









# Harnessing DSB repair to promote efficient homology-dependent and -independent prime editing

Martin Peterka<sup>1</sup>, Nina Akrap<sup>1,4</sup>, Songyuan Li<sup>1,4</sup>, Sandra Wimberger<sup>1,2,4</sup>, Pei-Pei Hsieh<sup>1</sup>, Dmitrii Degtev<sup>1</sup>, Burcu Bestas<sup>1</sup>, Jack Barr<sup>1</sup>, Stijn van de Plassche<sup>1</sup>, Patricia Mendoza-Garcia<sup>1</sup>, Saša Šviković<sup>1</sup>, Grzegorz Sienski<sup>1</sup>, Mike Firth<sup>3</sup> & Marcello Maresca<sup>1</sup>

Prime editing recently emerged as a next-generation approach for precise genome editing. Here we exploit DNA double-strand break (DSB) repair to develop two strategies that install precise genomic insertions using an *SpCas9* nuclease-based prime editor (PE<sub>n</sub>). We first demonstrate that PE<sub>n</sub> coupled to a regular prime editing guide RNA (pegRNA) efficiently promotes short genomic insertions through a homology-dependent DSB repair mechanism. While PE<sub>n</sub> editing leads to increased levels of by-products, it can rescue pegRNAs that perform poorly with a nickase-based prime editor. We also present a small molecule approach that yields increased product purity of PE<sub>n</sub> editing. Next, we develop a homology-independent PE<sub>n</sub> editing strategy, which installs genomic insertions at DSBs through the non-homologous end joining pathway (NHEJ). Lastly, we show that PE<sub>n</sub>-mediated insertions at DSBs prevent Cas9-induced large chromosomal deletions and provide evidence that continuous Cas9-mediated cutting is one of the mechanisms by which Cas9-induced large deletions arise. Altogether, this work expands the current prime editing toolbox by leveraging distinct DNA repair mechanisms including NHEJ, which represents the primary pathway of DSB repair in mammalian cells.

<sup>1</sup>Genome Engineering, Discovery Sciences, BioPharmaceuticals R&D Unit, AstraZeneca, Gothenburg, Sweden. <sup>2</sup>Department of Chemistry & Molecular Biology, University of Gothenburg, Gothenburg, Sweden. <sup>3</sup>Data Sciences and Quantitative Biology, Discovery Sciences, AstraZeneca, Cambridge, UK. <sup>4</sup>These authors contributed equally: Nina Akrap, Songyuan Li, Sandra Wimberger. ✉email: [martin.peterka@astrazeneca.com](mailto:martin.peterka@astrazeneca.com); [marcello.maresca@astrazeneca.com](mailto:marcello.maresca@astrazeneca.com)

Proper utilization of cellular DNA repair mechanisms is instrumental to any successful genome editing strategy. The activity of different DNA repair pathways is highly dependent on tissue and cell type, chromatin context, and the DNA sequence of the target locus<sup>1–4</sup>.

Targeted DNA insertions represent a particularly challenging type of precise genome modification but have a considerable therapeutic potential. About 25% of ClinVar human pathogenic variants are deletions, the majority of which are <25 bp in length and thus potentially actionable by prime editing<sup>5</sup>. A common approach to introduce DNA insertions is to induce a targeted DNA double-strand break (DSB) using a site-specific nuclease combined with the delivery of a donor DNA repair template to stimulate homology-directed repair (HDR) at the targeted locus. A major disadvantage of this strategy is the limited activity of homologous recombination, which is restricted to S/G2 phases of the cell cycle and is generally absent in postmitotic cells<sup>6</sup>.

Unlike homologous recombination, DNA end joining repair mechanisms such as non-homologous end joining (NHEJ) or alternative end joining (a-EJ) pathways remain active throughout the cell cycle and act as the major pathways of DSB repair in mammalian cells<sup>7–9</sup>. While being typically considered error prone, NHEJ can repair DSBs with high fidelity<sup>10,11</sup>. In contrast, the homology-dependent a-EJ pathway leads to deletions, which are highly predictable<sup>12,13</sup>. The respective precision and predictability of NHEJ and a-EJ have been successfully exploited for precise genome modifications including DNA insertions<sup>14–17</sup>. Harnessing DNA end joining pathways represents a valuable genome editing strategy, because most adult tissues are comprised of postmitotic cells unable to perform homologous recombination<sup>3,18</sup>.

The recently developed CRISPR-based prime editing can install a wide spectrum of genomic modifications including deletions, substitutions, and insertions without the need of a separate DNA template and without introducing DSBs<sup>5</sup>, therefore offering a major advantage over existing genome editing methods. The PE2 prime editor combines Cas9 (H840A) nickase with an engineered reverse transcriptase (RT) to install an edit encoded directly in the prime editing gRNA (pegRNA). The cascade of events leading to a successful prime editing outcome is comprised of (1) Cas9-mediated nicking of the target site, (2) hybridization of the pegRNA-encoded primer binding site (PBS) to the 3' end of the nick, (3) pegRNA-templated extension of the primed 3' end of the nick by RT resulting in a “flap” containing the desired edit, and (4) hybridization and ligation of the flap with the targeted locus. Inhibition of mismatch repair was recently shown to enhance prime editing efficiency<sup>19</sup>, but DNA repair mechanisms responsible for the upstream steps of successful incorporation of the 3' flap remain to be described in detail and might not be universally available in different cellular and genomic contexts, potentially limiting the scope and efficiency of the nickase-based PEs. Recent reports suggest a possible dependency of PE2 editing on cell cycle progression<sup>20,21</sup>. Thus, a prime editing strategy harnessing a wider spectrum of DNA repair pathways would be a valuable addition to the prime editing toolbox.

Here we introduce Prime Editor nuclease (PE<sub>n</sub>), which combines RT and the wild-type SpCas9 nuclease and show that prime editing can be performed at DSBs by utilizing DNA end joining repair pathways. We present two PE<sub>n</sub> strategies to robustly install small insertions via distinct DNA repair mechanisms. The first strategy utilizes regular pegRNAs to promote small DNA insertions by a homology-dependent DSB repair mechanism. This strategy worked robustly across different genomic loci as well as with pegRNAs displaying inefficient editing when combined with PE2. The second strategy relies on a modified sgRNA design to install small insertions through precise NHEJ. We also present a

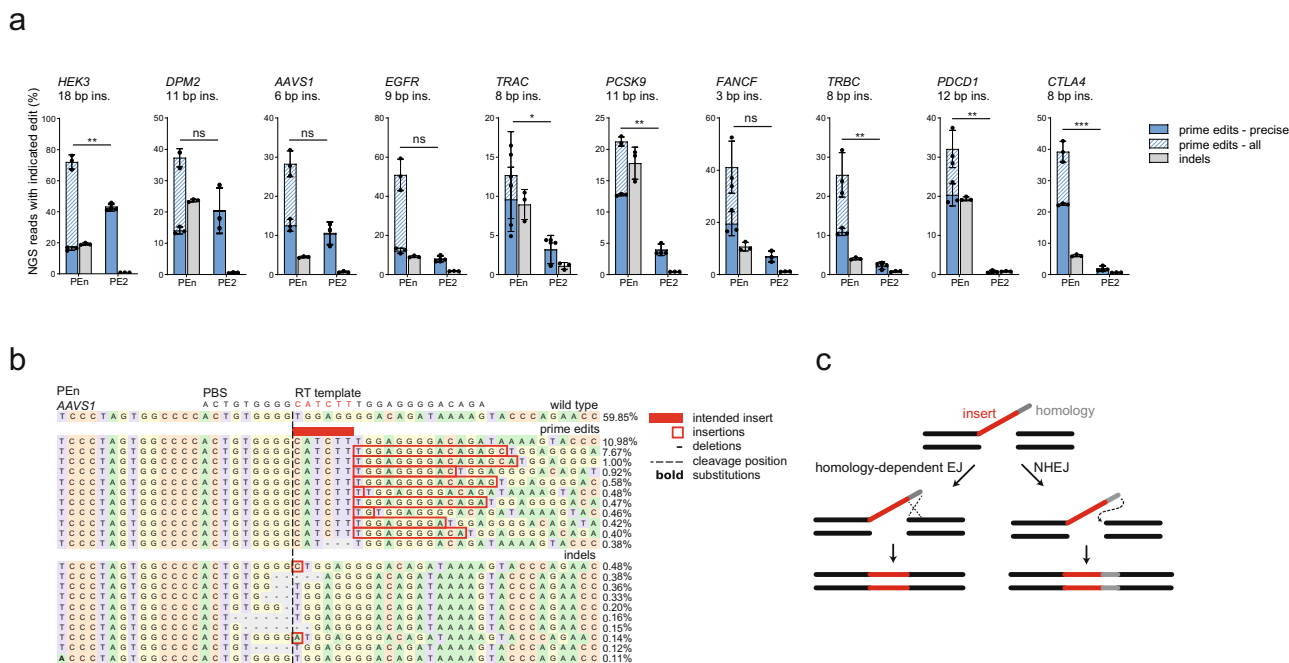
small molecule approach to decrease unintended by-products of PE<sub>n</sub> editing. Finally, we show that unlike editing with Cas9 alone, PE<sub>n</sub> does not induce large unintended on-target deletions, likely because PE<sub>n</sub>-mediated insertions at DSBs prevent NHEJ-mediated restoration of the wild-type sequence at the target locus. This suggests that the futile cycle of nuclease-mediated cut and NHEJ-mediated precise repair may be a possible cause of DSBs genotoxicity associated with Cas9 treatment.

## Results

**SpCas9 nuclease-based prime editing.** To test if a Cas9 nuclease-based prime editor can install small insertions at DSBs through a DNA end joining repair mechanism, we reverted the Cas9(H840A)-based PE2 into wtCas9-PE, designated here as PE<sub>n</sub>. We have constructed pegRNAs encoding small insertions of various sizes (6–18 bp) (Supplementary Data 1) against 10 genomic target sites and co-transfected HEK293T cells with a pegRNA and either PE<sub>n</sub> or PE2. NGS analysis of the editing outcomes at the targeted sites revealed successful intended insertions with varying frequencies and product purities for both PE<sub>n</sub> and PE2 (Fig. 1a). We classified the edited alleles into three categories; (1) all prime edits, representing any type of RT-templated insertions, (2) precise prime edits, that represent RT-templated insertions of intended size, and (3) other indels. As expected, PE2 editing resulted in high product purity, but also showed large site-to-site variability of insertion efficiency. PE<sub>n</sub> editing resulted in variable rates of precise prime edits but in general higher as compared to PE2. At some tested sites (*HEK3*, *DPM2*, *AAVS1*, *EGFR*) PE2 achieved similar or higher editing efficiency compared to PE<sub>n</sub> and clearly outperformed PE<sub>n</sub> in terms of precise editing purity. However, PE<sub>n</sub> installed insertions efficiently even with pegRNAs that were suboptimal for PE2 in our hands (Fig. 1a, *PCSK9*, *FANCF*, *TRBC*, *PDCD1*, *CTLA4*). We observed a similar trend in HeLa and HCT116 cells, where we tested PE2-optimal (*AAVS1*) as well as suboptimal (*CTLA4*) pegRNAs (Supplementary Fig. 1). As expected, due to the use of wild-type Cas9 that cuts both DNA strands, we also observed variable levels of PE<sub>n</sub>-induced imprecise prime edits and indels. Alignments of PE<sub>n</sub>-edited reads revealed that most of the imprecise prime edits represent additional integrations matching RT templates (Fig. 1b).

**Mechanism of PE<sub>n</sub>-based prime editing.** We reasoned that the integrated RT-templated homology tails might be products of DSB repair mediated by non-homologous end joining (NHEJ), while the precise insertions could occur through a homology-dependent process (Fig. 1c). If true, the inhibition of NHEJ could shift the outcomes of PE<sub>n</sub> editing by decreasing the frequency of imprecise prime edits. To test this, we performed PE<sub>n</sub> editing in HEK293T cells treated with AZD7648, a small molecule inhibitor of DNA-PK, an essential mediator of NHEJ<sup>22</sup>. Indeed, upon DNA-PK inhibition, the additional RT template integrations were abolished, and the remaining prime edits represented almost exclusively insertions of intended sizes (Fig. 2a, b). At several loci, DNA-PK inhibitor treatment also led to an increase of total rates of correct insertions (Fig. 2a – *DPM2*, *AAVS1*, *EGFR*).

Having pinpointed the contribution of NHEJ to PE<sub>n</sub> editing, we then investigated the mechanisms responsible for the homology-dependent DSB repair resulting in precise insertions. We reasoned that the short homology tails used in our pegRNA designs could utilize the a-EJ pathway, which typically uses DSB-proximal homology regions ~2–20 bp in length<sup>8</sup>. To test this hypothesis, we performed PE<sub>n</sub> editing using the *AAVS1* pegRNA with a 13 nt homology tail in HEK293T cells deficient in DNA Polymerase  $\theta$  (encoded by the *POLQ* gene), a crucial mediator of



**Fig. 1 SpCas9 nuclease-based prime editing.** **a** NGS analysis of PEn or PE2-mediated targeted DNA insertions of indicated sizes using 10 different pegRNAs targeting endogenous loci in HEK293T cells. Plots show mean  $\pm$  SD of  $n = 3$  biologically independent replicates. “prime edits - all” and “prime edits - precise” categories are superimposed.  $P$ -values were determined using Student’s paired  $t$  test (two-tailed) \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Calculated  $P$  values: *HEK3* = 0.0050, *DPM2* = 0.3075, *AAVS1* = 0.2569, *EGFR* = 0.0732, *TRAC* = 0.0433, *PCSK9* = 0.0028, *FANCF* = 0.0733, *TRBC* = 0.0021, *PDCC1* = 0.0055, *CTLA4* = 0.0003. **b** Representative alignment and allele frequencies of *AAVS1* locus edited with PEn and the indicated RT template in HEK293T cells. For each category, the top 10 variants are shown with a minimum frequency of 0.1%. **c** Model of homology-dependent and NHEJ modes of PEn-mediated insertions at DSBs. Source data for Fig. 1a are provided as a Source Data file.

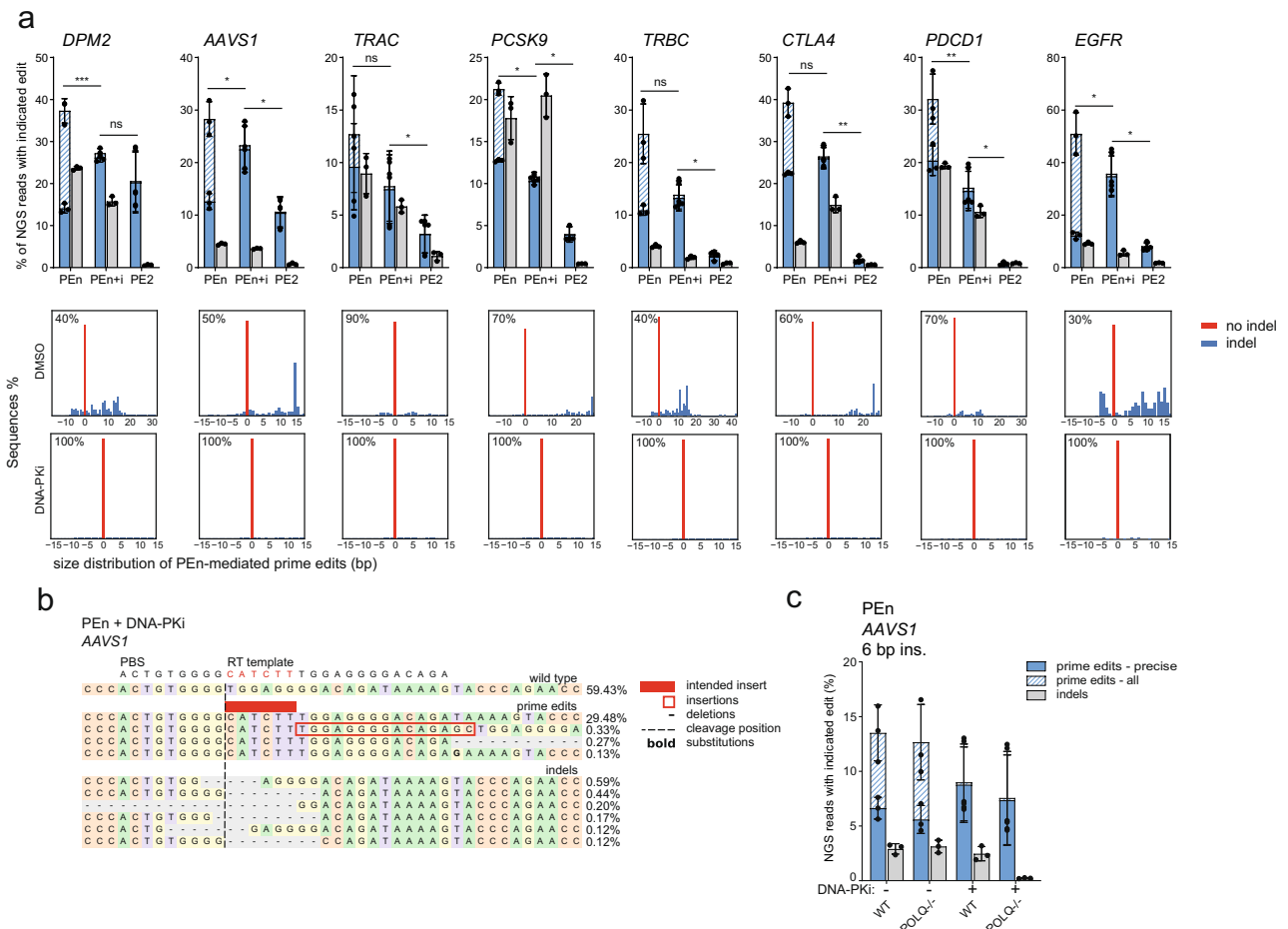
a-EJ<sup>8</sup>. First, to confirm a-EJ inhibition in *POLQ*<sup>-/-</sup> background, we performed Cas9 editing of the *AAVS1* locus in these cells with or without DNA-PK inhibition. As expected, no indels were detected at the targeted site upon DNA-PKi treatment of *POLQ*<sup>-/-</sup> cells (Supplementary Fig. 2), suggesting both NHEJ and a-EJ pathways were disabled. Despite this, PEn-mediated editing still proceeded efficiently, suggesting a mechanism independent of a-EJ (Fig. 2c).

Altogether, our data reveal that the imprecise PEn prime edits are mediated by NHEJ. Accordingly, DNA-PK inhibition improves the purity of PEn editing and leads to increased efficiency in a locus-dependent fashion. Interestingly, PEn-mediated precise insertions appear to be independent of the Pol  $\theta$ -mediated a-EJ pathway.

**PEn editing through NHEJ.** The observation that PEn-mediated imprecise edits were inserted via NHEJ prompted us to test whether a pegRNA design encoding the intended insertion, but no homology tail, could still perform precise primed insertions through NHEJ-mediated integration (Fig. 3a). This strategy could only be exploited to promote insertions at the cleavage site due to the end-to-end joining mechanism. To test this PRimed INsertions strategy (PRINS), we removed the homology region from the RT template of *AAVS1* pegRNA, resulting in a gRNA design with an extension containing only PBS and an intended insertion (Single PRimed Insertion gRNA, springRNA). PRINS editing of *AAVS1* using springRNA was able to install the intended insertion in HEK293T cells and was completely abrogated by DNA-PK inhibition, confirming that NHEJ is responsible for the PRINS-mediated insertions (Fig. 3b). The imprecise insertions constituted either truncated inserts or inserts longer than the intended size due to integrations of the gRNA scaffold sequence of various lengths (Fig. 3c). We have further tested this approach using a panel of springRNAs against different targets in HeLa and

HCT116 achieving variable but robust editing with up to 50% efficiencies (Fig. 3d, e). The unintended edits were sometimes prevalent such as in the case of *CTLA4*, where the top variant contained additional scaffold sequence (Fig. 3e and Supplementary Fig. 3). We have also tested all four 1 nt insertions at the *AAVS1* site and observed variable ratios of intended/scaffold-containing editing products, suggesting an effect of RT template and targeted DNA sequences on PRINS outcomes (Fig. 3d). Altogether, our results demonstrate that PEn can efficiently install precise insertions through NHEJ.

**Off-target analysis of PEn editing.** Integration of short double-stranded DNA fragments at DSBs has been exploited for Cas9 off-target detection<sup>23</sup> and integration of single-stranded DNA fragments was shown to increase both on- and off-target editing by Cas9<sup>24</sup>. Based on these studies, we reasoned that PEn might also show more pronounced off-target editing by actively modifying DSBs and in doing so preventing error-free DNA repair. To investigate PEn-mediated off-target editing, we have targeted three sites (*FANCF*, *HEK3* and *HEK4*) with gRNAs that were previously profiled for off-target editing with both Cas9 and PE2<sup>5,23</sup>. We used PEn and a matching PEn mutant (PEn-dRT) carrying previously reported RT-disabling mutations<sup>5</sup> with either pegRNAs or springRNAs against the three targets in HEK293T cells and analyzed both on-target editing and a total of 11 off-target sites by deep amplicon sequencing (Fig. 4). Compared to PEn-dRT, PEn induced up to 2-fold higher total on-target editing. Similarly, we observed that PEn increased off-target editing across most target sites. The increase ranged from moderate at *HEK4* off-targets (1.4–2.3-fold), to high at *FANCF* with off-target 1 reaching up to a 13-fold increase (Fig. 4). Examination of editing outcomes at these sites revealed that the increase was caused by RT-mediated insertions that constituted



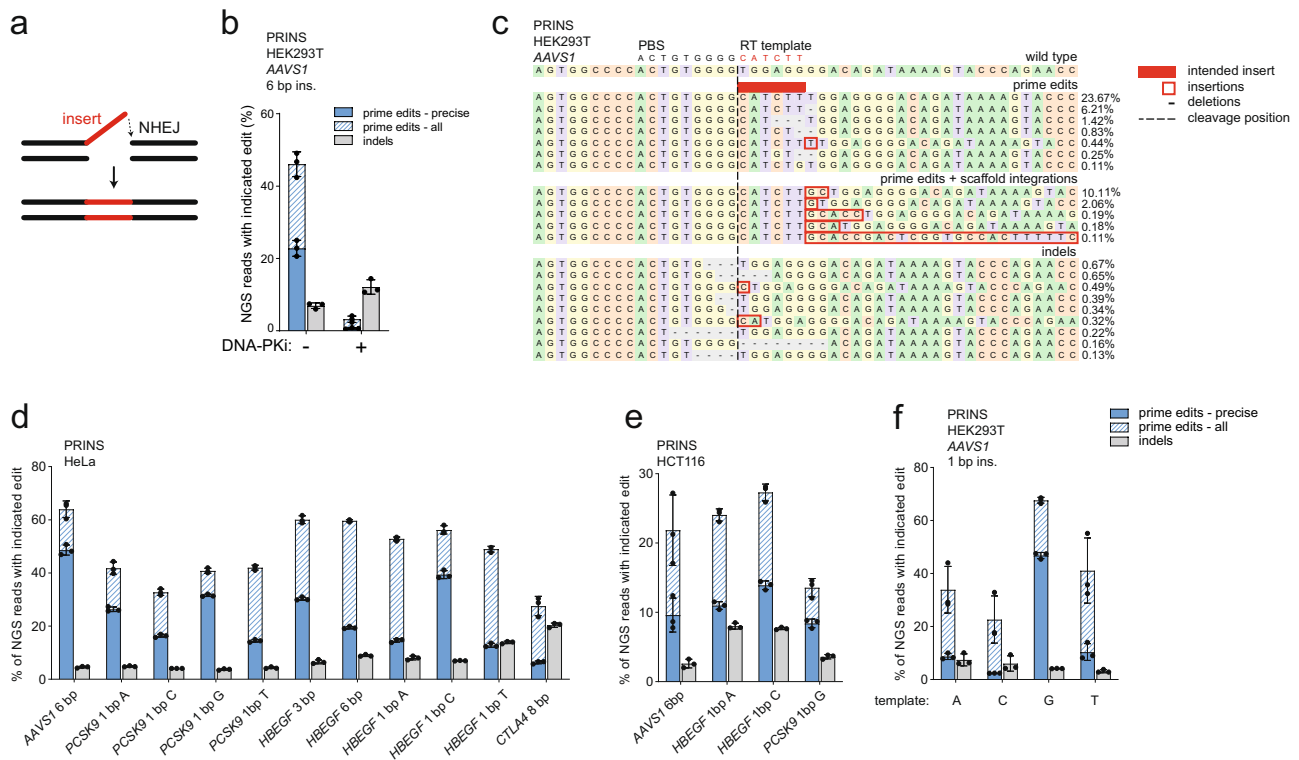
**Fig. 2 NHEJ mediates imprecise PEn editing.** **a** Selected data from Fig. 1a with additional DNA-PK inhibitor treatments of PEn samples. Plots represent PEn or PE2 editing using 8 different pegRNAs targeting endogenous loci in HEK293T. PEn edited cells were additionally pre-treated with DNA-PK inhibitor (PEn+i) or DMSO (PEn). Plots show mean ± SD of  $n = 3$  biologically independent replicates. “prime edits - all” and “prime edits - precise” categories are superimposed. Histograms below the bar plots represent percentages and size distribution of PEn prime edited alleles for each target site with or without DNA-PK inhibition.  $P$ -values were determined using Student’s paired  $t$  test (two-tailed) \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . Calculated  $P$  values: PEn vs. PEn+i ( $DPM2 = 0.00001$ ,  $AAVS1 = 0.0273$ ,  $TRAC = 0.0536$ ,  $PCSK9 = 0.0168$ ,  $TRBC = 0.1189$ ,  $CTLA4 = 0.1029$ ,  $PDCD1 = 0.0091$ ,  $EGFR = 0.0282$ ), PEn+i vs. PE2 ( $DPM2 = 0.3229$ ,  $AAVS1 = 0.0222$ ,  $TRAC = 0.0467$ ,  $PCSK9 = 0.0149$ ,  $TRBC = 0.0121$ ,  $CTLA4 = 0.0020$ ,  $PDCD1 = 0.0202$ ,  $EGFR = 0.0342$ ). **b** Representative alignment and allele frequencies of *AAVS1* locus edited with PEn and the indicated RT template in HEK293T cells treated with DNA-PK inhibitor. For each category, the top 10 variants are shown with a minimum frequency of 0.1%. **c** NGS analysis of PEn editing outcomes of *AAVS1* locus in wild-type and *POLQ*<sup>-/-</sup> and HEK293T cells with or without DNA-PK inhibitor treatment. The plot shows mean ± SD of  $n = 3$  biologically independent replicates. Source data for Fig. 2a, c are provided as a Source Data file.

the majority of edits across PEn-edited sites (Supplementary Fig. 4). Thus, these results show that efficient priming can increase PEn-mediated off-target editing and highlight a need for stringent peg/springRNAs and/or high fidelity Cas9 enzymes to be used with PEn.

**PEn-mediated insertions at DSBs mitigate Cas9-induced large deletions.** Cas9 editing has been shown to frequently cause large deletions spanning kilobase-sized regions surrounding the Cas9 target site<sup>25,26</sup>. This unintended consequence of Cas9 editing poses a potential roadblock for its therapeutic applications. As PEn generates DSBs, we wondered whether PEn editing also results in similar unwanted on-target editing. To test this, we used a diphtheria toxin (DT)-based selection system in HEK293T cells<sup>27</sup> to assay for large deletions induced by Cas9, PE2, and PEn. In this system, the disruption of the *HBEGF* coding sequence generates cells resistant to DT treatment, while cells carrying an intact copy of the *HBEGF* coding sequence are efficiently killed by DT (Fig. 5a). To monitor large deletions induced by different editors, we targeted

an intron of *HBEGF* with either Cas9, PE2 or PEn and subjected the edited cells to DT selection. The percentage of colonies surviving DT treatment normalized to the total *HBEGF* editing levels in each condition can be used to approximate the levels of large deletions in the cell population, as only cells carrying *HBEGF* deletions larger than ~600 bp acquire DT resistance. As expected, Cas9 editing led to a relatively high frequency of large deletions (Fig. 5b) confirming previous observations<sup>25,26</sup>. In contrast, nickase-based PE2 editing that does not induce DSBs only resulted in basal levels of large deletions. Surprisingly, similar to PE2, PEn editing with pegRNA or springRNA led to minimal levels of large deletions compared to Cas9 editing (Fig. 5b), despite efficient editing at the target site (Fig. S5a). We have analyzed large deletion patterns using PacBio long-read DNA sequencing<sup>26</sup> of the edited *HBEGF* locus prior to DT selection. The alignment of *HBEGF* long reads confirmed the presence of large deletions in Cas9-edited sample but not in PE2 or PEn-edited samples (Fig. 5c).

We hypothesized that Cas9-induced large deletions might be a result of cyclic targeted DNA cutting by Cas9 after precise DSB



**Fig. 3 Pen editing through NHEJ.** **a** Model of PRImed INsertions (PRINS) – an NHEJ-mediated mode of PE editing using springRNA. **b** NGS analysis of PRINS-mediated editing at *AAVS1* in HEK293T cells with or without DNA-PK inhibitor. **c** Representative alignment and allele frequencies of *AAVS1* locus edited with PRINS and the indicated RT template in HEK293T cells. For each category, the top 10 variants are shown with a minimum frequency of 0.1%. **d** NGS analysis of PRINS-mediated editing at indicated loci using a panel of springRNAs in HeLa cells. **e** NGS analysis of PRINS-mediated editing at indicated loci using a panel of springRNAs in HCT116 cells. **f** NGS analysis of PRINS-mediated 1 nt insertions at *AAVS1* using springRNAs with four different RT-templates in HEK293T cells. All plots show mean ± SD of *n* = 3 biologically independent replicates. “prime edits – all” and “prime edits – precise” categories are superimposed. Source data for Fig. 3b, d–f are provided as a Source Data file.

<i>FANCF</i>	target sequence/PBS	PAM	PE-n-dRT springRNA	PE-n-dRT pegRNA	PEn springRNA	PEn pegRNA	PE2 pegRNA
on-target	<b>GGAA<b>TCCCTTCTGCAGCACC</b></b>	<b>TGG</b>	33.63	35.77	63.50	52.28	7.08
off-target 1	GGAA <b>CCCCG</b> TCTGCAGCACC	AGG	1.50	1.34	19.60	12.12	0.03
off-target 2	GGAtTgCC <b>ATC</b> cGCAGCACC	TGG	0.16	0.12	0.30	0.20	0.01
off-target 3	GGAgTCCCT <b>cCT</b> aCAGCACC	AGG	0.04	0.04	0.24	0.14	0.01
off-target 4	aGAg <b>CCCCc</b> TCTGCAGCACC	AGG	0.07	0.04	0.35	0.24	0.01

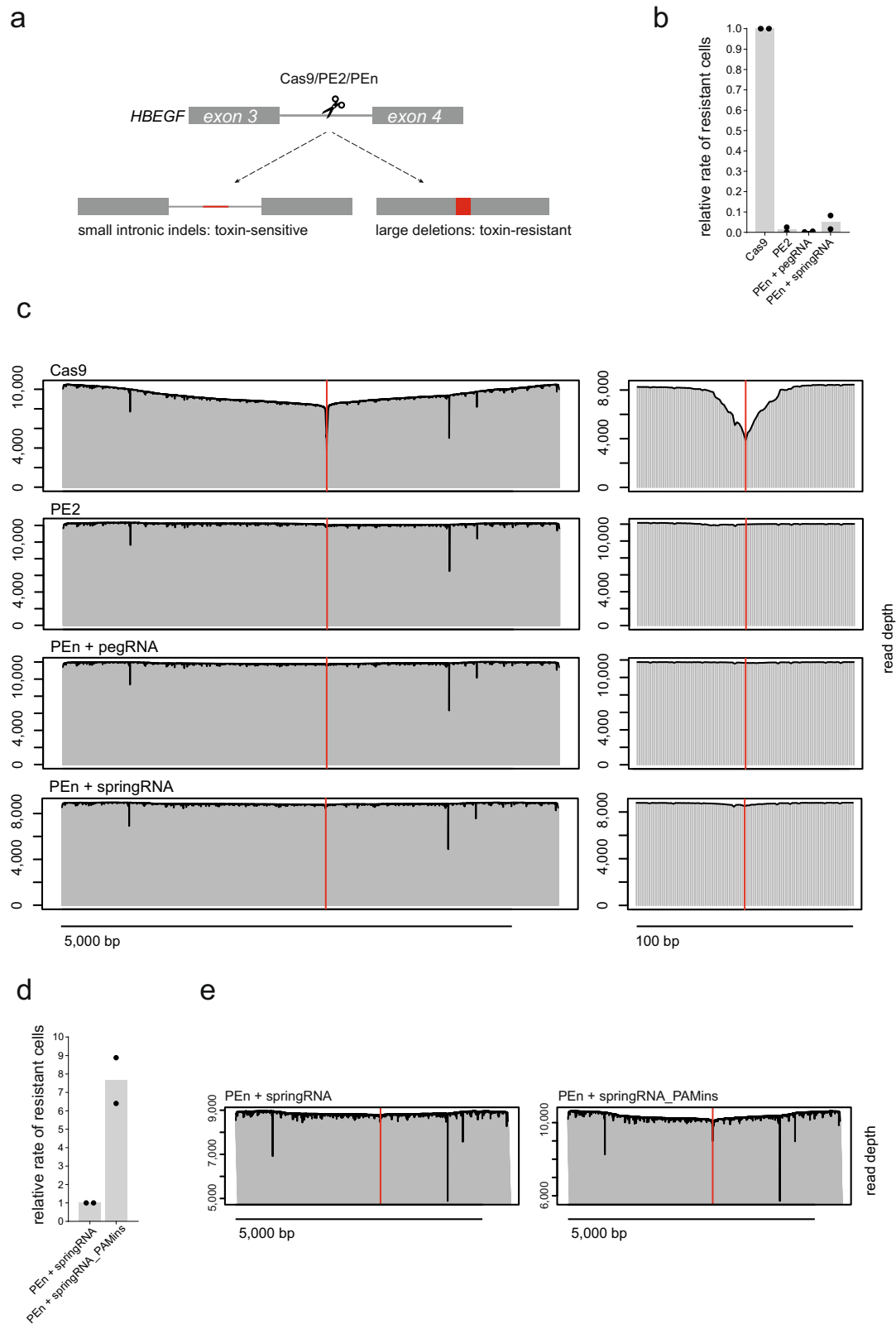
  

<i>HEK3</i>	target sequence/PBS	PAM	PE-n-dRT springRNA	PE-n-dRT pegRNA	PEn springRNA	PEn pegRNA	PE2 pegRNA
on-target	<b>GGCC<b>CAGACTGAGCAGTGA</b></b>	<b>TGG</b>	89.96	84.00	94.81	91.10	43.42
off-target 1	<b>ca</b> CCCAGACTGAGCAGT <b>Gc</b>	TGG	0.59	0.39	2.28	0.74	0.00
off-target 2	G <b>aCa</b> CAGAC <b>cGg</b> GCAGTGA	GGG	1.54	1.10	1.72	2.46	0.02
off-target 3	a <b>Gc</b> tCAGACTGAGCA <b>ag</b> TGA	GGG	0.35	0.28	0.47	0.69	0.15
off-target 4	a <b>Ga</b> CCAGACTGAGCA <b>aGa</b> G	GGG	0.01	0.01	0.02	0.02	0.00

<i>HEK4</i>	target sequence/PBS	PAM	PE-n-dRT springRNA	PE-n-dRT pegRNA	PEn springRNA	PEn pegRNA	PE2 pegRNA
on-target	<b>GGCA<b>CTGC<b>GGCTGGAGGTGG</b></b></b>	<b>GGG</b>	80.30	80.82	85.35	92.02	30.68
off-target 1	<b>t</b> GCAGTGC <b>GGc</b> GGAG <b>g</b> AGG	TGG	48.13	49.13	63.42	64.48	0.04
off-target 2	GGC <b>t</b> CTGC <b>GGc</b> TGGAG <b>g</b> GG	TGG	33.74	34.37	47.23	44.86	0.64
off-target 3	GGCA <b>tca</b> CGGCTGGAG <b>g</b> TGG	AGG	24.58	15.59	44.49	36.60	0.14

**Fig. 4 Off-target analysis of PE editing.** NGS analysis of editing outcomes at three on-target and eleven off-target sites with indicated editors and peg/springRNAs. Editing levels are shown as percentages of modified reads in each sample. The values represent the average of *n* = 3 biologically independent replicates. Mismatches to the on-target gRNA sequence are highlighted in red, the PBS region is highlighted in blue. Source data for Fig. 5 are provided as a Source Data file.



re-ligation by NHEJ. Since PEn does not rely on random indel generation by endogenous DSB repair system, it could efficiently disrupt this cycle by destroying the gRNA binding site upon successful RT-templated DNA insertion. To test this model, we have designed a springRNA encoding an insertion that reconstitutes the *HBEGF* gRNA binding site (PAMins), potentially allowing multiple rounds of PEn-mediated cutting. Indeed,

we have observed ~8-fold higher rates of DT-sensitive clones after PAMins springRNA editing relative to editing with a pegRNA encoding a random non-PAM insertion (Fig. 5d). Long-read sequencing of these two samples confirmed the more pronounced presence of large deletions upon PAMins editing (Fig. 5e). We have also performed long-read sequencing analysis post DT-selection to examine large deletion patterns in more detail

**Fig. 5 Large on-target deletion induction by Cas9, PE2, or PEn editing.** **a** Diphtheria toxin (DT) selection-driven assay to detect on-target large deletions induced by different genome editing systems. **b** Relative rates of surviving colonies after DT selection of cells edited by indicated genome editors. Data normalized to Cas9. The plot shows the mean of  $n = 2$  biologically independent replicates. **c** Alignment of long *HBEGF* reads from samples targeted with Cas9, PE2, or PEn and harvested before DT selection. Red lines denote Cas9 cleavage site. Scalebar 5000 bp. Panels on the right show window of 100 bp around the cleavage site. **d** Relative comparison of the rates of surviving colonies after DT selection of cells edited with PEn and either springRNA with a random insert of PAM-reconstituting springRNA. Data normalized to PEn + springRNA. The plot shows the mean of  $n = 2$  biologically independent replicates. **e** Alignment of long *HBEGF* reads from samples targeted with PEn and either springRNA with a random non-PAM insert of PAM-reconstituting springRNA (PAMins) harvested before DT selection. The y-axis is set from minimal to maximal read depth for each sample. Source data for Fig. 5b, d are provided as a Source Data file.

(Supplementary Fig. 5). PEn editing with a non-PAM-insert springRNA revealed a deletion landscape with discrete transitions in the coverage depth, suggesting that the detected large deletions originated from a small number of resistant clones, further confirming the rarity of PEn-induced large deletions. On the other hand, PAMins editing led to a complex and heterogeneous large deletion pattern resembling that of the Cas9-edited sample (Supplementary Fig. 5).

Altogether, our data show that PEn editing at the *HBEGF* locus does not induce considerable levels of unwanted large on-target deletions and thus might be a safer alternative compared to Cas9 editing. Additionally, we propose that multiple cycles of Cas9 cutting facilitated by precise repair of the target locus by NHEJ might be one of the mechanisms responsible for unwanted large deletions caused by Cas9 editing.

## Discussion

In this work, we present two different strategies to introduce precise genomic insertions using an *SpCas9* nuclease-based prime editor PEn. We showed that PEn promotes insertions through distinct DNA repair mechanisms, expanding the current nickase-based prime editing toolbox. In the first approach, we combined PEn with canonical pegRNAs to promote a homology-dependent DSB repair leading to precise insertions. Using PEn, we efficiently introduced insertions even with pegRNAs that performed poorly with PE2, suggesting that PEn can promote a more efficient DNA editing mechanism at the targeted locus. The highly efficient PEn editing also generated undesired consequences of DSB repair, such as indels, shorter prime edits and longer than intended prime edits that contained additional RT-template integrations. Similar bystander editing was also observed to various extents in the PE2 editing approach<sup>28</sup>. While the presence of the unintended integrations represents a downside of PEn editing, its high robustness and efficiency might be advantageous over the existing methods in situations where a seamless 3' end of the insertion to maintain an open reading frame of the target is not necessary, such as during the correction of frameshift mutations, gene disruption by defined stop codon integration or exon-intron junction editing. To control the DNA editing outcomes of PEn, we devised a strategy to remove the unintended prime edits by inhibiting DNA-PK, a crucial mediator of NHEJ<sup>8</sup>. For several genomic targets, the DNA-PK inhibitor treatment also led to a significant increase in precise editing levels. While this work was in revision, a study was published demonstrating that nuclease-based prime editing can outperform nickase-based prime editing at hard-to-edit targets as well as in mouse embryos, independently confirming and complementing our observations<sup>29</sup>. Additionally, a recent study utilized nuclease-based prime editing for the introduction of defined large genomic deletions<sup>30</sup>.

The mechanism of precise pegRNA-dependent PEn editing is likely a type of homology-dependent end joining DSB repair. We tested the involvement of a-EJ, a pathway that is utilizing small

homologies (2–20 bp) to repair DSBs. Nevertheless, our current data from a-EJ deficient cells suggest that Pol  $\theta$ -mediated a-EJ is not involved. Different homology-dependent modes of DSB repair such as single-strand annealing (SSA) or homologous recombination might be involved, but these are thought to require much longer homologies (>50 bp and >100 bp respectively)<sup>8</sup> than those present in our pegRNAs. Nevertheless, we cannot currently exclude those two possibilities. Future studies of PEn editing in systems with selectively inhibited different DNA repair enzymes will provide insights into its molecular mechanism.

The observation of NHEJ-mediated integrations of pegRNA RT templates during PEn editing led us to the development of the springRNAs. The springRNA does not require a homology sequence in the RT template and the intended insertion is installed through precise NHEJ. This mode of PEn editing (PRINS) could be of particular utility because NHEJ is a preferred type of DSB repair in most human cell types and acts independently on cell cycle progression<sup>3,8</sup>. NHEJ-driven precise genome editing has proved to be a valuable tool in the past, but unlike PRINS, the existing approaches rely on either separately provided dsDNA donors (larger than ~30 bp) or difficult-to-control indel generation<sup>14,15,31</sup>. Thus, to our knowledge, PRINS represents a unique way of installing small insertions via NHEJ.

Off-target analysis of PEn editing revealed that peg/springRNA-priming can increase the total editing levels at off-target sites to different extents. Further systematic investigation into peg/springRNA design and optimal high fidelity Cas9 utilization will be needed to fully understand and mitigate the off-target activity of PEn.

Our surprising observation that PEn does not induce large on-target deletions might provide a substantial advantage over Cas9 editing, where frequent large deletions can be of concern, especially in therapeutic applications<sup>25,26,32</sup>. Moreover, our data suggest a potential mechanism by which large deletions arise during Cas9-induced DSB generation. While the precision of NHEJ is controversial<sup>10,33</sup>, our data provide further evidence that NHEJ is inherently precise and possibly enables multiple cycles of target cleavage by Cas9. This “persistent” DSB may then increase the probability of faulty DNA repair leading to large deletion generation. This is in line with the observation in human embryos where long-lasting DSBs were suggested to be a potential cause of chromosomal loss or rearrangements<sup>34</sup>.

In conclusion, PEn editing is an effective method for introducing small genomic insertions and expands the spectrum of DNA repair mechanisms that can support prime editing, including NHEJ, which constitutes a major pathway of DSB repair in humans.

## Methods

**DNA constructs.** PE2, PEn, and *SpCas9* plasmids were generated by gene synthesis (GenScript). PE2 sequence including the backbone corresponds to the published CMV-PE2 construct (Addgene #132775). To generate PEn, the H840A Cas9 mutation in the PE2 construct was reversed to the original histidine. To generate the *SpCas9* construct, RT in PEn was replaced with eGFP. PEn dead-RT construct

was generated by introducing reported mutations in RT (M3-deadRT M-MLV RT(R110S, K103L, D200N, T330P, L603W)<sup>5</sup>. pegRNA constructs were generated by customizing protospacer, PBS and RT template in the target pMA-U6-pegRNA vector (GeneArt). Briefly, PCR fragments encoding pegRNAs flanked by 20 bp homology sequence matching pMA (Invitrogen) target backbone were generated by template-free PCR using two partially overlapping oligonucleotides. After PCR cleanup, the fragments were assembled into a linearized pMA backbone using HiFi DNA Assembly Master Mix (NEB) according to the manufacturer's protocol. All pegRNA sequences used in this work are listed in Supplementary Data 1.

**Cell culture, drug treatments, and transfections.** HEK293T (ATCC CRL-3216), HEK293T *POLQ*<sup>-/-</sup> (Synthego CRISPR KO pool, >90% indels), HCT116 (ATCC CCL-247) and HeLa (ATCC CCL-2) cells were cultured at 37 °C with 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal bovine serum. All cell lines were authenticated and regularly tested for mycoplasma. For gene editing experiments, cells were transfected using FuGENE HD reagent (Promega) as per manufacturer's instructions. For 96-well plate format, cells were seeded 24 h prior to transfection at 20,000 (HEK293T) or 10,000 (HeLa, HCT116) cells per well. Cells were transfected with 110 ng of plasmid DNA per well (55 ng of pegRNA/gRNA + 55 ng of pEn/PE2/Cas9). FuGENE:DNA ratio used for all transfections was 3:1. For larger wells, cell seeding numbers and transfected DNA amounts were scaled up accordingly. Cells were harvested for gene editing analysis after 72 h. In DNA-PKi experiments, AZD7648 (MedChemExpress, CAS No: 2230820-11-6) dissolved in DMSO was added to the growth medium 5 h prior to transfection to the final concentration of 1 μM.

**Genomic DNA extraction and sequencing analysis.** Cells were harvested using Quick Extract solution (Lucigen) according to manufacturer's instructions. Amplicons were generated using Phusion Flash High-Fidelity PCR Mastermix (F548, Thermo Scientific) in a 15 μL reaction, containing 1.5 μL of genomic DNA extract and 0.5 μM of target-specific primers with NGS adapters (primers #1-50, as listed in the Supplementary Data 1). Applied PCR cycling conditions: 98 °C for 3 min, 30x (98 °C for 10 s, 60 °C for 5 s, 72 °C for 5 s). PCR products were purified using HighPrep PCR Clean-up System (MagBio Genomics). Size, purity, and concentration of amplicons were determined using a fragment analyzer (Agilent). Amplicons were subjected to the second round of PCR to add unique Illumina indexes. Indexing PCR was performed using KAPA HiFi HotStart Ready Mix (Roche), 1 ng of PCR template and 0.5 μM of indexed primers in the total reaction volume of 25 μL. PCR cycling conditions: 72 °C for 3 min, 98 °C for 30 s, 10x (98 °C for 10 s, 63 °C for 30 s, 72 °C for 3 min), 72 °C for 5 min. Indexed amplicons were purified using HighPrep PCR Clean-up System (MagBio Genomics) and analyzed using a fragment analyzer (Agilent). Samples were quantified using Qubit 4 Fluorometer (Life Technologies) and subjected to sequencing using Illumina NextSeq system according to manufacturer's instructions. For off-target analysis, amplicons were generated using Q5 Hot Start High-Fidelity 2x Master Mix (M0494, NEB). Amplicons for long-read sequencing were generated with Q5 High-Fidelity polymerase (M0492S, NEB) using primers #51-52 (Supplementary Data 1) and the following PCR protocol: 98 °C 30 s 30x (98 °C 10 s 70 °C 10 s 72 °C 6 min) 72 °C 6 min.

**Bioinformatic analysis.** Demultiplexing of the NGS sequencing data was performed using bcl2fastq software. The fastq files were analyzed using CRISPResso2<sup>35</sup> in the prime editing mode with the quantification window of 5 starting from the 3' end of intended inserts. Detailed parameters are listed in the Supplementary Data 1. Prime edited override sequences were used for each site. To generate the representative alignments, the window was extended to 30 to visualize homology arm integrations of different lengths. Histograms in Fig. 2a were generated using CRISPResso2. Barplots were generated using GraphPad Prism 9 (GraphPad Software, Inc) or JMP 14.1.0 (SAS Institute Inc.). Long-read sequencing was performed by GeneWiz using PacBio platform. Resulting CCS reads were aligned to the reference sequence using minimap2<sup>36</sup> (2.2.15 with "--MD -a -xsplice -C5 -O6,24 -B4" options). The resulting sam files were processed using a custom python3 script to extract the read depth and location of deletions. The coverage plots were produced using R (3.4.2).

**Diphtheria toxin selection assay.** To assess the rate of large deletions induced by genome editing, HEK293T cells were transfected with different combinations of pEn/PE2/Cas9 and gRNA/springRNA/pegRNA followed by a survival assay based on DT selection. In the survival assay, transfected cells (>50% confluence) were treated with DT (Sigma-Aldrich) at 20 ng/mL. Cell viability was measured using the AlamarBlue cell viability reagent (ThermoFisher) before and after DT selection. The ratio of cell viability before/after the selection was calculated to indicate the rate of large deletions. Genomic DNA was harvested from each sample before and after DT selection and indel rates for each sample were analyzed by NGS.

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

## Data availability

Data supporting the results of this study are presented within the article and supplementary figures. NGS data are available in the NCBI Sequence Read Archive database (BioProject accession code PRJNA803881). Additional details and data to support the findings of this study are available from the corresponding authors upon reasonable request. Source data for Figs. 1a, 2a, c, 3b, d-f, 4, 5b, d, S1, S2, S4, S5a are provided as Source Data file. Source data are provided with this paper.

Received: 3 September 2021; Accepted: 7 February 2022;

Published online: 24 March 2022

## References

- Sun, S., Osterman, M. D. & Li, M. Tissue specificity of DNA damage response and tumorigenesis. *Cancer Biol. Med.* **16**, 396–414 (2019).
- Schep, R. et al. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol. Cell* **81**, 2216–2230.e10 (2021).
- Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* **21**, 1468–1478 (2019).
- Allen, F. et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64–72 (2019).
- Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
- Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).
- Yun, M. H. & Hiom, K. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. *Nature* **459**, 460–463 (2009).
- Zhao, B., Rothenberg, E., Ramsden, D. A. & Lieber, M. R. The molecular basis and disease relevance of non-homologous DNA end joining. *Nat. Rev. Mol. Cell Biol.* **21**, 765–781 (2020).
- Taleei, R. & Nikjoo, H. Biochemical DSB-repair model for mammalian cells in G1 and early S phases of the cell cycle. *Mutat. Res./Genet. Toxicol. Environ. Mutagen.* **756**, 206–212 (2013).
- Bétermier, M., Bertrand, P. & Lopez, B. S. Is non-homologous end joining really an inherently error-prone process? *PLoS Genet.* **10**, e1004086 (2014).
- Bhargava, R. et al. C-NHEJ without indels is robust and requires synergistic function of distinct XLF domains. *Nat. Commun.* **9**, 2484 (2018).
- Martínez-Gálvez, G., Manduca, A. & Ekker, S. C. MMEJ-based precision gene editing for applications in gene therapy and functional genomics. *bioRxiv* <https://doi.org/10.1101/2020.04.25.060541> (2020).
- van Overbeek, M. et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* **63**, 633–646 (2016).
- Maresca, M., Lin, V. G., Guo, N. & Yang, Y. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res.* **23**, 539–546 (2013).
- Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
- Shen, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
- Cristea, S. et al. In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.* **110**, 871–880 (2013).
- Saha, K. et al. The NIH somatic cell genome editing program. *Nature* **592**, 195–204 (2021).
- Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652.e5629 (2021).
- Schene, I. F. et al. Mutation-specific reporter for the optimization and enrichment of prime editing. *bioRxiv* <https://doi.org/10.1101/2021.05.08.443062> (2021).
- Wang, Q. et al. Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery. *Nucleic Acids Res.* **49**, 11986–12001 (2021).
- Fok, J. H. L. et al. AZD7648 is a potent and selective DNA-PK inhibitor that enhances radiation, chemotherapy and olaparib activity. *Nat. Commun.* **10**, 5065 (2019).
- Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
- Richardson, C. D., Ray, G. J., Bray, N. L. & Corn, J. E. Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. *Nat. Commun.* **7**, 12463 (2016).



25. Adikusuma, F. et al. Large deletions induced by Cas9 cleavage. *Nature* **560**, E8–E9 (2018).
26. 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).
27. Li, S. et al. Universal toxin-based selection for precise genome engineering in human cells. *Nat. Commun.* **12**, 497 (2021).
28. Petri, K. et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **40**, 189–193 (2021).
29. Adikusuma, F. et al. Optimized nickase- and nuclease-based prime editing in human and mouse cells. *Nucleic Acids Res.* **49**, 10785–10795 (2021).
30. Jiang, T., Zhang, X.-O., Weng, Z. & Xue, W. Deletion and replacement of long genomic sequences using prime editing. *Nat. Biotechnol.* **40**, 227–234 (2022).
31. Román-Rodríguez, F. J. et al. NHEJ-mediated repair of CRISPR–Cas9-induced DNA breaks efficiently corrects mutations in HSPCs from patients with fanconi anemia. *Cell Stem Cell* **25**, 607–621.e607 (2019).
32. Alanis-Lobato, G. et al. Frequent loss of heterozygosity in CRISPR–Cas9-edited early human embryos. *Proc. Natl Acad. Sci. USA* **118**, e2004832117 (2021).
33. Brinkman, E. K. et al. Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. *Mol. Cell* **70**, 801–813.e806 (2018).
34. Zuccaro, M. V. et al. Allele-specific chromosome removal after Cas9 cleavage in human embryos. *Cell* **183**, 1650–1664.e1615 (2020).
35. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
36. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

## Acknowledgements

We thank Steve Rees, Mohammad Bohlooly, and Mike Snowden for supporting this work. We thank Amelia Smith for proofreading the manuscript. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 765269 (S.W.). M.P. is a PostDoc fellow of the AstraZeneca R&D PostDoc program.

## Author contributions

M.P. and M.M. conceptualized the study. M.P., N.A., S.L., and S.W. performed most of the experimental work with help from P.H., D.D., J.B., S.v.d.P., P.M-G., S.Š, and G.S.;

M.F. performed bioinformatical analyses. M.P. prepared the manuscript with input from all authors. M.M. supervised the study.

## Competing interests

M.P., N.A., S.L., S.W., P.H., D.D., J.B., S.v.d.P., P.M-G., S.Š, G.S., M.F., and M.M. are employees and shareholders of AstraZeneca. B.B. is a former employee of AstraZeneca. M.M. is listed as inventor in an AstraZeneca patent application (WO2021204877A2) related to this work.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28771-1>.

**Correspondence** and requests for materials should be addressed to Martin Peterka or Marcello Maresca.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022