflect the mean of two independent biological replicates.

deletion strategies. Using twinPE, we tested three deletion strategies, including the previously reported PD strategy[34]. TwinPE achieved precise deletions of up to 780 nt in *DMD*, including the deletion of exon 51 with up to 28% efficiency and only 5.1% indels. Although Cas9 nuclease-mediated deletion at this site achieved 50% perfect deletion efficiency, it generated 45% indels in the process. TwinPE



might be especially useful for applications in which uncontrolled indel byproducts or other DSB consequences are problematic.

When combined with site-specific serine recombinases, twinPE can support the integration of large DNA cargo and the inversion of large DNA sequences. TwinPE and Bxb1 in a single transfection successfully inserted a 5.6-kb DNA donor plasmid into the genomes of human cells at *AAVS1*, *CCR5* and *ALB* with 6.8%, 6.1% and 2.6% efficiency, respectively. This RNA-programmable method for targeted gene-sized insertions has several advantages over other methods that use zinc-finger recombinases[32], DSBs[14,15] and HDR[48] (Supplemental Note 4). Given that twinPE insertion of *attB* and *attP* sequences can already exceed 80%, improving recombinase efficiency represents a promising strategy for enhancing targeted knock-in efficiencies.

We used multiplex twinPE insertion of *attP* and *attB* and Bxb1 to perform a 40-kb sequence inversion at *IDS* and *IDS2* that is associated with Hunter syndrome with 9.6% and 2.6% efficiency via sequential transfection and one-step RNA nucleofection, respectively (Fig. 4e). Although nuclease approaches have achieved targeted sequence inversions previously[49,50], DSB-induced repair pathways generate undesired products and can lead to de novo structural variants, including deletion of the targeted DNA sequence[20–22] and chromosomal abnormalities[23,24]. TwinPE and Bxb1 achieve similar efficiency while circumventing the uncontrolled nature of DSB repair.

In summary, twinPE expands the capabilities of prime editing to include targeted deletion, replacement and, when combined with site-specific recombinases, gene-sized integration and inversion without requiring double-strand breaks. These new capabilities might enable strategies to study and treat genetic diseases arising from loss-of-function or complex structural mutations.

Online content

Any methods, additional references, Nature Research 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/s41587-021-01133-w.

Received: 1 July 2021; Accepted: 22 October 2021;
Published online: 9 December 2021

References

1. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
2. Landrum, M. J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
3. Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* **14**, 125–138 (2013).
4. Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).
5. Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).
6. 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).
7. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
8. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
9. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
10. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
11. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
12. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
13. Canver, M. C. et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289**, 21312–21324 (2014).
14. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
15. Wang, B. et al. Highly efficient CRISPR/HDR-mediated knock-in for mouse embryonic stem cells and zygotes. *Biotechniques* **59**, 201–202, 204, 206–208 (2015).
16. Pawelczak, K. S., Gavande, N. S., VanderVere-Carozza, P. S. & Turchi, J. J. Modulating DNA repair pathways to improve precision genome engineering. *ACS Chem. Biol.* **13**, 389–396 (2018).
17. Branzei, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nat. Rev. Mol. Cell Biol.* **9**, 297–308 (2008).
18. Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010).
19. Gasperini, M. et al. CRISPR/Cas9-mediated scanning for regulatory elements required for *HPRT1* expression via thousands of large, programmed genomic deletions. *Am. J. Hum. Genet.* **101**, 192–205 (2017).
20. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
21. Alanis-Lobato, G. et al. Frequent loss of heterozygosity in CRISPR–Cas9-edited early human embryos. *Proc. Natl Acad. Sci. USA* **118**, e2004832117 (2021).
22. Song, Y. et al. Large-fragment deletions induced by Cas9 cleavage while not in the BEs system. *Mol. Ther. Nucleic Acids* **21**, 523–526 (2020).
23. Brunet, E. & Jasin, M. Induction of chromosomal translocations with CRISPR–Cas9 and other nucleases: understanding the repair mechanisms that give rise to translocations. *Adv. Exp. Med. Biol.* **1044**, 15–25 (2018).
24. Nahmad, A. D. et al. Frequent aneuploidy in primary human T cells following CRISPR–Cas9 cleavage. Preprint at https://www.biorxiv.org/content/10.1101/2021.08.20.457092v1.abstract (2021).
25. Leibowitz, M. L. et al. Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021).
26. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
27. Ihry, R. J. et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
28. Enache, O. M. et al. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat. Genet.* **52**, 662–668 (2020).
29. Merrick, C. A., Zhao, J. & Rosser, S. J. Serine integrases: advancing synthetic biology. *ACS Synth. Biol.* **7**, 299–310 (2018).
30. Karpinski, J. et al. Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat. Biotechnol.* **34**, 401–409 (2016).
31. Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H. & Liu, D. R. A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Res.* **44**, 9758–9770 (2016).
32. Gaj, T. et al. Enhancing the specificity of recombinase-mediated genome engineering through dimer interface redesign. *J. Am. Chem. Soc.* **136**, 5047–5056 (2014).
33. Kim, A. I. et al. Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis groEL1* gene. *Mol. Microbiol.* **50**, 463–473 (2003).
34. Choi, J. et al. Precise genomic deletions using paired prime editing. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020 (2021).
35. Lin, Q. et al. High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-00868-w (2021).
36. Scriver, C. R. The *PAH* gene, phenylketonuria, and a paradigm shift. *Hum. Mutat.* **28**, 831–845 (2007).
37. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-01039-7 (2021).
38. Flanigan, K. M. et al. Mutational spectrum of *DMD* mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. *Hum. Mutat.* **30**, 1657–1666 (2009).
39. Aartsma-Rus, A. et al. Development of exon skipping therapies for duchenne muscular dystrophy: a critical review and a perspective on the outstanding issues. *Nucleic Acid Ther.* **27**, 251–259 (2017).
40. Kim, D. Y., Moon, S. B., Ko, J.-H., Kim, Y.-S. & Kim, D. Unbiased investigation of specificities of prime editing systems in human cells. *Nucleic Acids Res.* **48**, 10576–10589 (2020).
41. Jin, S. et al. Genome-wide specificity of prime editors in plants. *Nat. Biotechnol.* **39**, 1292–1299 (2021).
42. Duportet, X. et al. A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* **42**, 13440–13451 (2014).
43. Jusiak, B. et al. Comparison of integrases identifies Bxb1-GA mutant as the most efficient site-specific integrase system in mammalian cells. *ACS Synth. Biol.* **8**, 16–24 (2019).

44. Sharma, R. et al. In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777–1784 (2015).
45. Nathwani, A. C. et al. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N. Engl. J. Med.* **371**, 1994–2004 (2014).
46. Bessen, J. L. et al. High-resolution specificity profiling and off-target prediction for site-specific DNA recombinases. *Nat. Commun.* **10**, 1937 (2019).
47. Bondeson, M. L. et al. Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. *Hum. Mol. Genet.* **4**, 615–621 (1995).
48. Chen, X. et al. In trans paired nicking triggers seamless genome editing without double-stranded DNA cutting. *Nat. Commun.* **8**, 657 (2017).
49. Park, C. Y. et al. Targeted inversion and reversion of the blood coagulation factor 8 gene in human iPS cells using TALENs. *Proc. Natl Acad. Sci. USA* **111**, 9253–9258 (2014).
50. Li, J. et al. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J. Mol. Cell. Biol.* **7**, 284–298 (2015).

Methods

General methods. DNA amplification was conducted by PCR using Phusion U Green Multiplex PCR Master Mix (Thermo Fisher Scientific) or Q5 Hot Start High-Fidelity 2× Master Mix (New England Biolabs), unless otherwise noted. DNA oligonucleotides were obtained from Integrated DNA Technologies (IDT). Plasmids expressing sgRNAs were constructed by ligation of annealed oligonucleotides into *Bsm*BI-digested acceptor vector, as previously described[12]. Plasmids expressing pegRNAs were constructed by Gibson assembly or Golden Gate assembly, as previously described[12]. Sequences of sgRNA and pegRNA constructs used in this work are listed in Supplementary Table 1. All vectors for mammalian cell experiments were purified using Plasmid Plus Midiprep kits (Qiagen), PureYield Plasmid Miniprep kits (Promega) or QIAprep Spin Miniprep kits. Synthetic pegRNAs were ordered from IDT without high-performance liquid chromatography purification.

General mammalian cell culture conditions. HEK293T (American Type Culture Collection (ATCC) CRL-3216), U2OS (ATTC HTB-96), K562 (ATCC CCL-243) and HeLa (ATCC CCL-2) cells were purchased from ATCC and cultured and passaged in DMEM plus GlutaMAX (Thermo Fisher Scientific), McCoy's 5A Medium (Gibco), RPMI Medium 1640 plus GlutaMAX (Gibco) or EMEM (ATCC), respectively, each supplemented with 10% (v/v) FBS (Gibco, qualified) and 1× penicillin–streptomycin (Corning). Huh7 cells (a gift from Erik Sontheimer's group, originated from ATCC) were cultured and passaged in DMEM plus GlutaMAX (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS (Gibco, qualified) and 1× penicillin–streptomycin. All cell types were incubated, maintained and cultured at 37 °C with 5% CO_2. Cell lines were authenticated by their respective suppliers and tested negative for mycoplasma.

HEK293T, HeLa and Huh7 tissue culture transfection protocol and genomic DNA preparation. HEK293T cells were seeded on 48-well poly-D-lysine-coated plates (Corning). Next, 16–24 h after seeding, cells were transfected at approximately 60% confluency with 1 µl of Lipofectamine 2000 (Thermo Fisher Scientific), according to the manufacturer's protocols, and 750 ng of PE2 plasmid DNA, 125 ng of pegRNA 1 plasmid DNA and 125 ng of pegRNA 2 plasmid DNA (for twinPE transfections); 750 ng of PE2 plasmid DNA, 250 ng of pegRNA plasmid DNA and 83 ng of sgRNA plasmid DNA (for PE3 transfections); or 750 ng of Cas9 plasmid DNA and 125 ng of sgRNA 1 plasmid DNA and 125 ng of sgRNA 2 plasmid DNA (for paired Cas9 nuclease transfections). Unless otherwise stated, cells were cultured 3 d after transfection, after which the media were removed; the cells were washed with 1× PBS solution (Thermo Fisher Scientific); and genomic DNA was extracted by the addition of 150 µl of freshly prepared lysis buffer (10 mM Tris-HCl, pH 7.5; 0.05% SDS; 25 µg ml^{-1} of proteinase K (Thermo Fisher Scientific)) directly into each well of the tissue culture plate. The genomic DNA mixture was incubated at 37 °C for 1–2 h, followed by an 80 °C enzyme inactivation step for 30 min. Primers used for mammalian cell genomic DNA amplification are listed in Supplementary Table 2. For HeLa cell transfections, cells were grown and seeded in 96-well plates (Falcon, cat. no. 353075). Then, 16–24 h after seeding, cells were transfected with 0.75 µl of TransIT-HeLaMONSTER transfection reagent (Mirus) using 190 ng of PE2-P2A-Blast plasmid DNA, 31.5 ng of pegRNA 1 plasmid DNA and 31.5 ng of pegRNA 2 plasmid DNA (for twinPE transfections). Twenty-four hours after transfection, cells were treated with blasticidin to a final concentration of 10 µg ml^{-1}. Genomic DNA isolation was performed as described above for HEK293T cells using 50 µl of lysis buffer. Huh7 cells were seeded at 150,000 cells per well in poly-D-lysine-coated 24-well plates (Corning). Then, 16–24 h after seeding, cells were transfected with 2 µl of Lipofectamine 2000 (Thermo Fisher Scientific), according to the manufacturer's protocols, and up to 800 ng of plasmid DNA (same ratios as in HEK293T transections described above, scaled proportionally). Genomic DNA isolation was performed as described above for HEK293T transfections.

High-throughput DNA sequencing of genomic DNA samples. Genomic sites of interest were amplified from genomic DNA samples and sequenced on an Illumina MiSeq, as previously described, with the following modifications[10]. In brief, amplification primers containing Illumina forward and reverse adapters (Supplementary Table 2) were used for a first round of PCR (PCR 1) amplifying the genomic region of interest. Next, 25-µl PCR 1 reactions were performed with 0.5 µM of each forward and reverse primer, 1 µl of genomic DNA extract and 12.5 µl of Phusion U Green Multiplex PCR Master Mix. PCR reactions were carried out as follows: 98 °C for 2 min and then 30 cycles of 98 °C for 10 s, 61 °C for 20 s and 72 °C for 30 s, followed by a final 72 °C extension for 2 min. Unique Illumina barcoding primer pairs were added to each sample in a secondary PCR reaction (PCR 2). Specifically, 25 µl of a given PCR 2 reaction contained 0.5 µM of each unique forward and reverse Illumina barcoding primer pair, 1 µl of unpurified PCR 1 reaction mixture and 12.5 µl of Phusion U Green Multiplex PCR 2× Master Mix. The barcoding PCR 2 reactions were carried out as follows: 98 °C for 2 min and then 12 cycles of 98 °C for 10 s, 61 °C for 20 s and 72 °C for 30 s, followed by a final 72 °C extension for 2 min. PCR products were evaluated analytically by electrophoresis in a 1.5% agarose gel. PCR 2 products (pooled by common amplicons) were purified by electrophoresis with a 1.5% agarose gel using a QIAquick Gel Extraction Kit (Qiagen), eluting with 40 µl of water. DNA concentration was measured by fluorometric quantification (Qubit, Thermo Fisher Scientific) or qPCR (KAPA Library Quantification Kit-Illumina, KAPA Biosystems) and sequenced on an Illumina MiSeq instrument using Illumina MiSeq control software (version 3.1), according to the manufacturer's protocols.

Sequencing reads were demultiplexed using MiSeq Reporter (Illumina). Alignment of amplicon sequences to a reference sequence was performed using CRISPResso2 (ref. [51]). For all prime editing yield quantification, prime editing efficiency was calculated as: percentage of [no. of reads with the desired edit that do not contain indels] ÷ [no. of total reads]. For quantification of editing, CRISPResso2 was run in HDR mode using the desired allele as the expected allele (e flag) and with 'discard_indel_reads' on. Any sequence containing an indel with respect to the allele to which it aligns was counted separately and did not contribute to the correctly edited allele percentage. The percent editing was quantified as the number of non-discarded reads aligning to the anticipated edited allele (not containing indels) divided by the total number of sequencing reads (which includes those aligned to reference with or without indels and those aligned to the desired edit with or without indels). Indels were quantified as the total number of discarded reads (from either the original or edited allele alignments) divided by the total number of sequencing reads. Editing yield was then calculated as: [no. of non-discarded HDR aligned reads] ÷ [total reads]. Indel yields were calculated as: [no. of indel-containing discarded reads] ÷ [total reads].

Unique molecular identifiers (UMIs) were applied to quantify the deletion efficiency and assess PCR bias (see Supplementary Note 1 for discussion) in a three-step PCR protocol. In brief, in the first step of linear amplification, 1 µl of genomic DNA extract was linearly amplified by 0.1 µM of only the forward primer containing a 15-nt or 16-nt UMI with Phusion U Green Multiplex PCR Master Mix in a 25-µl reaction (ten cycles of 98 °C for 1 min, 61 °C for 25 s and 72 °C for 1 min). The PCR products were then purified by 1.6× AMPure beads (Beckman Coulter) and eluted in 20 µl of Qiagen elution buffer. In the second step, 1 µl or 2 µl of purified linearly amplified PCR products were then amplified for 30 cycles with 0.5 µM of each forward and reverse primer with Phusion U Green Multiplex PCR mix in a 25-µl reaction, as described above. In this case, the forward primer anneals to the P5 Illumina adaptor sequence located at the 5′ end of the UMI primer and upstream of the UMI sequence. For the *DMD* locus library preparation, the PCR products were purified by 1× AMPure beads and eluted in 25 µl of elution buffer. In the third step, the purified PCR products (1 µl) were amplified for 12 cycles as described above for adding unique Illumina barcodes and adaptors. To assess large deletions at the *DMD* locus, the top band (unedited large amplicon) and the bottom band (edited amplicons with deletions) were excised separately from a 1.5% agarose gel and loaded on two separate MiSeq runs to avoid biased clustering of amplicons. For library preparation at other loci, 1 µl of PCR product was used directly for the barcoding PCR step without 1× AMPure beads purification. For UMI-based PCR bias assessment described in Supplementary Note 1, the editing efficiency was calculated from the libraries prepared following the UMI protocols analyzed with UMI deduplication or without UMI deduplication.

Raw sequencing reads were UMI deduplicated using AmpUMI[52]. For paired-end reads, SeqKit[53] was used to concatenate (merge without overlap) R1s with the reverse complement of R2s. The concatenated R1 + R2s were UMI deduplicated using the UMI at the 5′ end of R1. UMI-deduplicated R1s or concatenated R1 + R2s were analyzed using CRISPResso2. For analyzing concatenated R1 + R2s, an appropriate concatenated reference amplicon sequence was provided to minimize sequencing alignment artifacts due to the concatenation.

Nucleofection of U2OS and K562. Nucleofection was performed in all experiments that used K562 and U2OS cells. In total, 200,000 cells were used per nucleofection. Counted cells were pelleted and washed with PBS and then resuspended in nucleofection solution following the recommendation of the Lonza SE Cell Line 4D-Nucleofector Kit. After nucleofection of the cells, the cells were allowed to incubate in the cuvette at room temperature for 10 min. After this time, the contents of the cuvette were transferred to a 48-well plate containing pre-warmed (37 °C) media. Genomic DNA was extracted and prepared for Illumina MiSeq preparation, as described above.

Single-step twinPE and Bxb1-mediated DNA donor knock-in and inversions. For single-step knock-in, HEK293T cells were transfected with 500 ng of PE2 plasmid DNA, 50 ng of each pegRNA plasmid DNA (two in total), 200 ng of codon-optimized Bxb1 plasmid DNA and 200 ng of DNA donor using Lipofectamine 2000, as described above. For multiplex knock-in, HEK293T cells were transfected with 200 ng of PE2 plasmid, 50 ng of each pegRNA plasmid DNA (four in total), 300 ng of Bxb1 plasmid DNA and 150 ng of each DNA donor (two in total). Donor plasmid sequences can be found in Supplementary Sequences 1. The codon-optimized Bxb1 gene was purchased from GenScript and can be found in Supplementary Sequence 2. It was assembled into a pCMV expression vector using standard molecular cloning techniques.

For inversion experiments, sequential plasmid transfection and single-step mRNA nucleofection were performed. In the sequential plasmid transfection experiment, HEK293T cells were transfected with 750 ng of PE2 plasmid DNA and 62.5 ng of each pegRNA plasmid DNA (four in total) using Lipofectamine 2000,

as described above. After 3 d, cells were detached and plated in 24-well plates and then serially passaged for about 7 d. Next, 48-well plates were seeded with 20,000 cells per well and transfected 16–24 h later with Lipofectamine 2000 and 500 ng of BxB1 plasmid DNA. Genomic DNA was extracted and prepared for Illumina MiSeq, as described above. For single-step mRNA nucleofection, 200,000 HEK293T cells were nucleofected with 1,000 ng of PE2 mRNA, 30 pmol each of synthetic pegRNAs (four pegRNAs in total) and 750 ng of BxB1 mRNA using 20 µl of Lonza buffer and cells with program CM-130. Cells were recovered with 80 µl of pre-warmed (37 °C) media for 5 min. Then, 25-µl samples from the nucleofection cuvette were added to each well of the 48-well plate and incubated at 37 °C for 72 h before isolation of genomic DNA, as described above.

ddPCR analysis of knock-in efficiency. Genomic DNA from crude cell lysates was column purified (Zymo), and DNA concentrations were determined by NanoDrop (Thermo Fisher Scientific). ddPCR was used to determine the abundance of an amplicon containing the genome–donor junction in comparison to a reference gene (*ACTB*). Next, 100–200 ng of DNA was added to a reaction mixture containing ddPCR Supermix for Probes (Bio-Rad, 1863026), HindIII-HF (0.25 U µl^{-1}, New England Biolabs, R3104L), *ACTB* primers and probes (Supplementary Table 3; 900 nM each primer, 250 nM probe) and genome–donor junction primers and probes (Supplementary Table 3; 900 nM each primer, 250 nM probe), according to the manufacturer's protocol. Droplets were generated using a QX200 Manual Droplet Generator (Bio-Rad, 186-4002). ddPCR was performed as follows: 95 °C for 10 min and then 50 cycles of 94 °C for 30 s and 58 °C for 2 min. After PCR cycles, a final incubation was conducted at 98 °C for 10 min. Droplets were read on a QX200 Droplet Reader (Bio-Rad, 1864001) using QuantaSoft (version 1.4, Bio-Rad). Data were analyzed using QuantaSoft (version 1.4).

Factor IX expression from *ALB* locus in Huh7 cells. Huh7 cells were transfected with 160 ng of Bxb1 plasmid DNA, 160 ng of donor plasmid DNA (*attP*-splice acceptor-cDNA of *F9* exons 2–8), 400 ng of prime editor plasmid DNA and pegRNA plasmid DNA for installation of *attB* in the first intron of *ALB* or at *CCR5* (40 ng each). Three days after transfection, cells were passaged and allowed to grow to confluence. Their media were changed, and they were left to condition the fresh media, with aliquots taken at days 4, 7 and 10. Factor IX concentration in conditioned media was measured by ELISA (Innovative Research Human Total Factor IX ELISA Kit, IHUFIXKTT).

Lentivirus production, GFP reporter cell line construction and flow cytometry analysis. The lentiviral transfer vector described in Extended Data Fig. 9 and Supplementary Sequence 3 was assembled using standard molecular cloning techniques. The lentivirus was produced as previously described[54,55]. In brief, in a T-75 flask, rapidly dividing HEK293T cells (ATCC) were transfected at 80–90% confluence with lentivirus helper plasmids pVSV-G and psPAX2 along with the constructed lentiviral transfer vector using FuGENE HD (Promega), according to the manufacturer's protocol. After 48 h, supernatant was collected, centrifuged at 3,000*g* for 15 min to remove cellular debris and filtered using a 0.45-µm PVDF filter. Filtered supernatant was concentrated using PEG-it Virus Precipitation Solution (System Biosciences), according to the manufacturer's instructions. The resulting pallet was resuspended in Opti-MEM (Thermo Fisher Scientific) using 1% of the original medium volume. Resuspended pellet was used to transduce the HEK293T cells with multiplicity of infection. Successfully transduced cells that had stable GFP reporter integration were selected with 2.5 µg ml^{-1} of puromycin for more than three passages and then seeded into 48-well plates for transfection with PE2 plasmid DNA (750 ng), four *AAVS1*-targeting pegRNAs (62.5 ng each) for multiplexed *attP* and *attB* insertion and Bxb1 plasmid DNA (100 ng, 200 ng, 500 ng or 1,000 ng) in a single transfection step. Cells were then collected after 3 d and analyzed by CytoFLEX S Flow Cytometer (Beckman Coulter) and FlowJo version 10.

Production of PE2 and Bxb1 mRNA. The mRNA was produced using the previously reported protocol[37,56]. In brief, the PE2 editor or codon-optimized Bxb1 coding sequences were cloned into a plasmid encoding a mutated T7 promoter that cannot initiate with mononucleotides. This plasmid was used as template in a PCR reaction (Phusion U 2× Master Mix) using PAGE-purified primers (IDT) forward: 5′-TCGAGCTCGGTACCTAATACGACTCACTATAAGG-3′ and reverse: 5′-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTCTTCCTAC TCAGGCTTTATTCAAAGACCA-3′, and the resulting PCR product was purified and used as template for in vitro transcription. The HiScribe T7 High-Yield RNA Synthesis Kit (New England Biolabs) was used according to the company's recommended instructions but with full substitution of N1-methyl-pseudouridine for uridine and co-transcriptional capping with CleanCap AG (TriLink BioTechnologies). RNA was precipitated with lithium chloride (2.5 M final concentration), and the pellet was washed with ice-cold 70% ethanol and then resuspended in water.

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

Data availability

Source data are provided with this paper. High-throughput sequencing data have been deposited in the NCBI Sequence Read Archive database under accession PRJNA770428. Additional data used in the study can be accessed on Figshare (https://doi.org/10.6084/m9.figshare.c.5674708).

References

51. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
52. Clement, K., Farouni, R., Bauer, D. E. & Pinello, L. AmpUMI: design and analysis of unique molecular identifiers for deep amplicon sequencing. *Bioinformatics* **34**, i202–i210 (2018).
53. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).
54. Levy, J. M. & Nicoll, R. A. Membrane-associated guanylate kinase dynamics reveal regional and developmental specificity of synapse stability. *J. Physiol.* **595**, 1699–1709 (2017).
55. Koblan, L. W. et al. In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. *Nature* **589**, 608–614 (2021).
56. Gaudelli, N. M. et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **38**, 892–900 (2020).

Acknowledgements

We thank E. Sontheimer's group for sharing Huh7 cells and members of the Liu Laboratory for helpful discussions. This work was supported by the Merkin Institute of Transformative Technologies in Healthcare, National Institutes of Health grants U01 AI142756, RM1 HG009490 and R35 GM118062, the Bill and Melinda Gates Foundation, and the Howard Hughes Medical Institute. A.V.A. acknowledges a Jane Coffin Childs postdoctoral fellowship through the Howard Hughes Medical Institute.

Author contributions

A.V.A., X.D.G., C.J.P., A.T.N. and J.M.L. designed experiments. A.V.A., X.D.G., C.J.P., A.T.N., L.W.K., A.R. and J.A.M.M. performed experiments and analyzed data. A.V.A., X.D.G., C.J.P. and D.R.L. wrote the manuscript. D.R.L. supervised the research.

Competing interests

D.R.L. is a consultant and equity holder of Beam Therapeutics, Prime Medicine, Pairwise Plants and Chroma Medicine, companies that use genome editing or genome engineering technologies. A.V.A., C.J.P. and J.M.L. are currently employees at Prime Medicine. A.V.A., X.D.G., C.J.P., J.M.L. and D.R.L. have filed patent applications on twinPE and prime editing through the Broad Institute.

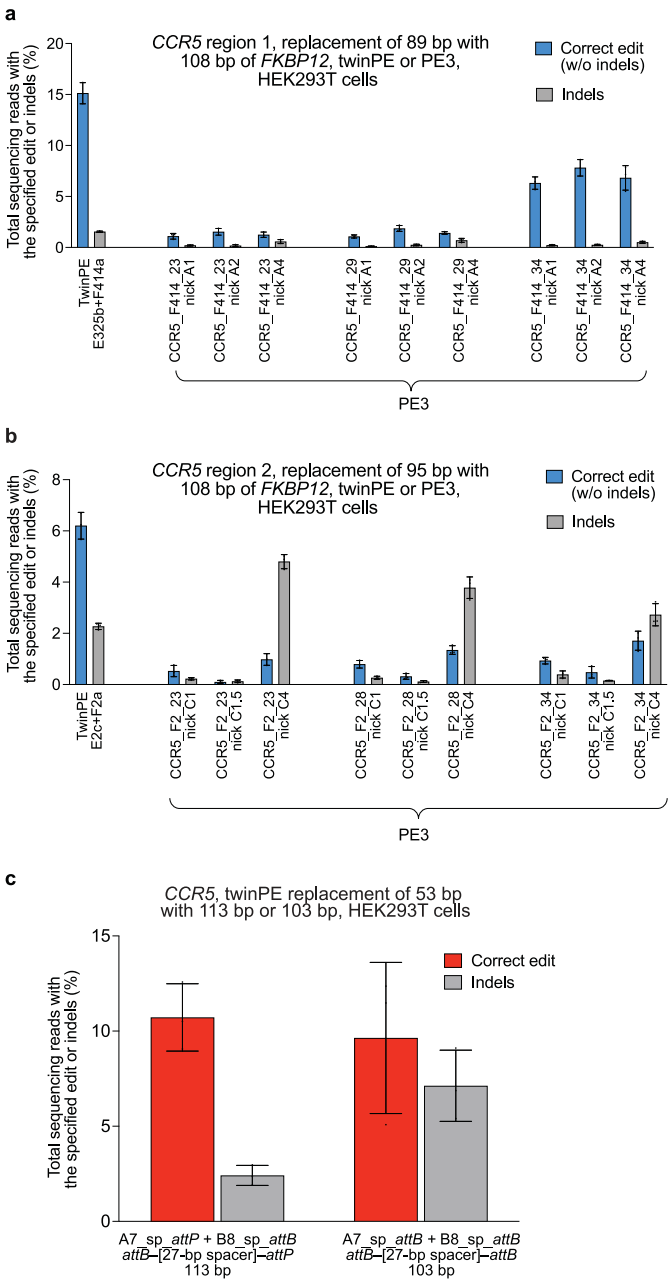
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41587-021-01133-w.

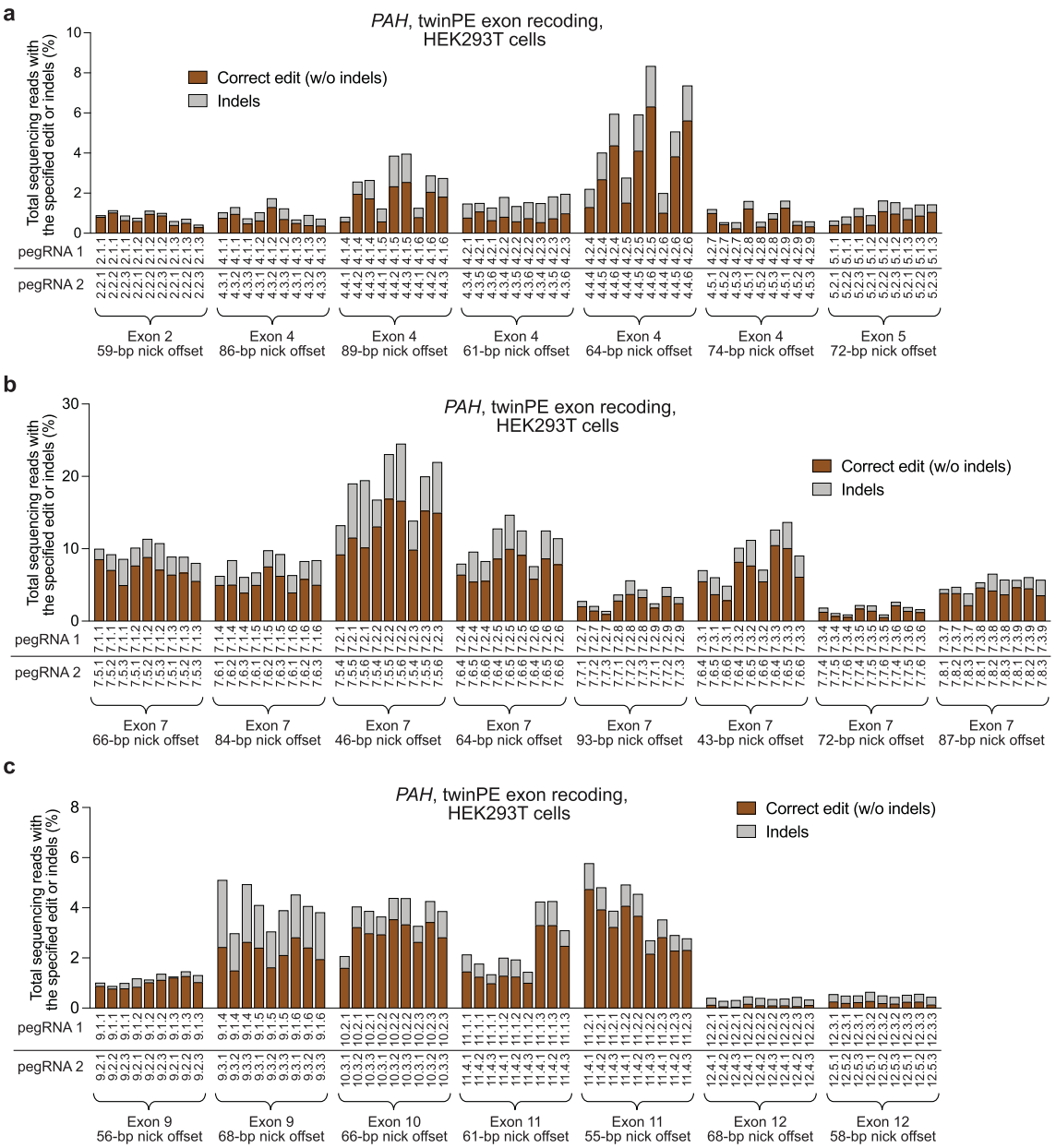
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-01133-w.

Correspondence and requests for materials should be addressed to David R. Liu.

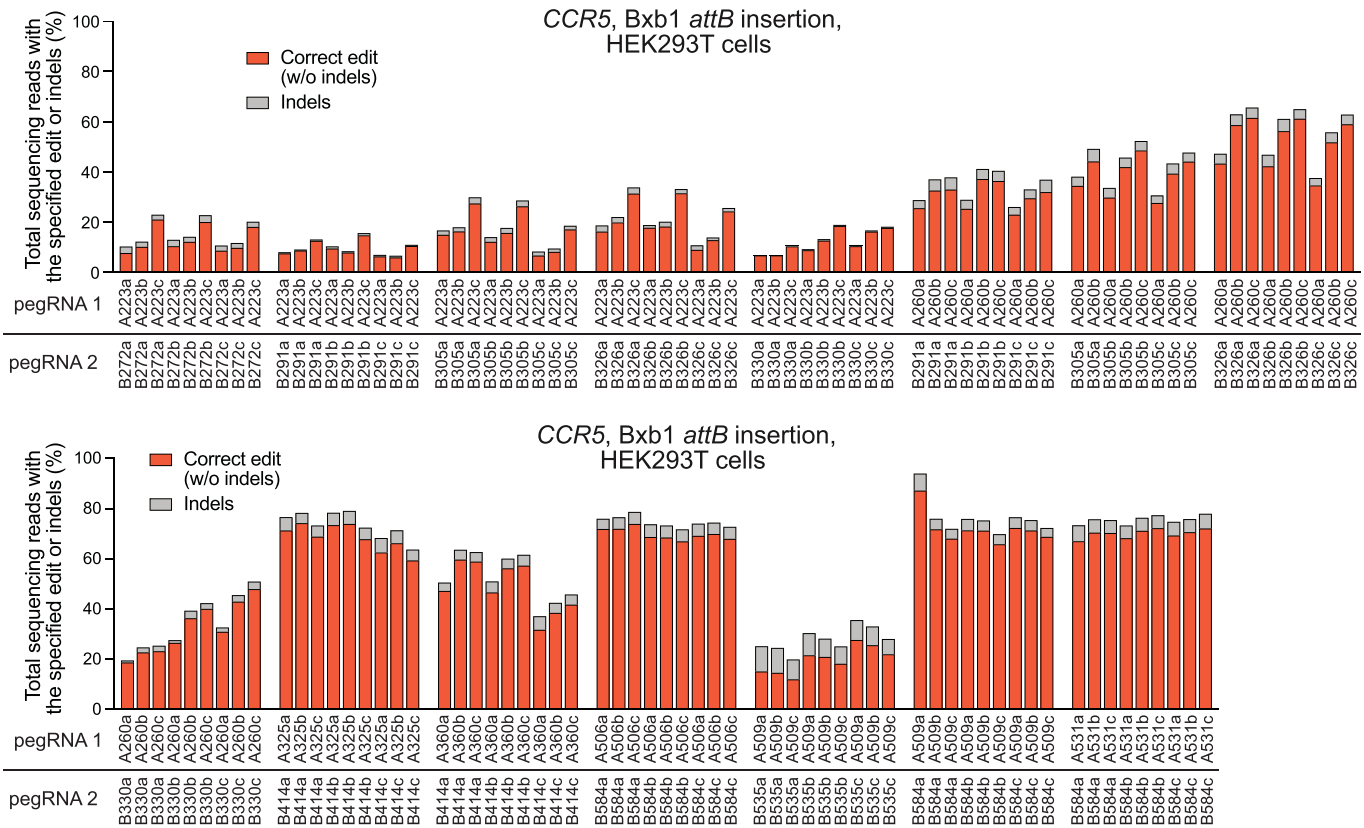
Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.



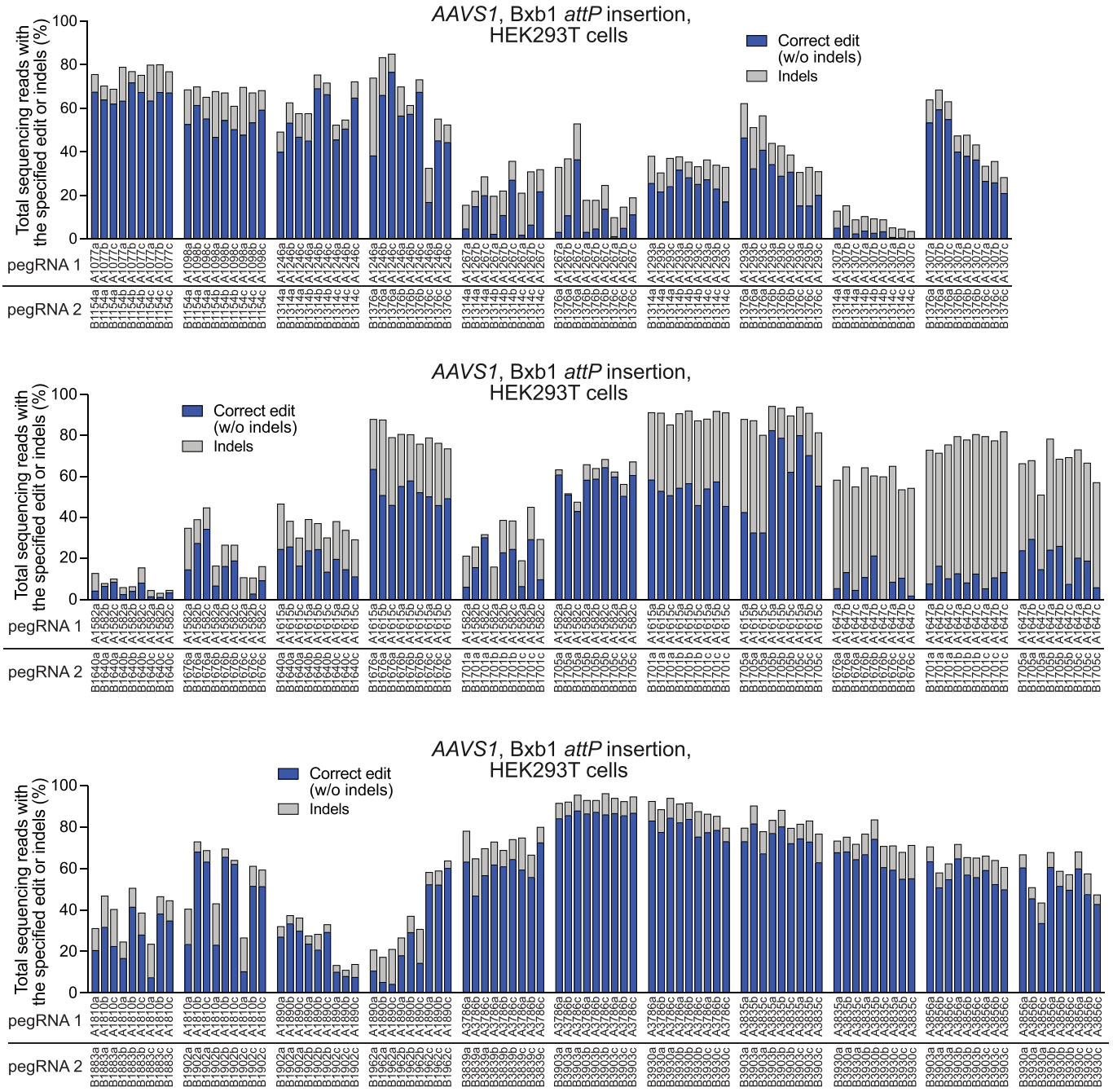
Extended Data Fig. 1 | Twin prime editing mediates sequence replacements at *CCR5*. (**a**) Replacement of endogenous sequence within *CCR5* region 1 with a 108-bp fragment of *FKBP12* cDNA using twinPE (*FKBP12* sequence oriented in the forward direction,) or PE3 (*FKBP12* sequence oriented in the reverse direction). For PE3 editing, pegRNA RT templates were designed to encode 108 base pairs of *FKBP12* cDNA sequence and one of three different target-site homology sequence lengths. For PE3 edits, each pegRNA was tested with three nicking sgRNAs. (**b**) Replacement of endogenous sequence within *CCR5* region 2 with a 108-bp fragment of *FKBP12* cDNA sequence using twinPE (*FKBP12* sequence oriented in the forward direction) or PE3 (*FKBP12* sequence oriented in the reverse direction). As in (a), PE3 edits were tested with pegRNAs containing RT templates that were designed to encode 108 base pairs of *FKBP12* cDNA sequence and one of three different target-site homology sequence lengths. For PE3 edits, each pegRNA was tested with three nicking sgRNAs. Values and error bars reflect the mean and s.d. of three independent biological replicates. (**c**) Transfection of HEK293T cells with a pair of pegRNAs targeting *CCR5* leads to replacement of 53 base pairs of endogenous sequence with 113 base pairs (*attB*–[27-bp spacer]–*attP*) or 103 base pairs (*attB*–[27-bp spacer]–*attB*) of exogenous sequence. Values and error bars reflect the mean and s.d. of three independent biological replicates.



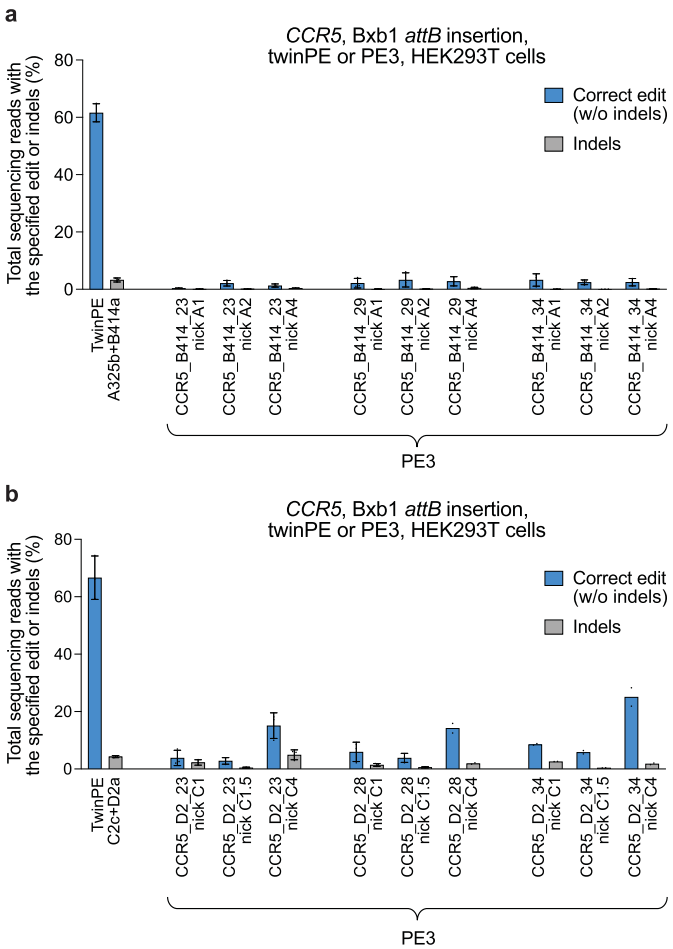
Extended Data Fig. 2 | Recoding of *PAH* exon sequences in HEK293T cells via twinPE. Screen of pegRNA pairs targeting *PAH* for recoding of (**a**) exons 2, 4 and 5, (**b**) exon 7, and (**c**) exons 9, 10, 11, and 12. RT templates of pegRNAs encoded partially recoded exonic sequence to optimize orthogonality to the endogenous gene sequence. For each spacer pair, nine pegRNA combinations were tested using three PBS variants for each spacer in a three-by-three matrix, with RT templates encoding the recoded exonic sequence, which was held constant for given spacer pairs. Sequences of pegRNAs are listed in Supplementary Table 1. Sequences of recoded exonic sequences are listed in Supplementary Table 4. Values in (a), (b) and exon 9 in (c) reflect single biological replicates. Values for exons 10, 11 and 12 in (c) reflect the mean of three independent biological replicates.



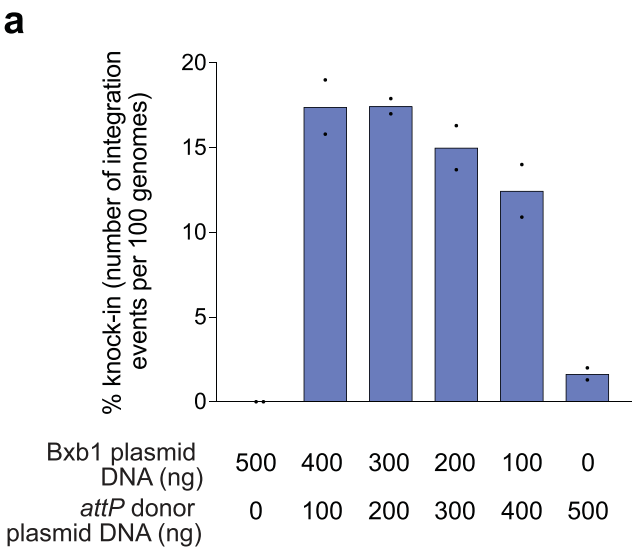
Extended Data Fig. 3 | Installation of a 38-bp Bxb1 *attB* site at *CCR5* with twinPE. Spacer pairs targeting the *CCR5* locus were designed for twinPE-mediated insertion of the Bxb1 *attB* attachment site. For each spacer, three pegRNAs were designed having three different PBS lengths and a fixed RT template that encodes the full-length Bxb1 *attB* sequence (38 bp). Sequences of pegRNAs are listed in Supplementary Table 1. For each spacer pair, a three-by-three matrix of pegRNA combinations was tested by plasmid DNA co-transfection with PE2 in HEK293T cells. Each pegRNA pair is specified below the x-axis. Values reflect single biological replicates.

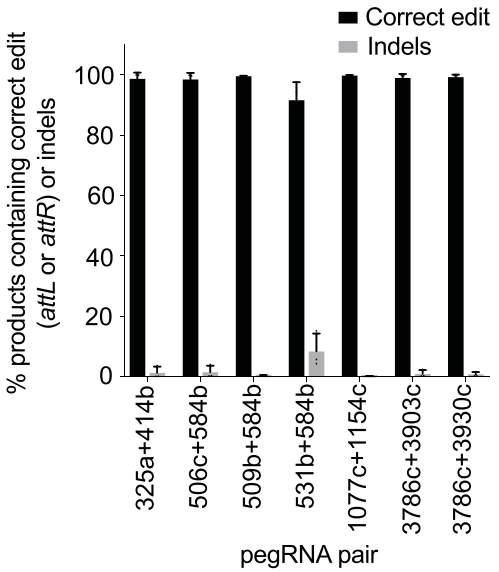


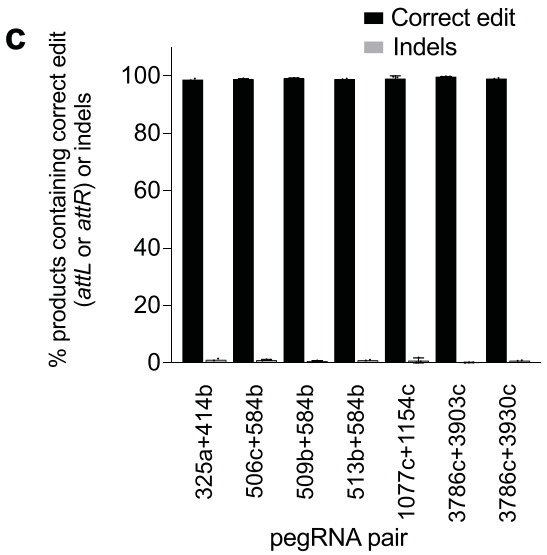
Extended Data Fig. 4 | Installation of a 50-bp Bxb1 *attP* site at *AAVS1* with twinPE. Spacer pairs targeting the *AAVS1* locus were designed for twinPE-mediated insertion of the Bxb1 *attP* attachment site. For each spacer, three pegRNAs were designed having three different PBS lengths and a fixed RT template that encodes a portion (43-44 bp) of the Bxb1 *attP* sequence. Sequences of pegRNAs are listed in Supplementary Table 1. For each spacer pair, a three-by-three matrix of pegRNA combinations was tested by plasmid DNA co-transfection with PE2 in HEK293T cells. Each pegRNA pair is specified below the x-axis. Values reflect single biological replicates.

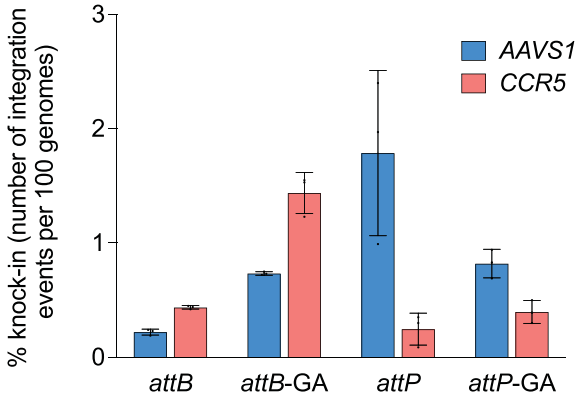


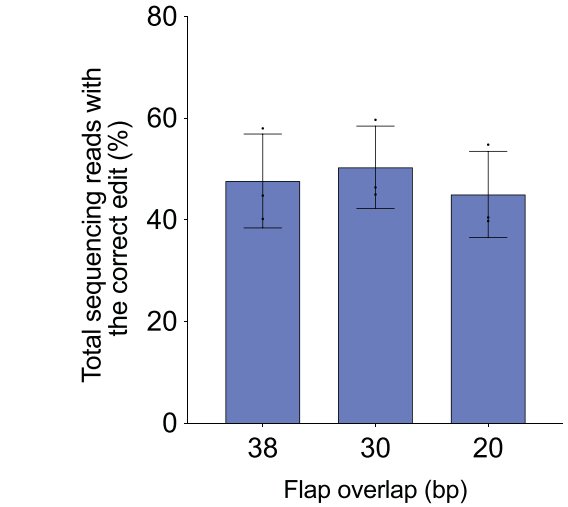
Extended Data Fig. 5 | Comparison of twinPE and PE3 for Bxb1 *attB* insertion at *CCR5*. (**a**) Replacement of endogenous sequence within *CCR5* region 1 with the Bxb1 *attB* site using twinPE or PE3. For PE3 editing systems, pegRNA RT templates were designed to encode the Bxb1 *attB* sequence and one of three different target-site homology sequence lengths. For PE3 edits, each pegRNA was tested with three nicking sgRNAs. (**b**) Replacement of endogenous sequence within *CCR5* region 2 with the Bxb1 *attB* sequence using twinPE or PE3. As in (a), PE3 edits were tested with pegRNAs containing RT templates that were designed to encode the Bxb1 *attB* sequence and one of three different target-site homology sequence lengths and tested with three nicking sgRNAs. Values and error bars in (a) and TwinPE edits, PE3 edits of *CCR5*_D2_23, *CCR5* D2_28 with nicking guide RNA C1 and C1.5 in (b) reflect the mean and s.d. of three independent biological replicates. Values of *CCR5* D2_28 with nicking guide RNA C4 and *CCR5* D2_34 in (b) reflect the mean of two independent biological replicates.

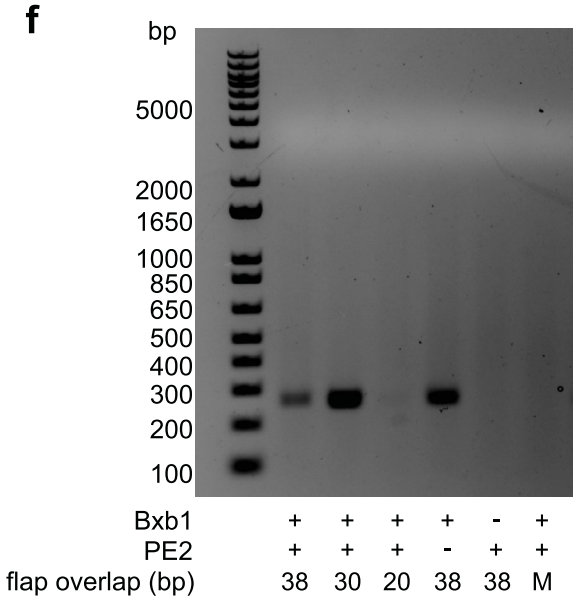








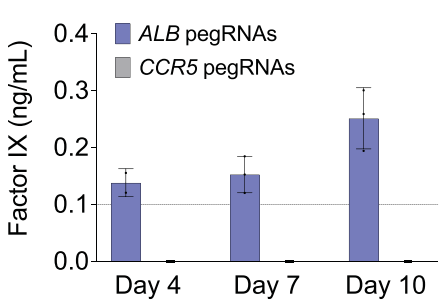




Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | TwinPE combined with Bxb1 recombinase for targeted knock-in of donor DNA plasmids. (**a**) Bxb1-mediated DNA donor knock-in in clonal HEK293T cell lines. Transfection of a HEK293T clonal cell line containing homozygous *attB* site insertion at *CCR5* with varying amounts of Bxb1-expressing plasmid and *attP*-containing donor DNA plasmid. Knock-in efficiency was quantified by ddPCR. Values and error bars reflect the mean of two independent biological replicates. (**b**) Assessment of genome-donor junction purity by high-throughput sequencing. Genomic DNA from single-transfection knock-in experiments was amplified with a forward primer that binds the genome and a reverse primer that binds within the donor plasmid (Supplementary Table 2). Values and error bars reflect the mean and s.d. of three independent biological replicates. (**c**) Assessment of genome-donor junction purity at the other junction by high-throughput sequencing as performed in (b). Values and error bars in 506c + 584b, 509b + 584b, 1077c + 1154c, and 3786c + 3903c reflect the mean and s.d. of three independent biological replicates. Values in 325a + 414b, 513b + 584b, 3786c + 3930c reflect the mean of two independent biological replicates. (**d**) Multiplexed single-transfection knock-in at *AAVS1* and *CCR5*. HEK293T cells were transfected with plasmids encoding PE2, Bxb1, a pair of pegRNAs for the insertion of *attP* at *AAVS1*, an *attB*-donor, a pegRNA pair for the insertion of one of four attachment sites (*attB*, *attB-GA*, *attP*, or *attP-GA*) at *CCR5*, and a corresponding donor. Knock-in was observed at both target loci under all four conditions. Insertion of *attP* at *AAVS1* and *attB* at *CCR5* gave the lowest knock-in efficiencies overall (0.2% at *AAVS1*, 0.4% at *CCR5*). Insertion of *attP* at both sites yielded the highest levels of knock-in at *AAVS1* (1.8%) but low levels (0.2%) at *CCR5*. When an orthogonal edit (*attB*-GA or *attP*-GA) was introduced at *CCR5*, *AAVS1* knock-in was 0.7-0.8%. Higher knock-in at *CCR5* was observed with *attB*-GA (1.4%) than with *attP*-GA (0.4%), consistent with our single locus knock-in results. Values and error bars reflect the mean and s.d. of three independent biological replicates. (**e**) and (**f**) Effects of reducing pegRNA overlap on twinPE efficiency and donor/pegRNA recombination. (**e**) The editing efficiencies of pairs of pegRNAs for insertion of Bxb1 *attB* at *CCR5* were measured by high-throughput sequencing. The pairs differed in the amount of overlap shared between their flaps, from 38 bp (full-length *attB* sequence) down to 20 bp. Editing efficiency of the pairs with shorter overlaps was comparable to the pair with full-length overlap. Values and error bars reflect the mean and s.d. of three independent biological replicates. (**f**) Assessment of recombination between *attB*-containing pegRNA plasmids and *attP*-containing donor plasmids. Following transfection of HEK293T cells with the indicated samples, isolated DNA was amplified with a forward primer that binds the pegRNA expression plasmid (TTGAAAAAGTGGCACCGAGT) and a reverse primer that binds the donor plasmid (CTCCCACTCATGATCTA). A positive 256-bp PCR band confirms recombination between the two plasmids. When the pegRNA encodes full-length *attB* (38-bp) or a truncated version of *attB* with 30-bp of overlap between flaps, a band is observed; however, recombination is not observed when the pegRNAs encode a truncated *attB* with only 20-bp of flap overlap. The 'No PE2' control uses the 38-bp overlap pegRNA pair. No recombination is observed in the absence of Bxb1 or if the donor and pegRNA plasmids both bear *attB* (Mismatch, 'M'). Three independent biological replicates were performed and a representative image from one of the replicates is shown.

a

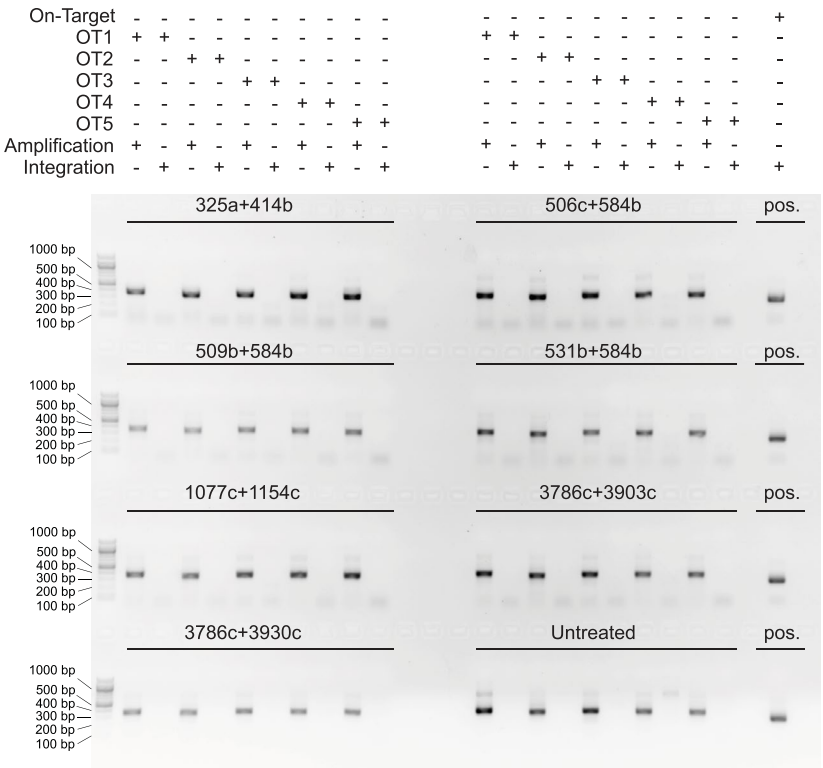


b

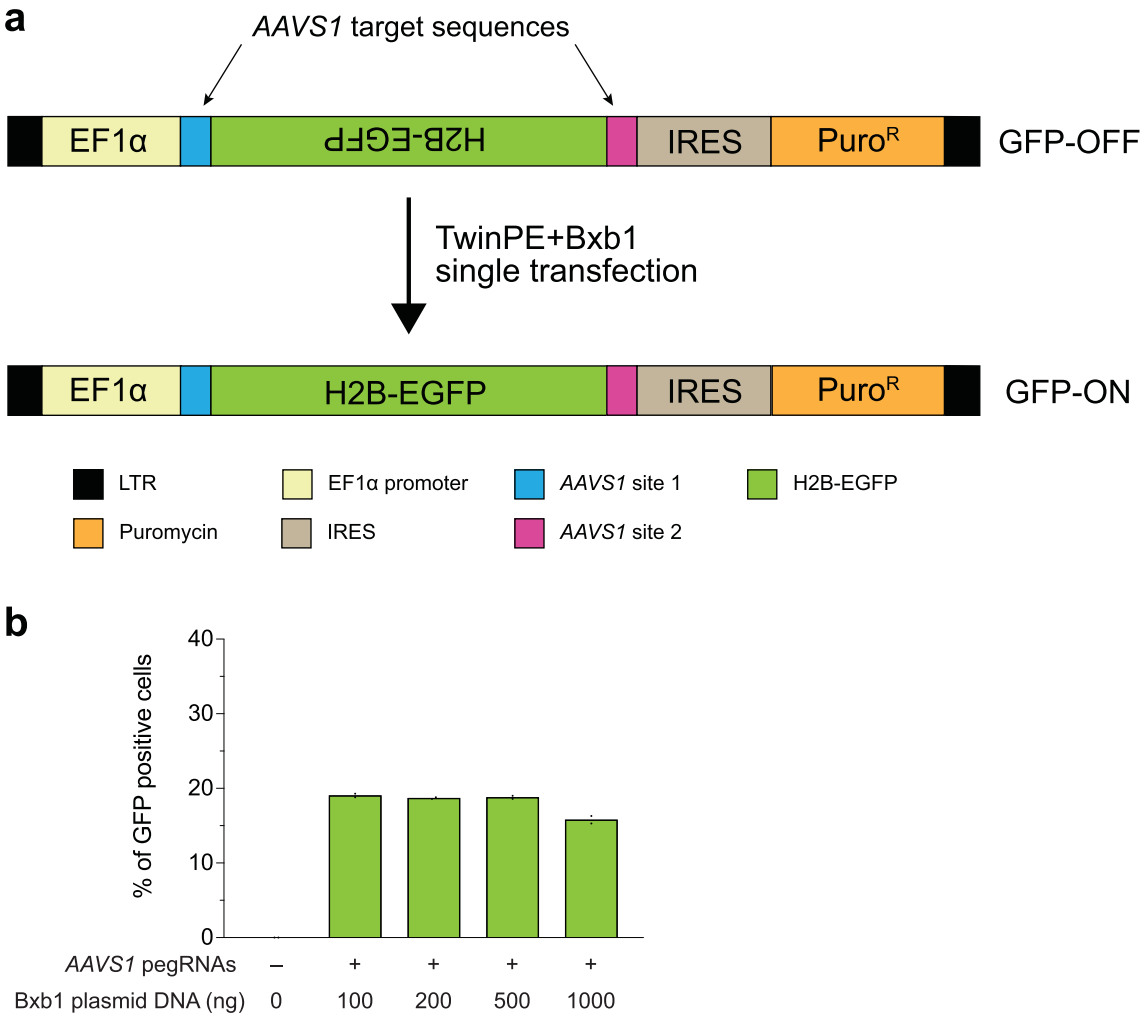
twinPE+Bxb1 mediated % editing at off-target sites with Bxb1 minimal substrate motif

		Targeted insertion at *CCR5*				Targeted insertion at *AAVS1*		
	Untreated	325a+414b	506c+584b	509b+584b	531b+584b	1077c+1154c	3786c+3903c	3786c+3930c
on-target site	<0.1	1.36	4.97	6.10	3.47	3.87	6.23	6.77
off-target site 1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
off-target site 2	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
off-target site 3	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
off-target site 4	<0.1	0.117	<0.1	<0.1	<0.1	<0.1	<0.1	0.141
off-target site 5	<0.1	<0.1	<0.1	0.157	<0.1	<0.1	<0.1	<0.1

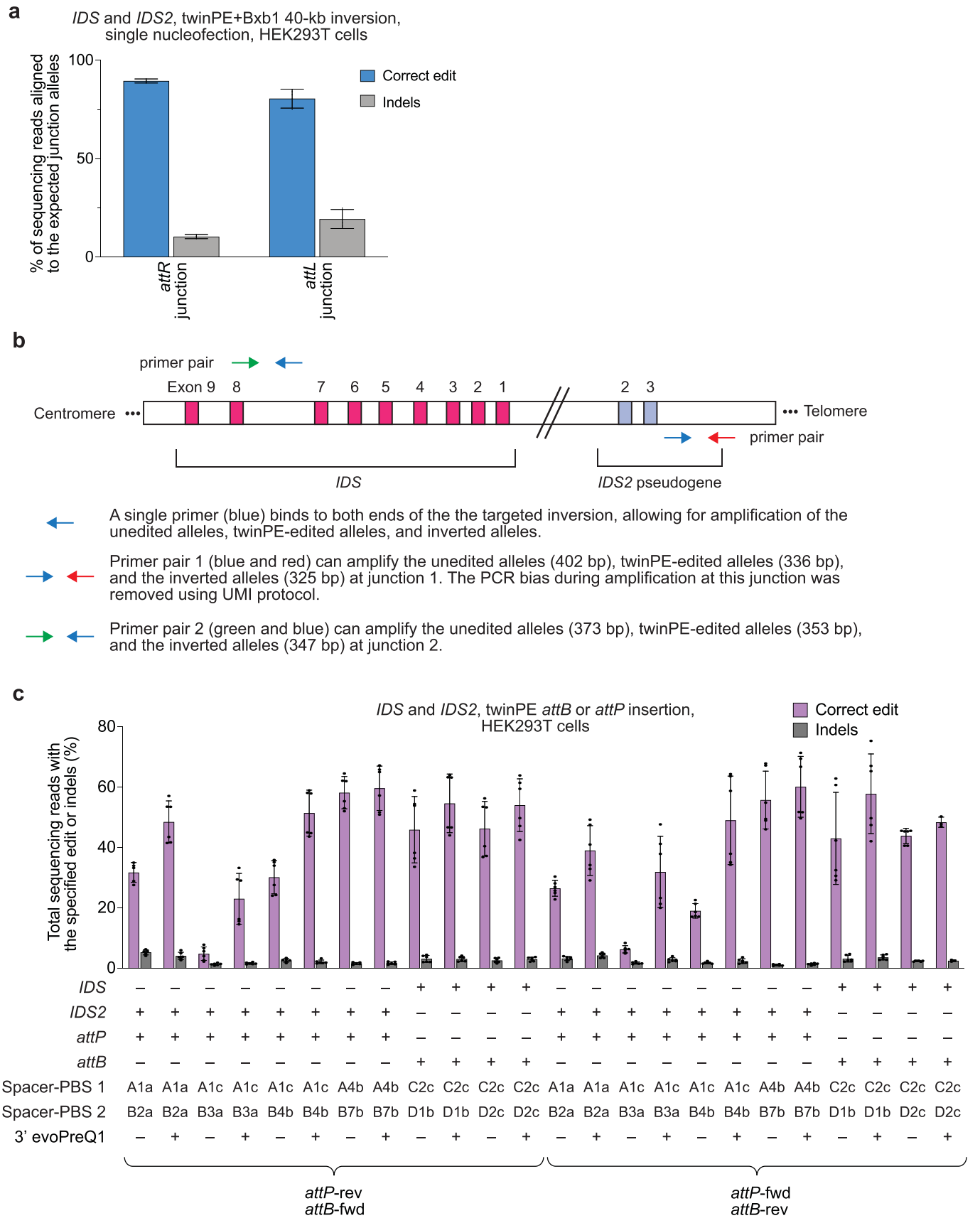
c



Extended Data Fig. 7 | Expression of human Factor IX from the *ALB* promoter following twinPE-recombinase knock-in and characterization of Bxb1 off-target editing. (**a**) Huh7 cells were transfected with Bxb1, donor (*attP*-splice acceptor-cDNA of *F9* exons 2-8), PE2, and pegRNAs for installation of *attB* in the first intron of *ALB* or at *CCR5*. Three days post-transfection, cells were split and allowed to grow to confluence. Their media was changed, and they were left to condition the fresh media, with aliquots taken at days 4, 7, and 10. Factor IX was present at detectable levels by ELISA (dashed line represents the lower limit of detection) in two of three samples treated with *ALB* pegRNAs at Day 4, and in all samples treated with *ALB* pegRNAs at Day 7 and Day 10. Factor IX was never detected in the conditioned media of any samples treated with *CCR5* pegRNAs. Values and error bars reflect the mean and s.d. of two or three independent biological replicates. (**b**) Targeted amplicon sequencing was performed for each of the five nominated pseudo-sites (OT1-OT5) from seven different samples treated with 5.6-kb donor DNA plasmid, twinPE reagents targeting *CCR5 or AAVS1*, and Bxb1 recombinase. The indels in all five pseudo-sites are either below the limit of detection (<0.1%) or near-background compared to untreated controls. The integration efficiency at the on-target site was measured by ddPCR as shown in Fig. 3d (**c**) To capture potential donor plasmid integration events at nominated pseudo-sites, primers were used to amplify predicted integration junctions. The gel depicts PCR reactions performed for each off-target site as indicated in the above legend. Confirmation of on-target donor integration from the samples is shown in the right-most column of the gel. In (b) and (c), two or three independent biological replicates were performed.

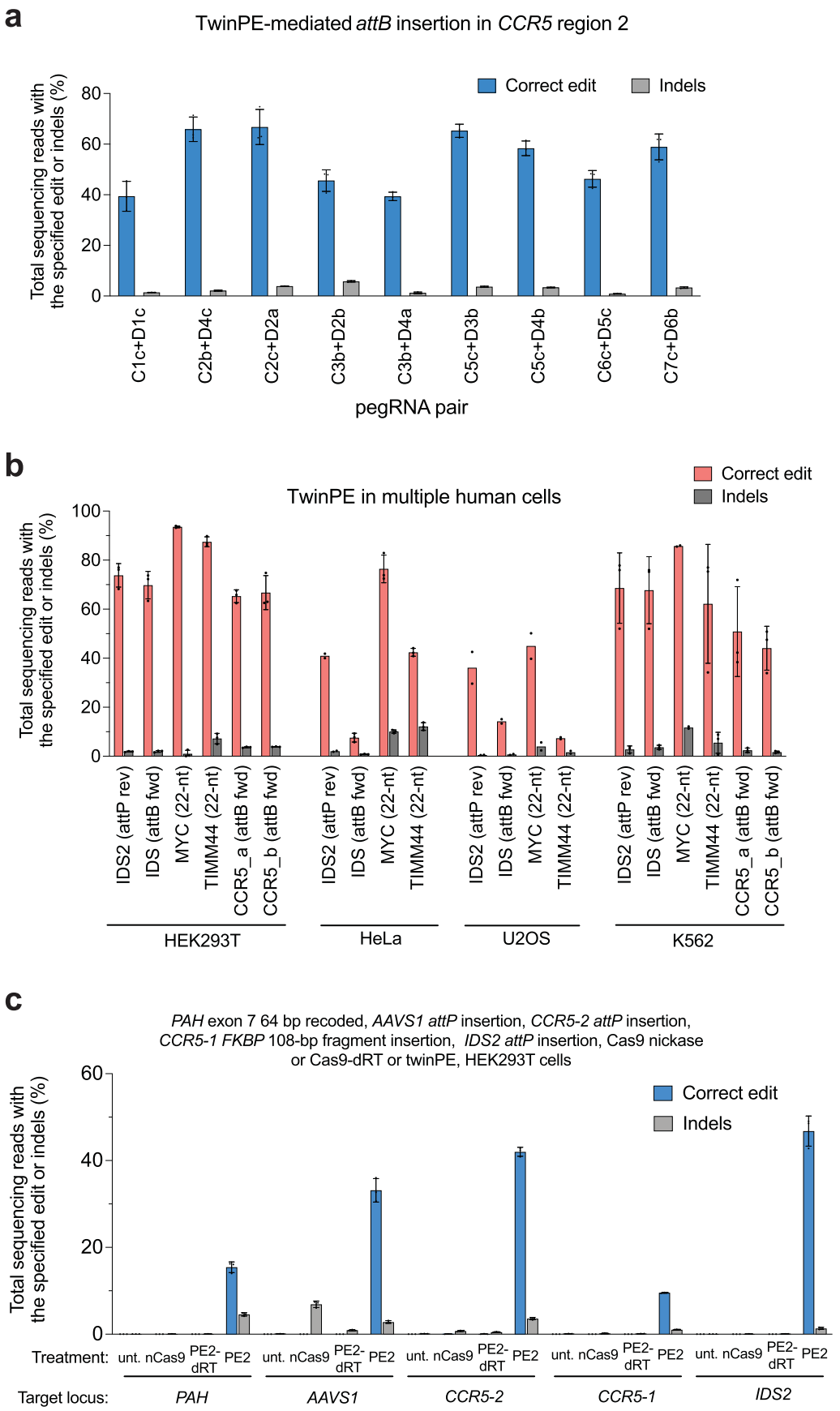


Extended Data Fig. 8 | TwinPE and Bxb1-mediated inversion in HEK293T GFP reporter cells. (**a**) The lentiviral fluorescent reporter construct used to assess inversion efficiency with twinPE and Bxb1 recombinase. The reporter contains an EF1α promoter followed by an inverted H2B-EGFP coding sequence that is flanked by partial *AAVS1* DNA sequence, an internal ribosome entry site (IRES), and a puromycin resistance gene. Successful installation of opposite-facing *attB* (left) and *attP* (right) sequences at the *AAVS1* target sequences and subsequent inversion by Bxb1 corrects the orientation of GFP for functional expression. (**b**) The fluorescent reporter construct was stably integrated into HEK293T cells via lentiviral transduction and puromycin selection. The polyclonal GFP reporter cell line was then transfected with twinPE plasmid components (PE2 and four pegRNAs) and varying amounts of Bxb1 plasmid for single-transfection inversion. Cells were analyzed by flow cytometry and gated for live single cells. Quantification of GFP positive cells by flow cytometry. Values and error bars reflect the mean of two independent biological replicates.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | TwinPE and Bxb1 recombinase-mediated inversion between *IDS* and *IDS2*. (**a**) Assessment of the inverted *IDS* junction purity by high-throughput sequencing in HEK293T cells. Frequency of expected junction sequences containing *attR* and *attL* recombination products after twinPE and BxB1-mediated single-step inversion. The product purities range from 81-89%. Values and error bars reflect the mean and s.d. of three independent biological replicates. (**b**) Schematic diagram of the designed PCR strategies for quantifying *IDS* inversion efficiency. Primer pair 1 (green forward and blue reverse primer) can amplify the unedited alleles (403 bp), twinPE-edited alleles (337 bp), and the inverted alleles (326 bp) at junction 1 in a single PCR reaction. Due to the size difference, a UMI protocol was applied to eliminate PCR bias during quantification of inversion efficiency. Similarly, using primer pair 2 (red forward and blue reverse primer), the unedited alleles (346 bp), twinPE-edited alleles (326 bp), and inverted alleles (320 bp) at junction 2 can be amplified in a single PCR reaction. Amplicons can then be sequenced by standard high-throughput sequencing protocols for amplicon sequencing. (**c**) Screening of pegRNA pairs for the insertion of Bxb1 *attB* and *attP* sequences at *IDS* and *IDS2*. TwinPE editing was tested with standard pegRNAs and epegRNAs containing a 3′ evoPreQ1 motif. Values and error bars reflect the mean and s.d. of three independent biological replicates.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Twin prime editing mediated insertion in *CCR5* region 2 in HEK293T cells, twin prime editing in multiple human cell lines, and editing activity of Cas9 nickase and PE2-dead RT variants. (**a**) TwinPE-mediated endogenous sequence replacement with Bxb1 *attB* attachment site in *CCR5* region 2 in HEK293T cells. (**b**) TwinPE-mediated endogenous sequence replacement with *attP*, *attB*, or 22-nt DNA sequences in multiple human cell lines. Six different pegRNA pairs targeting five loci were tested in HEK293T, HeLa, U2OS and K562 cells. HEK293T and HeLa cell were transfected with PE2 and pegRNA plasmids via Lipofectamine 2000 (Thermo Fisher) and *Trans*IT-HeLaMonster (Mirus), respectively. U2OS and K562 cells were nucleofected using Lonza 4D-Nucleofector and SE kit. DNA loci and the specified insertion edits are shown in the x-axis. (**c**) HEK293T cells were transfected with twinPE pegRNA pairs and either Cas9–H840A nickase (nCas9), PE2-dRT (a PE2 variant that contains K103L and R110S inactivating mutations to the RT domain), or PE2. Treatment with either nCas9 or PE2-dRT did not result in desired edits, while PE2 installed the specified edits as indicated. Values and error bars in (a) and (c) reflect the mean and s.d. of three independent biological replicates. Values and error bars in (b) reflect the mean and s.d. of at least two independent biological replicates except editing in *IDS2* in HeLa cells, editing in U2OS cells, and editing in *MYC* in K562 cells, which represent two independent biological replicates.

nature portfolio

Corresponding author(s): David R. Liu

Last updated by author(s): Oct 20, 2021

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	Illumina Miseq Control software (3.1) was used on the Illumina Miseq sequencers to collect the high-throughput sequencing data. Droplet digital PCR data was collected using QuantaSoft (v. 1.4, Bio-Rad).
Data analysis	Crispresso2 was used to analyze HTS data for quantifying editing activity at genomic sites. Droplet digital PCR data was analyzed using QuantaSoft (v. 1.4, Bio-Rad).

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

High-throughput sequencing data have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA770428.

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	Sample sizes were determined based on literature precedence for genome editing experiments (e.g. Anzalone et al., Nature 2019).
Data exclusions	No data was excluded.
Replication	All main text figure experiments were repeated at least once and all attempts at replication were successful. PegRNA screening experiments were performed in single replicate in some cases, but significant results were repeated successfully.
Randomization	Mammalian cells used in this study were grown under identical conditions; no randomization was used.
Blinding	Mammalian cells used in this study were grown under identical conditions; blinding was not used.

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
☒	☐ Human research participants
☒	☐ 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

Cell line source(s)	HEK293T (ATCC), U2OS (ATCC), K562 (ATCC), HeLa (ATCC), Huh7 (gift from Erik Sontheimer Lab at UMMS via ATCC).
Authentication	HEK293T, U2OS, K562, and HeLa were authenticated by the supplier using STR analysis. Huh7 cells were authenticated using ATCC's cell authentication.
Mycoplasma contamination	All cell lines tested negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	None used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	25,000 HEK293T stable GFP reporter cells were seeded onto 48-well poly-D-lysine coated plates (Corning). 16-24 h post-seeding, cells were transfected with 1uL of Lipofectamine 2000 (Thermo Fisher Scientific) using the protocol described in the method session and 750 ng of PE2, 250 ng AAVS1 targeting pegRNAs (62.5 ng each), and Bxb1 plasmid DNA (100 ng, 200 ng, 500 ng, or 1000 ng). The untreated and twinPE+Bxb1 treated cells were cultured for 72 hours, and then collected for flow cytometry analysis.
Instrument	CytoFLEX S Flow Cytometer (Beckman Coulter).
Software	FlowJo v10.
Cell population abundance	The abundances of twinPE+Bxb1 treated surviving single cells that showed GFP positive were 19.05%, 18.70%, 18.80%, and 15.8% at 100ng Bxb1, 200ng Bxb1, 500ng Bxb1, and 1000ng Bxb1 plasmid DNA conditions, respectively. The abundance of untreated surviving single cells that showed GFP positive were 0.01%.
Gating strategy	HEK293T stable GFP reporter cells were first gated (Gate A) based on forward (FSC-A) and side scattering (SSC-A) to remove dead cells and other debris. A second gate (Gate B) was used to select singlets based on FSC-H and FSC-A. Finally, GFP positive and GFP negative cells were gated (Gate C) and analyzed via FITC channel.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.