

TAPE-seq is a cell-based method for predicting genome-wide off-target effects of prime editor

Received: 4 March 2022

Accepted: 21 December 2022

Published online: 29 December 2022

 Check for updatesJeonghun Kwon¹, Minyoung Kim¹, Seungmin Bae¹, Anna Jo¹, Youngho Kim¹ & Jungjoon K. Lee¹✉

Prime editors (PEs) are powerful tools that widen the possibilities for sequence modifications during genome editing. Although methods based on the analysis of Cas9 nuclease or nickase activity have been used to predict genome-wide off-target activities of PEs, no tool that directly uses PEs for this purpose has been reported yet. In this study, we present a cell-based assay, named TAgmentation of Prime Editor sequencing (TAPE-seq), that provides genome-wide off-target candidates for PEs. TAPE-seq analyses are successfully performed using many different versions of PEs. The TAPE-seq predictions are compared with results from two other off-site prediction methods, Cas9 nuclease-based GUIDE-seq and Cas9 nickase-based Digenome-seq (nDigenome-seq). TAPE-seq shows a lower miss rate, and a higher area under the receiver operating characteristic curve compared to the other methods. TAPE-seq also identified valid off-target sites that were missed by the other methods.

CRISPR-Cas9 can introduce double-strand breaks (DSBs) at off-target as well as on-target sites, and various experimental protocols have been developed to predict such off-target activities at a genome-wide level. The methods can be categorized into three types depending on their mechanism of action: Cell-based (GUIDE-seq¹, GUIDE-tag², BLISS³, BLESS⁴, DISCOVER-seq⁵, integrase-defective lentiviral vector-mediated DNA break capture⁶, HTGTS⁷, CReVIS-Seq⁸, ITR-seq⁹, TAG-seq¹⁰, and INDUCE-seq¹¹), in vitro (e.g., Digenome-seq¹², DIG-seq¹³, CHANGE-seq¹⁴, CIRCLE-seq¹⁵, and SITE-seq¹⁶), and in silico (e.g., Cas-OFFinder¹⁷, CRISPOR¹⁸, and CHOPCHOP¹⁹). Because each has pros and cons, two or three methods have been used in combination to predict genome-wide off-target activities of CRISPR-based therapeutics^{20–22}.

These tools can also be used to predict genome-wide off-target activities of cytidine base editors (CBEs)²³ and adenine base editors (ABEs)²⁴. However, the development of more sophisticated versions of these prediction tools, such as the cell-based methods ONE-seq²⁵ and Detect-seq²⁶ and the in vitro methods CBE Digenome-seq²⁷, ABE Digenome-seq²⁸, and EndoV-seq²⁹ enabled more direct predictions, because these tools either use the same molecular mechanisms as the base editors or mimic these mechanisms.

Prime editor 2 (PE2) is a versatile genome editing tool that can insert, delete, or substitute nucleotides in target genomic DNA sequences³⁰. It consists of a fusion between catalytically impaired Cas9 nickase and an engineered reverse transcriptase (RT) complexed with a prime editing guide RNA (pegRNA), which contains a spacer sequence, a primer binding site (PBS), and a RT template that contains the desired edit. The Cas9 nickase, guided by the spacer sequence in the pegRNA, nicks the non-target DNA strand. The PBS in the pegRNA then binds to the single-stranded DNA released from the nicked strand, the end of which then primes reverse transcription of DNA using the RT template in the pegRNA. The newly synthesized DNA ultimately hybridizes with the uncleaved complementary DNA strand after cleavage of a 5' flap sequence, which lacks the edit, and is ligated with the nicked DNA strand. The mismatch in the heteroduplex is repaired via cellular repair mechanisms, resulting in the insertion of the RT template sequence at the target locus.

Because the first step of the PE2 mechanism is Cas9 nickase-induced nicking of the non-target DNA strand, it has been expected that the off-target activity of PE2 would resemble that of Cas9 or Cas9 nickase. Therefore, the genome-wide off-target activity of PE2 has

¹Toolgen, Seoul, Republic of Korea. ✉ e-mail: jj.lee@toolgen.com

been estimated using GUIDE-seq³⁰, nDignome-seq³¹, and in silico prediction tools like CAS-OFFinder^{17,32}, which measure or predict DSB or nickase activity of Cas9 nuclease or nickase. However, a method that directly measures the off-target activity of PE2 has not been reported. Because Cas9 and PE2 are different enzymes, a new method that directly measures the genome-wide off-target activity of PE is needed.

In this study, we develop a cell-based genome-wide off-target prediction tool for PEs named TAPE-seq, which involves direct analysis of PE activity in live cells. We optimize TAPE-seq by using various versions of PEs that had previously been analyzed using GUIDE-seq and nDignome-seq, allowing comparisons to be made between the three methods.

Results

Optimization of the tagmentation rate

Experimental genome-wide off-target prediction methods can be categorized as either cell-based or in vitro based³³. Because prime editing is a multi-step process involving many cellular enzymes, including flap endonuclease, exonuclease, and ligase, it is difficult to develop an in vitro-based assay that closely mimics this complex cellular process. On the other hand, most of the cell-based methods introduce tag sequences into on- and off-target loci so that they can be amplified by PCR during a later step. Since PE2 nicks its target without causing a DSB, it is not possible to insert double-stranded oligonucleotides or viral DNA fragments as tags for amplification purposes.

However, PE2 itself has the ability to insert any short sequence into the target site. We, therefore, designed pegRNAs with an additional 34-bp tag sequence between the PBS and RT template sequences. For the tag, we chose the same sequence that is used in GUIDE-seq¹, because it has been proven to work in cells from many different origins. We also chose PBS and RT template sequences that were used in validation experiments after GUIDE-seq³⁰ and nDignome-seq³¹ were used as prediction tools (Supplementary Data 1).

The signal-to-noise ratio of the developed off-target prediction method would be proportional to the efficiency of tag insertion at on- and off-target loci. We therefore first optimized the experimental conditions for tag integration into the on-target site. When plasmids encoding PE2 and a *HEK4*-targeting pegRNA (incorporating a +2 G to T edit, numbered relative to the nick) containing the tag sequence were transiently transfected into HEK293T cells, a tag integration rate of only 0.011% was observed. To improve this rate, we constructed an all-in-one vector encoding PE2 and the pegRNA in the piggyBac system³⁴. A stable cell line was constructed via transfection of this vector with transposase; in this situation, the tag integration (tagmentation) rate increased to more than 2% (Fig. 1a) after 14 days of puromycin selection. [Puromycin selection for 14 days successfully enriched green fluorescent protein (GFP) positive cells following transfection with a GFP-piggyBac construct (Supplementary Figure 1)]. The improvements in the number of targets found were not significant even if we prolonged the incubation time from 2 to 7 weeks (Supplementary Figure 2a). [We have assigned a similar number of Miseq reads to the 2 week (5329899), 4 week (5313548), 6 week (2324242), and 7 week (4021702) samples (Supplementary Data 4). The higher number of on-target reads in the 2 week sample (62565) compared to the 4 week (2369), 6 week (1060), and 7 week (1594) samples (Supplementary Data 3) could simply indicate a higher signal-to-noise ratio from the TAPE-seq analysis of the 2 week sample compared to the other samples.] Therefore, puromycin selection was performed for 2 weeks in subsequent studies.

We further optimized the tagmentation rates by finding the optimum amount of piggyBac vector to co-transfect with the transposase plasmid, testing amounts ranging from 50 ng to 1000 ng. When the copy number of the piggyBac vector was measured, 1000 ng

resulted in the highest value (Supplementary Figure 2b). In addition, 1000 ng consistently resulted in high tagmentation rates at on-target (Supplementary Figure 2c) and off-target sites (Supplementary Figure 2d). Therefore, we transfected 1000 ng of piggyBac vector in subsequent TAPE-seq experiments.

Next, we tested various lengths of the probe sequence, ranging from 19 to 34 bp, as it is possible that a shorter probe sequence could result in a higher tagmentation rate. Indeed, for the on-target site of the *HEK4* (+2 G to T) pegRNA, a 19-bp probe sequence resulted in higher integration rates compared to a 34-bp probe sequence (Supplementary Figure 2e–j). However, for one of the off-target sites, the opposite trend was observed (Supplementary Figure 2h–j). We chose to use a 34-bp sequence for subsequent analyses, because the objective of TAPE-seq is the tagmentation of off-target sites and because the GUIDE-seq experiment and analysis was optimized using a 34-bp tag sequence. [Both GUIDE-seq^{1,35} and its predecessor, the anchored multiplex PCR (AMP)³⁶ method, involve a nested PCR step to ensure high specificity, which is achieved by using two unidirectional primers. When primers were optimized for 17 different targets for AMP analysis³⁷, the use of the two tandem primers yielded target priming sites ranging from 35 bp to 71 bp in length, with an average of 46 bp and a median of 44 bp. We reasoned that the reduction in length of the target priming site from 34 bp to 19 bp would eliminate the high specificity obtained with the nested PCR step in GUIDE-seq and the AMP method. Indeed, when the length of the probe sequence was reduced from 34 bp to 19 bp, the number of Nucleotide BLAST^{38,39} hits surged from 1 to approximately 4000, suggesting a 4000 times higher chance of genome-wide mis-priming, which would result in a lower signal-to-noise ratio. Because the 34-bp probe sequence used in GUIDE-seq successfully tagged on-target and off-target sites on six different occasions, we chose to use the 34-bp sequence for subsequent analyses.]

When the tagmentation rates were measured for samples incubated under optimized conditions, each with nine different pegRNAs that contained a tag sequence and that targeted different genes, tagmentations were observed at all of the targets (Fig. 1b). The tagmentation efficiencies were not directly proportional to the PE2 efficiencies, which were measured from the stable cell lines expressing PE2 and the corresponding pegRNAs without the tag sequences. We also compared the tagmentation rates of one on-target and five off-target loci that had previously been identified by nDignome-seq (Supplementary Figure 3a). Because one of the off-target loci showed a ~100% tagmentation rate, we proceeded to the next step with the aforementioned conditions for the tagmentation step.

Analysis of on-target and off-target tagmentation patterns

Next, we compared the prime editing pattern at the on-target loci for each prime-edited sample obtained using pegRNAs with the tag sequence. The addition of the tag sequence to the pegRNA results in two alternative integration scenarios. In the first case (Case 1), the 34-bp tag sequence is added without perturbing the rest of the prime editing pattern, such that if the 34-bp probe sequence is removed from this pattern, it is identical to that induced by the pegRNA without the tag. In the second case (Case 2), the tag integration perturbs the prime editing pattern, such that if the 34-bp tag is removed, it is different than that induced by the pegRNA without the tag. When the tag integration patterns at on- and off-target loci for nine different pegRNAs were analyzed with targeted deep sequencing analysis and PE-Analyzer⁴⁰, the majority of the tagmented samples corresponded to the Case 1 scenario (Fig. 1c). In addition, further analysis of Case 1 samples revealed that most of them included both the tag and the prime editing; only a small fraction was tagmented without prime editing (Supplementary Figure 3b, Supplementary Data 2). From these results, we concluded that the presence of the tag sequence

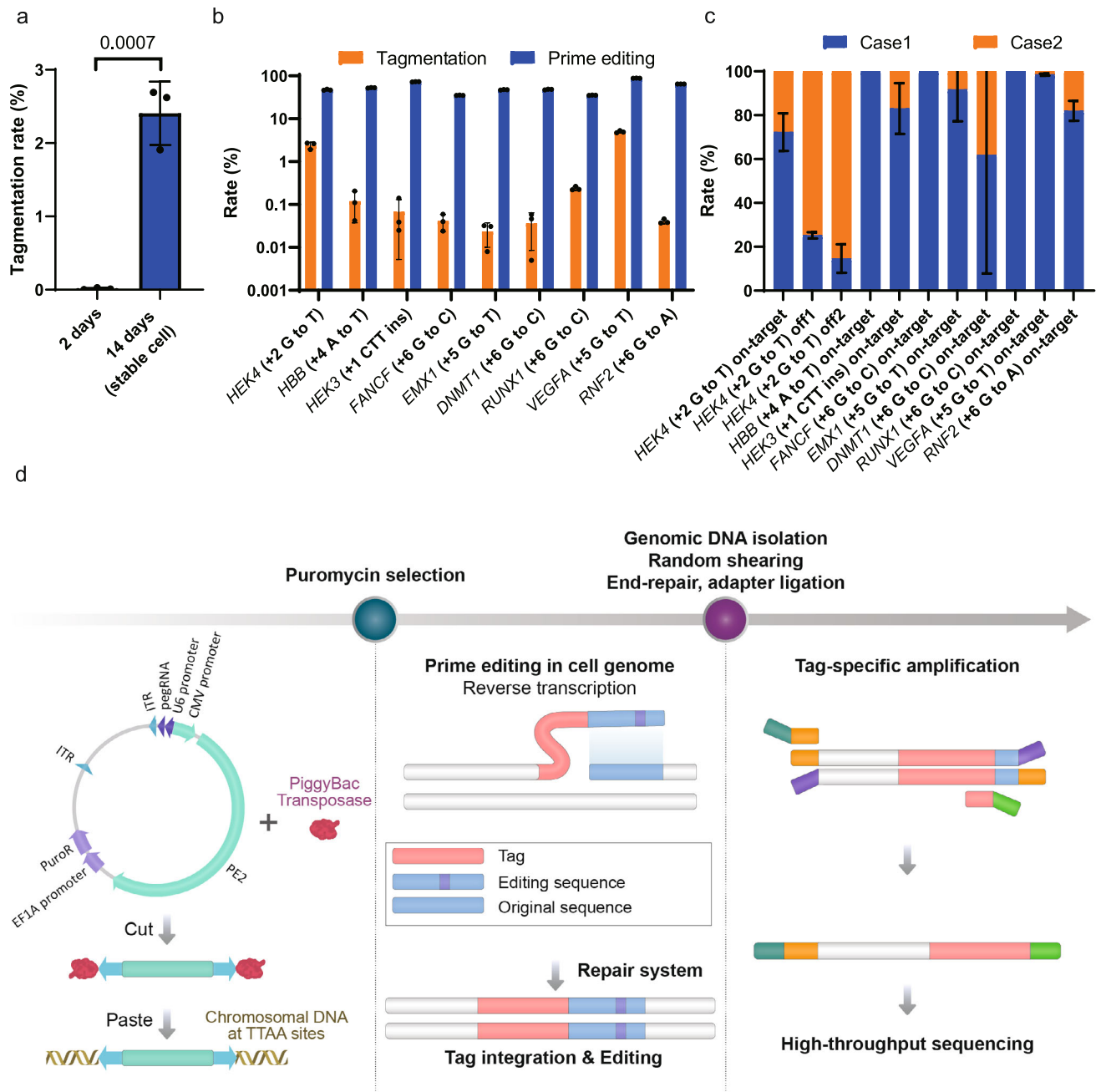


Fig. 1 | Optimization of the tagmentation rate and analysis of the tagmentation pattern. **a** Tagmentation rates after 2 or 14 days; two-sided unpaired student *t* test. **b** Tagmentation and prime editing rates at nine different on-target sites. **c** Ratios of

Case 1 and Case 2 editing determined by targeted deep sequencing and PE Analyzer. **d** Schematic of the TAPE-seq method. The bars represent the mean. Error bars indicate standard deviation (*n* = 3 independent transfections).

has minimal effect on the prime editing pattern at on- and off-target sites.

Analysis of tagged genomic DNA to predict the genome-wide off-target effects of PE2

We purified the tagged genomic DNA and processed it using the protocol from GUIDE-seq^{1,35} for tag-specific amplification to produce a TAPE-seq library. In the previous analysis³¹, *HEK4*-targeted pegRNAs were associated with a large number of validated off-target sites compared to pegRNAs targeting other sites. We, therefore, optimized the TAPE-seq protocol using the *HEK4* site as a case study. First, we analyzed the TAPE-seq library made from the same genomic DNA pool, produced after cells were transfected with plasmids encoding PE2 and

the *HEK4* (+2 G to T) pegRNA, with MiSeq and HiSeq, and summarized the results in a Venn diagram (Supplementary Figure 4a). HiSeq (53,771,178 reads) did not reveal more candidate off-target sites, indicating that the read number for MiSeq (2,251,379 reads) is large enough for this analysis. However, the HiSeq and MiSeq results each missed some of the other's predicted off-target sites even when the TAPE-seq library made from the same genomic DNA sample was used. We speculate that due to low tagmentation efficiencies at off-target sites, the tag-specific amplifications of some of these off-target sites were not replicated in each run. We also compared the TAPE-seq results for the *HEK4* (+2 G to T) and *HEK4* (+3 ATT ins) pegRNAs, which were also previously analyzed. The results, summarized in a Venn diagram, show that TAPE-seq analysis of the *HEK4* (+2 G to T) pegRNA-

treated sample correctly predicted a validated off-target site of the *HEK4* (+3 ATT ins) pegRNA, which was missed by TAPE-seq analysis for the *HEK4* (+3 ATT ins) pegRNA (Supplementary Figure 4b), and vice versa (Supplementary Figure 4c). We speculate that the off-target profile of the *HEK4* (+2 G to T) pegRNA is similar to that of the *HEK4* (+3 ATT ins) pegRNA, so the difference between the TAPE-seq results for these two samples may be caused by the same replication issue found for the HiSeq and MiSeq samples following treatment with the *HEK4* (+2 G to T) pegRNA: the low tagmentation rate of off-target sites. We, therefore, combined all three sets of TAPE-seq results for the *HEK4* pegRNA for later analysis and labeled the combined results *HEK4* [combined] for simplicity.

Comparisons of TAPE-seq prediction results with those from GUIDE-seq and nDigenome-seq

TAPE-seq analyses were performed with the optimized protocol for ten different pegRNAs (Supplementary Data 3 and 4) and compared with the previous predictions made by GUIDE-seq and nDigenome-seq. Validation experiments were performed for all of the off-target candidates (referred to herein as off1, off2, etc.) predicted by TAPE-seq using a HEK293T cell line that stably expressed PE2 and the appropriate pegRNA (Supplementary Data 5). Some of the targets identified by nDigenome-seq that were determined to be false positives were validated in our experiment (Supplementary Data 6). This result may be due to the prolonged incubation period in our protocol (4 weeks) compared to the transient transfection used in the nDigenome-seq validation experiments (96 h). We also performed validation experiments for the validated target loci identified by the methods from previous papers even if they were missed by TAPE-seq. Venn diagrams showing the overlap of off-target sites predicted by TAPE-seq, GUIDE-seq, and nDigenome-seq and validated loci were constructed (Fig. 2a–j). TAPE-seq predicted far fewer off-target sites than GUIDE-seq and nDigenome-seq. However, TAPE-seq also missed fewer off-target sites than either of the other methods (Fig. 2k), suggesting that TAPE-seq predictions show high accuracy.

TAPE-seq analysis using PE2 and PE4 in different cell lines

Later versions of PEs have been developed and have been reported to show higher prime editing efficiencies than earlier versions. We reasoned that TAPE-seq could be further optimized by using Prime Editor 4 (PE4), which is a modified version of PE2 that exhibits higher prime editing efficiencies due to the inclusion of a plasmid that encodes dominant negative MLH1 to inhibit mismatch repair⁴¹. It is possible that the higher efficiency of PE4 would also lead to a higher number of off-target candidates compared to that seen with PE2. In addition, we wanted to check whether performing TAPE-seq in different cell lines would produce better predictions. To this end, we performed TAPE-seq using PE2 and PE4 in HEK293T, HeLa, and K562 cells (Supplementary Data 3). No significant differences were seen in the tagmentation rates at the on-target and one of the off-target loci of the *HEK4* (+2G to T) pegRNA in the three cell lines (Supplementary Figure 5a, b). We validated the predicted off-target locus via targeted deep sequencing. Venn diagrams show that PE4 missed more validated off-target sites than PE2 (Fig. 3a–f, g). In addition, TAPE-seq performed in HEK293T cells missed fewer validated off-target sites compared to analysis in the other two cell lines (Fig. 3h).

Next, we determined whether the candidate off-target sites in the HEK293T, HeLa, and K562 cell lines could be validated and compared the validation results with the TAPE-seq predictions for the respective cell lines using Venn diagrams (Supplementary Figure 5c–h). Far fewer validated off-target sites were found in HeLa and K562 cells compared to HEK293T cells; furthermore, only a few off-target sites were missed by TAPE-seq in each cell line (Supplementary Figure 5i). We speculate that the TAPE-seq predictions in each cell line are accurate. In addition, TAPE-seq

predictions made using the HEK293T cell line identified all valid off-target sites for HeLa and K562 cells. Therefore, we excluded PE4 and used HEK293T cells for all subsequent experiments.

TAPE-seq analysis using PE2-nuclease and PEmax-nuclease with engineered pegRNAs

Prime editor nucleases, which contain wild-type Cas9 nuclease instead of Cas9 nickase, have also been reported to exhibit higher prime editing efficiencies than PE2⁴². However, these PEs also result in a higher indel ratio as a side effect. We reasoned that the use of these prime editor nucleases would result in higher tagmentation rates at off-target loci, increasing the success rate of TAPE-seq for identifying novel off-target loci.

Optimization experiments showed that transient transfections were sufficient for PE2-nuclease⁴² and PEmax-nuclease with engineered pegRNAs (epegRNAs)^{41,43}, confirmed by the high tagmentation rates compared to that found with TAPE-seq performed with PE2 (Fig. 4a). [Although the on-target tagmentation rates of PE2-nuclease and PEmax-nuclease with epegRNAs were significantly higher than that of PE2 (Fig. 4a), there were only 1110 on-target TAPE-seq reads for PE2-nuclease and 906 for PEmax-nuclease with epegRNAs, compared to 62565 for the PE2 (2 week) sample (Supplementary Data 3). Nevertheless, PE2-nuclease and PEmax-nuclease with epegRNAs led to the identification of 30 and 27 candidates, respectively, compared to 8 candidates identified in the PE2 (2-week) sample.]

We undertook TAPE-seq with ten different pegRNAs using PE2-nuclease and PEmax-nuclease with epegRNAs and compared the results with that of TAPE-seq using PE2 with Venn diagrams (Fig. 4b–f, Supplementary Figure 6a–e). Venn diagrams were also used to compare predictions from TAPE-seq using PEmax-nuclease with epegRNAs with those from GUIDE-seq and nDigenome-seq (Fig. 4g–k, Supplementary Figure 6g–j). To summarize the results, we compared the miss rates (defined as the number of missed validated off-target sites divided by the total number of validated off-target sites) of TAPE-seq performed using PE2, PE2-nuclease, and PEmax-nuclease with epegRNAs to those of GUIDE-seq and nDigenome-seq for ten different pegRNAs (Supplementary Data 7, Fig. 4l). TAPE-seq using PEmax-nuclease with epegRNAs showed the lowest miss rate. It should be noted that the number of missed validated off-target sites for PE2 has increased compared to the results shown in Fig. 2k, because TAPE-seqs performed using PE2-nuclease and PEmax-nuclease with epegRNAs have identified novel validated off-target sites.

TAPE-seq analysis using PEmax-nuclease with epegRNAs shows the highest area under the receiver operating characteristic (ROC) curve

A ROC curve is a plot that shows the diagnostic ability of a binary classifier. We reasoned that by constructing ROC curves for TAPE-seq analyses using PE2, PE2-nuclease, and PEmax-nuclease with epegRNAs to compare with those for GUIDE-seq and nDigenome-seq, we could quantitatively compare the diagnostic ability of TAPE-seq's metric (copy number) with that of GUIDE-seq (copy number) and nDigenome-seq (DNA cleavage score). When the area under the ROC curves were compared to each other (Fig. 5a–e, Supplementary Figure 7a–e), TAPE-seq using PEmax-nuclease with epegRNAs showed the highest value (Fig. 5f). This result suggests that the TAPE-seq metric shows superior diagnostic ability when compared to that of GUIDE-seq and nDigenome-seq in predicting off-target sites.

Editing patterns at validated off-target sites

The editing patterns at all of the validated off-target sites were analyzed by comparing targeted deep sequencing results for the *HEK4* (+2 G to T), *HEK4* (+3 TAA ins), *HBB* (+4 A to T), *DNMT1* (+6 G to C), and *VEGFA* (+5 G to T) pegRNAs used with PE2, PE2-nuclease, and PEmax-

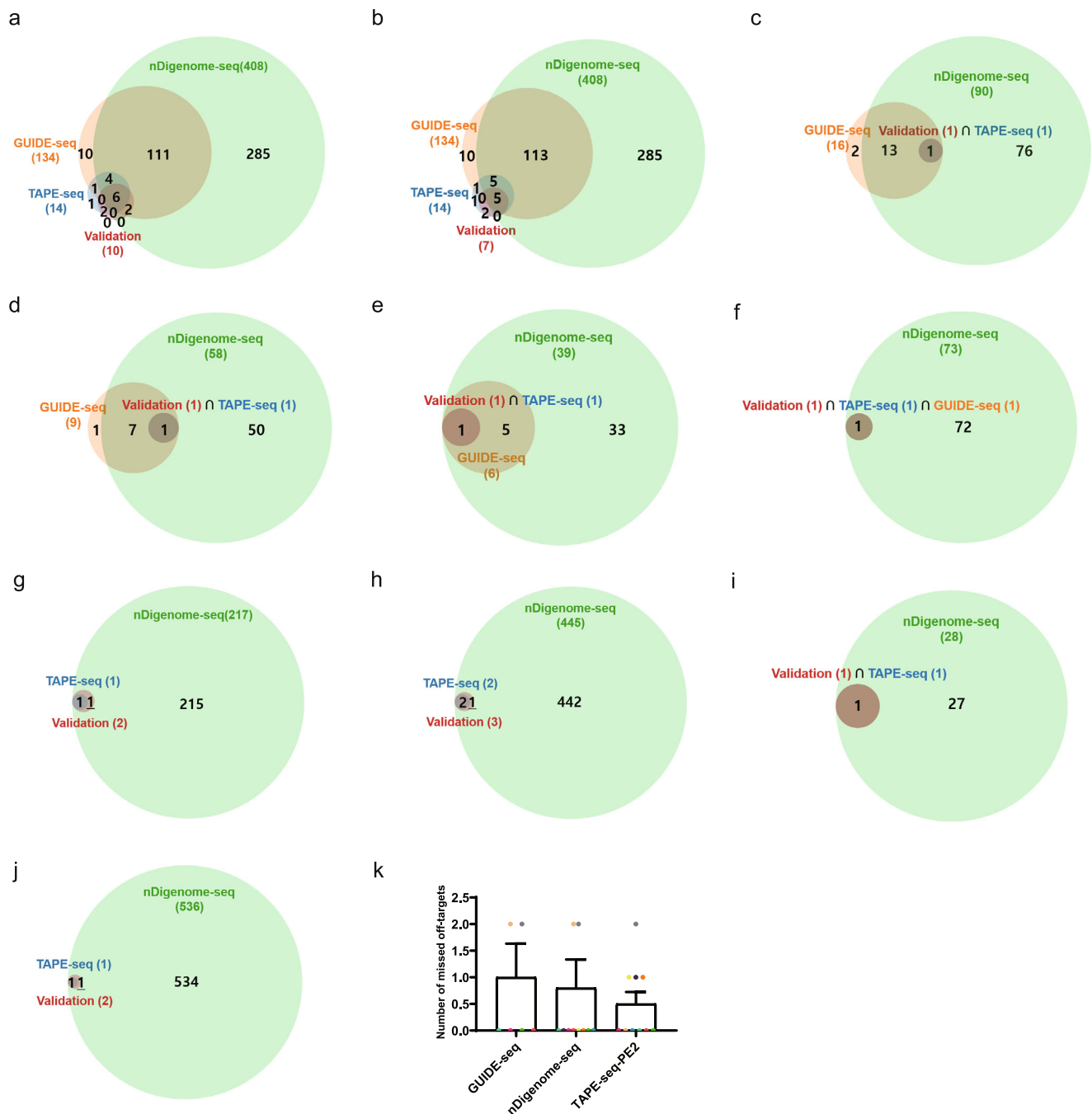


Fig. 2 | Comparisons of TAPE-seq prediction results with those from GUIDE-seq and nDigenome-seq. a–j Venn diagrams showing the number of and overlap between off-target loci predicted by nDigenome-seq, GUIDE-seq, and TAPE-seq and validated off-target sites for **a** *HEK4* (+2 G to T), **b** *HEK4* (+3 TAA ins), **c** *EMX1* (+5 G to T), **d** *FANCF* (+6 G to C), **e** *HEK3* (+1 CTT ins), **f** *RNF2* (+6 G to A), **g** *DNMT1* (+6 G to C), **h** *HBB* (+4 A to T), **i** *RUNX1* (+6 G to C), and **j** *VEGFA* (+5 G to T) pegRNAs. **k** The

number of off-target sites missed by nDigenome-seq ($n = 10$ independent experiments each represented by different color), GUIDE-seq ($n = 6$ independent experiments each represented by different color), and TAPE-seq ($n = 10$ independent experiments each represented by different color). Some numbers have been underlined to differentiate neighboring numbers (**g**, **h**, **j**). The bars represent the mean. Error bars indicate standard deviation (**k**).

nuclease with epegRNAs in HEK293T, HeLa, and K562 cells (Supplementary Data 9). At the *HEK4*-off3 site predicted by TAPE-seq, only the *HEK4* (+3 TAA ins) pegRNA induced editing (Fig. 6a), whereas for the off-target sites *HEK4*-off7, *HEK4*-off10, and *HEK4*-off22 predicted by TAPE-seq, only the *HEK4* (+2 G to T) pegRNA gave rise to off-target effects (Fig. 6b). These results suggest that off-target effects are also dependent on the RT template sequence. This phenomenon may partly explain the higher area under ROC curve for TAPE-seq compared to GUIDE-seq or nDigenome-seq, as these two methods are performed with single guide RNAs lacking a RT template sequence.

Mismatch analysis by region

Next, we tabulated mismatch numbers in the PBS, RT template, and spacer regions in the pegRNAs for the on-target and off-target sites and listed them together with validation results (Supplementary Data 8). ROC curves were constructed using the number of mismatches instead of the copy number as the metric to predict validation results as a binary classification (Fig. 7a–i; *RNF2* was excluded as it had only one sample). In most cases, the area under the ROC curve for mismatches in the RT template region was higher than that for mismatches in the PBS and target regions (Fig. 7j). In addition, when the mismatch rates of

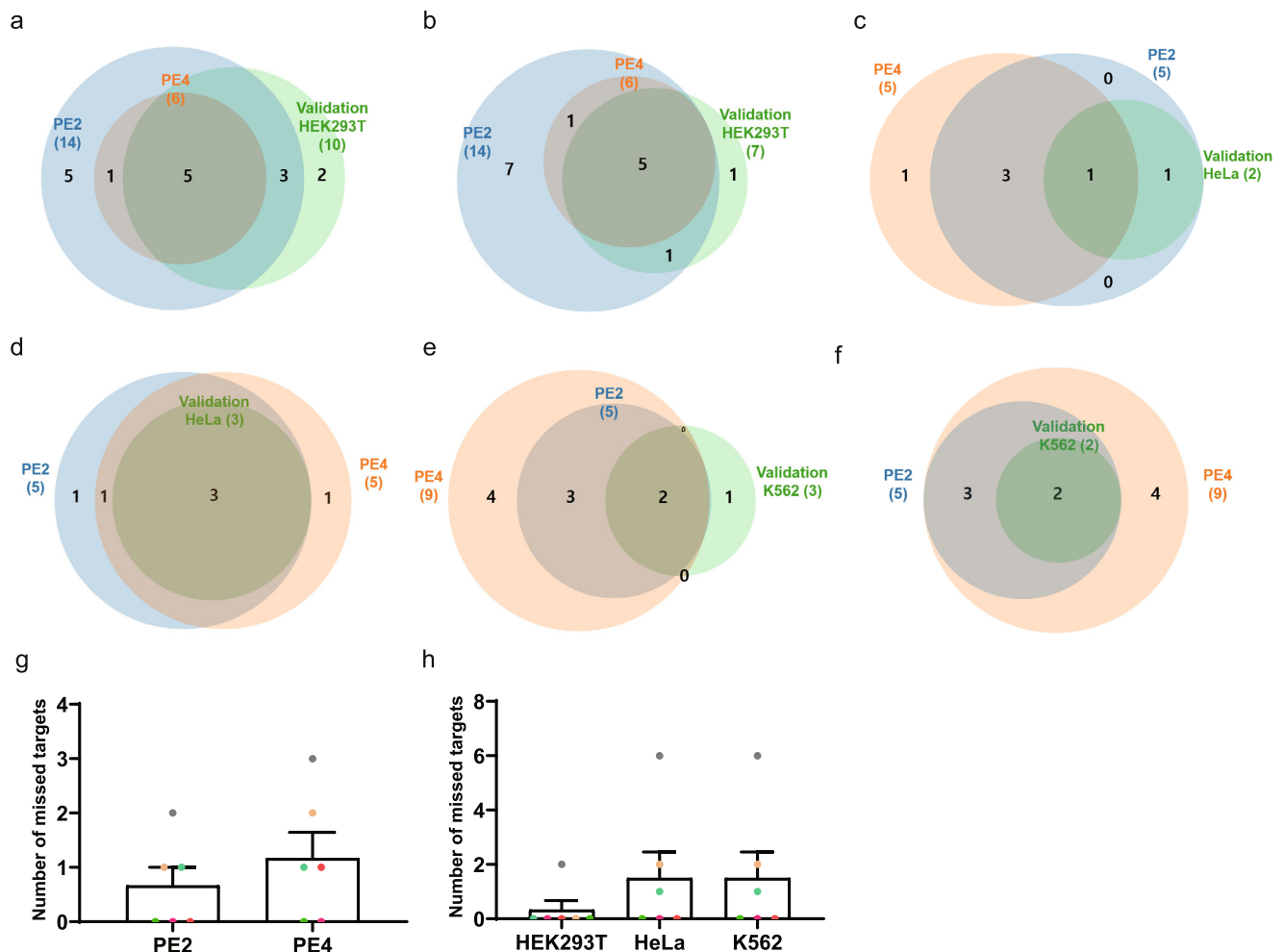


Fig. 3 | TAPE-seq analysis using PE2 and PE4 in different cell lines. a–f Venn diagrams showing the number of and overlap between validated off-target sites in PE2-transfected cell lines and the off-target sites predicted by TAPE-seq using PE2 or PE4 with the following pegRNAs and cell lines: **a** *HEK4* (+2 G to T) in HEK293T cells, **b** *HEK4* (+3 TAA ins) in HEK293T cells, **c** *HEK4* (+2 G to T) in HeLa cells, **d** *HEK4* (+3 TAA ins) in HeLa cells, **e** *HEK4* (+2 G to T) in K562 cells, and **f** *HEK4*

(+3 TAA ins) in K562 cells. **g** The number of off-target sites missed by the different versions of PE (PE2 or PE4) used for TAPE-seq. **h** The number of missed off-target sites in the different cell lines (HEK293T, HeLa, and K562) used for TAPE-seq. The bars represent the mean. Error bars indicate standard deviation ($n = 6$ independent experiments each represented by different color).

false and validated targets in the PBS, target, and RT template regions were compared, rates for false were significantly higher than that for validated in the target and RT template regions, not in the PBS region (Fig. 7k). All in all, RT template mismatches seem to show as much diagnostic ability as target mismatches for predicting the validity of potential off-target sites. Unlike TAPE-seq, GUIDE-seq and nDigenome-seq do not involve RT in their protocols, limiting their ability to accommodate the molecular mechanism of RT in their off-target prediction processes. We speculate that TAPE-seq's higher diagnostic ability originates from its recruitment of RT and subsequent elimination of false positive off-target sites.

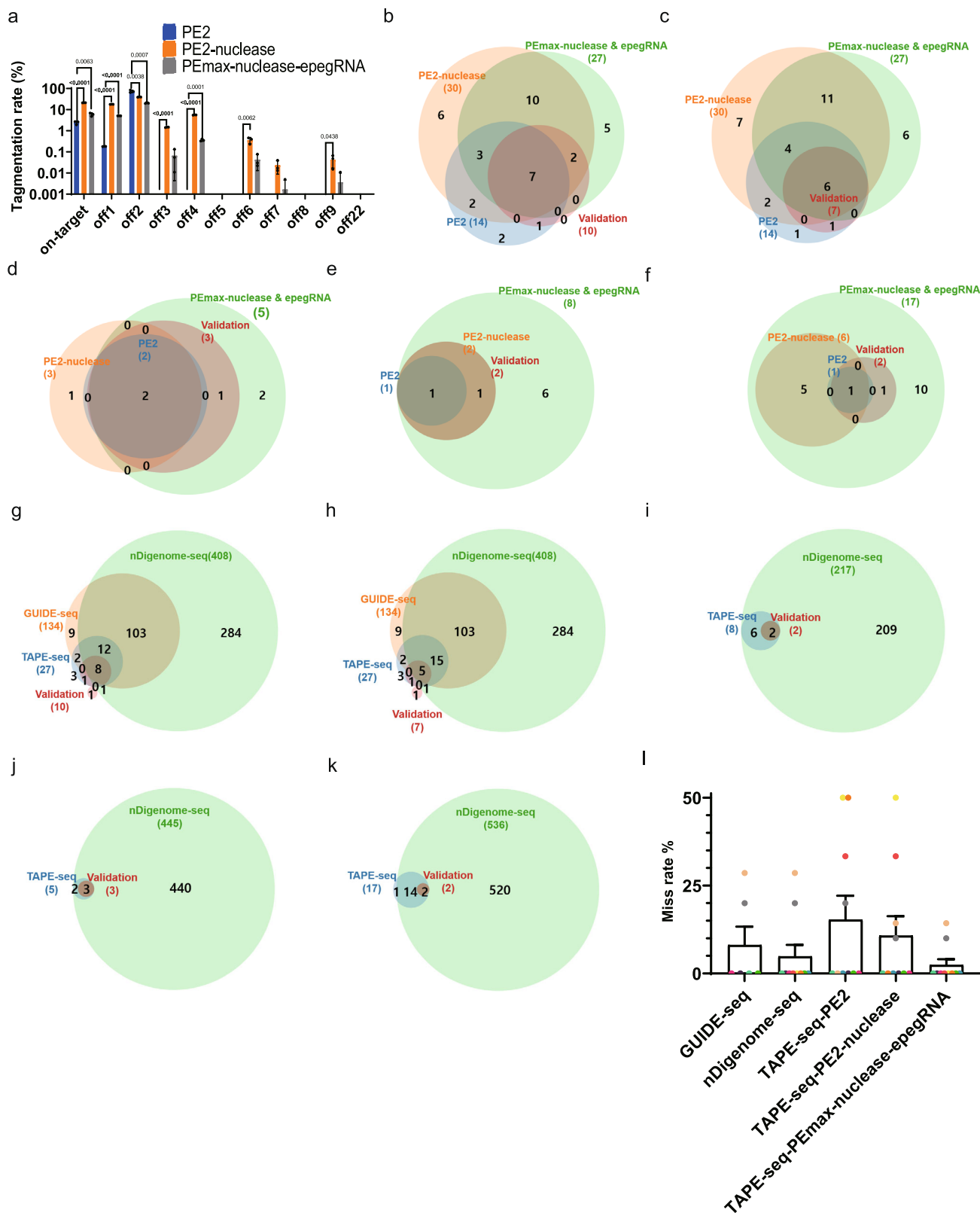
Discussion

In this paper, we describe the development of TAPE-seq, which shows high predictive power for genome-wide off-target effects of PE2. Similar to results described in the previous papers^{1–3}, TAPE-seq also identified fewer off-target loci for PE2 compared to those associated with DSB-inducing Cas9 targeted to the same sites. Recently, various techniques have been developed^{41,42,44–47} to increase the efficiency of PE2. Some of these techniques^{41–43} have been applied to the TAPE-seq protocol to increase the tagmentation rate, which also increased the sensitivity of TAPE-seq for identifying novel off-target loci that were

missed by previous methods. It is anticipated that increasing the sensitivity of TAPE-seq will result in the identification of more, previously missed, off-target loci. In addition, the tagmentation condition could also be further optimized to increase the sensitivity of TAPE-seq.

The potential advantages of TAPE-seq include that it is an unbiased cell-based method that can detect cell type-specific prime editing events with high validation rates, low miss rates, and high areas under the ROC curve. This method directly measures PE genome editing activities by accommodating the RT mechanism, unlike other methods such as nDigenome-seq and GUIDE-seq that only give indirect measures of the nickase or DSB activities of Cas9. In addition, whereas the biggest limitation of GUIDE-seq is the necessity for transfecting a double-stranded oligodeoxynucleotide (dsODN) tag, which could be toxic to some intolerant cells or not possible in an animal model, the TAPE-seq tag sequence is included in the pegRNA itself, so that toxicities due to dsODNs are irrelevant; furthermore, in vivo delivery of TAPE-seq vectors is also possible.

There are several potential limitations of TAPE-seq. First, performing TAPE-seq in a 'surrogate' cell line to predict off-target loci in other cell types could result in a high off-target prediction miss rate due to cell type-specific activities. Additionally, because the pegRNA is single-stranded, the tag sequence could form a secondary structure with the



neighboring RT or PBS sequence. Such an occurrence could be detected by a low on-target tagmentation rate, in which case the tag sequence should be modified before the definitive TAPE-seq analysis is performed; the reverse complementary sequence of 34-bp tag sequence could be used or another tag sequence that does not form the secondary structure could be designed, a process that could be assisted by prediction tools such as Vienna2.0⁴⁸, which is used to engineer pegRNA designs⁴³.

For recently developed CRISPR-based therapeutics like EDIT-101²¹ and NTLA-2001²⁰, off-target prediction results from cell-based, in vitro, and in silico methods were combined for the Initiation of New Drug (IND) applications. We expect that as more PE-based therapeutics are developed^{44,49-52}, TAPE-seq will become one of the powerful cell-based methods for studying the safety of PE-based drugs before clinical trials.

Fig. 4 | TAPE-seq analysis using PE2-nuclease and PEmax-nuclease with engineered pegRNAs. **a** Tagmentation rates at the *HEK4* on-target site and various associated off-target sites after TAPE-seq using PE2, PE2-nuclease, or PEmax-nuclease with epegRNAs was performed; two-sided unpaired student *t* test. ($n = 3$ independent transfection). **b–f** Venn diagrams showing the number of and overlap between off-target loci predicted by TAPE-seq using PE2, PE2-nuclease, or PEmax-nuclease with epegRNAs, and validated off-target sites for the **b** *HEK4* (+2 G to T), **c** *HEK4* (+3 TAA ins), **d** *HBB* (+4 A to T), **e** *DNMT1* (+6 G to C), and **f** *VEGFA* (+5 G to T)

pegRNAs. **g–k** Venn diagrams showing the number of and overlap between off-target loci predicted by TAPE-seq using PEmax-nuclease with epegRNAs, GUIDE-seq, and nDigenome-seq and validated off-target sites for the **g** *HEK4* (+2 G to T), **h** *HEK4* (+3 TAA ins), **i** *HBB* (+4 A to T), **j** *DNMT1* (+6 G to C), and **k** *VEGFA* (+5 G to T) pegRNAs. **l** Miss rates of the five different off-target prediction methods ($n = 6$ independent experiments for GUIDE-seq and $n = 10$ independent experiments for the rest of the methods each represented by different color). The bars represent the mean. Error bars indicate standard deviation.

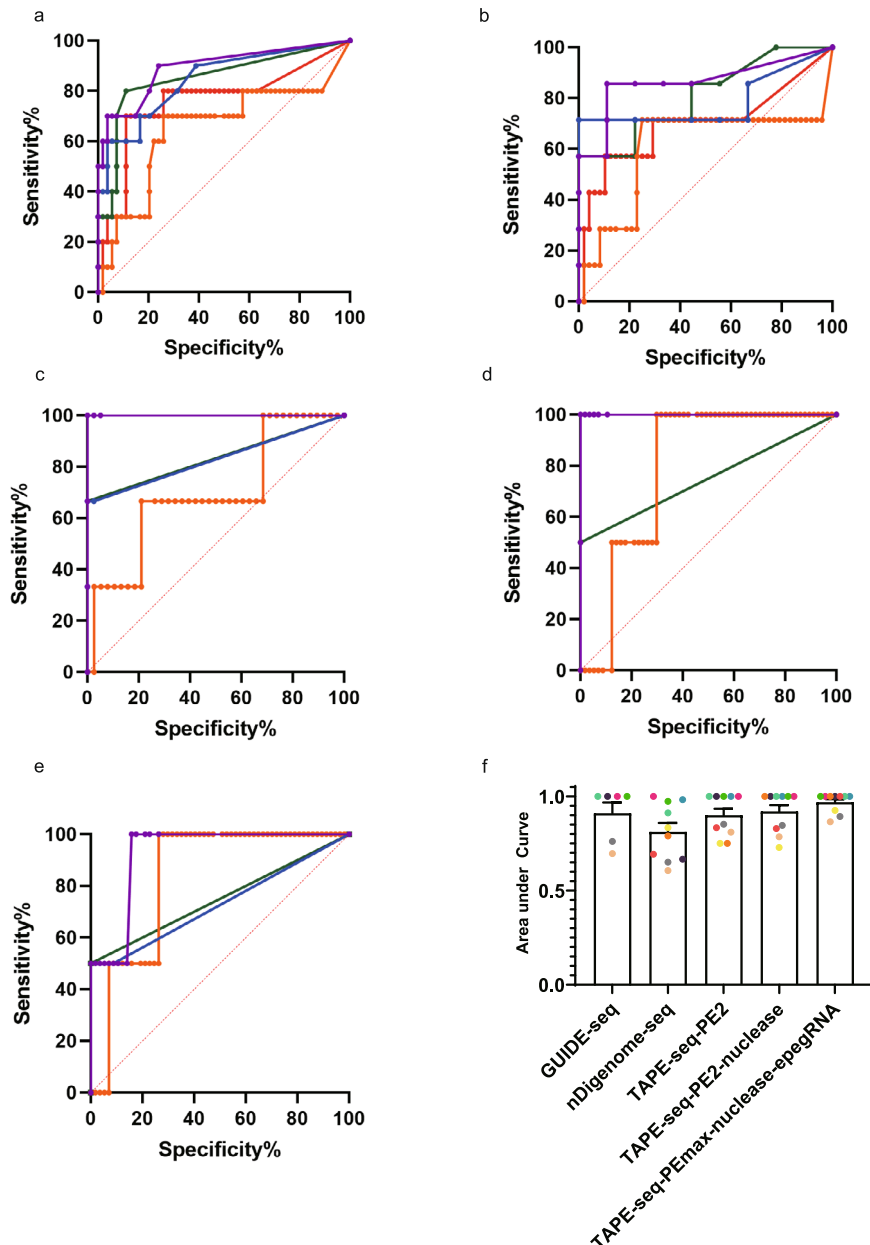


Fig. 5 | ROC curve analysis. **a–e** ROC curves for GUIDE-seq (red), nDigenome-seq (orange), and TAPE-seq using PE2 (green), PE2-nuclease (blue), and PEmax-nuclease with epegRNAs (purple) for the **a** *HEK4* (+2 G to T), **b** *HEK4* (+3 TAA ins), **c** *HBB* (+4 A to T), **d** *DNMT1* (+6 G to C), and **e** *HEK3* (+1 CTT ins) pegRNAs. **f** Area under ROC

curves ($n = 6$ independent experiments for GUIDE-seq and $n = 10$ independent experiments for the rest of the methods each represented by different color). The bars represent the mean. Error bars indicate standard deviation.

Methods

Plasmid construction

The sgRNA-expressing plasmid pRG2 (addgene #104174) was modified to create a pegRNA-expressing plasmid (pRG2-pegRNA) by Gibson assembly following cleavage at the BsmBI restriction site at the 3' terminus of the sgRNA scaffold. The plasmid was

modified to contain a BsaI site (for incorporation of a spacer sequence) and a BsmBI site (for incorporation of the pegRNA 3' extension). To create the piggyBac PE2 all-in-one plasmid (pAllin1-PE2), the piggyBac PE2-expressing plasmid DNA was synthesized and cloned to make a vector (piggy-PE2). It was then digested with Mlu I. The pegRNA-encoding sequence was amplified from

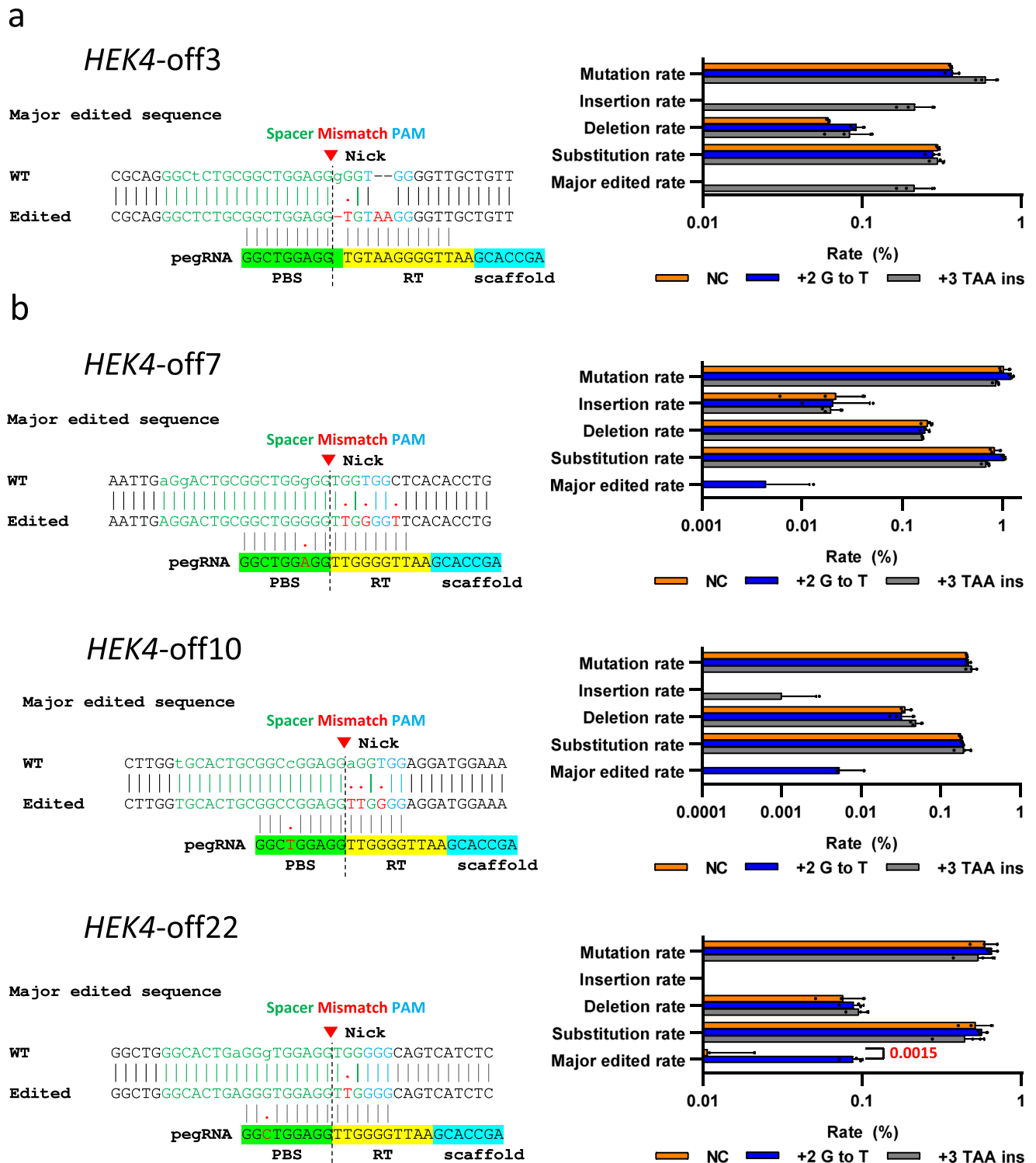


Fig. 6 | Off-target prime editing patterns. a Editing patterns induced by the *HEK4* (+3 TAA ins) pegRNA at the *HEK4*-off3 site predicted by TAPE-seq. **b** Editing patterns induced by the *HEK4* (+2 G to T) pegRNA at the *HEK4*-off7, *HEK4*-off10, and *HEK4*-off22 sites predicted by TAPE-seq; two-sided unpaired student *t* test. NC negative

control. The bars represent the mean. Error bars indicate standard deviation ($n = 3$ independent transfection). Small letters indicate mismatches compared to the pegRNA. Major edited rate corresponds to the frequency of the 'Edited' sequence.

pRG2-pegRNA by PCR to generate the insert fragment, which was cloned into the digested piggyBac PE2 vector via Gibson assembly. Other PE all-in-one plasmids (pAllin1-PE4, pAllin1-PE2-nuclease, and pAllin1-PEmax-nuclease) were constructed using the same procedure that was used to construct pAllin1-PE2. The pRG2-epgRNA vector was constructed using the same procedure that was used to construct pRG2-pegRNA. The DNA sequences of

all of the constructed vectors (pRG2-pegRNA, pAllin1-PE2, piggy-PE2, pRG2-epgRNA, pAllin1-PE4, pAllin1-PE2-nuclease, and pAllin1-PEmax-nuclease) are available in Supplementary Data 10.

Human cell culture and transfection

HEK293T (ATCC CRL-1268), HeLa (ATCC CCL-2), and K562 (Sigma 89121407) cells were maintained in the appropriate medium

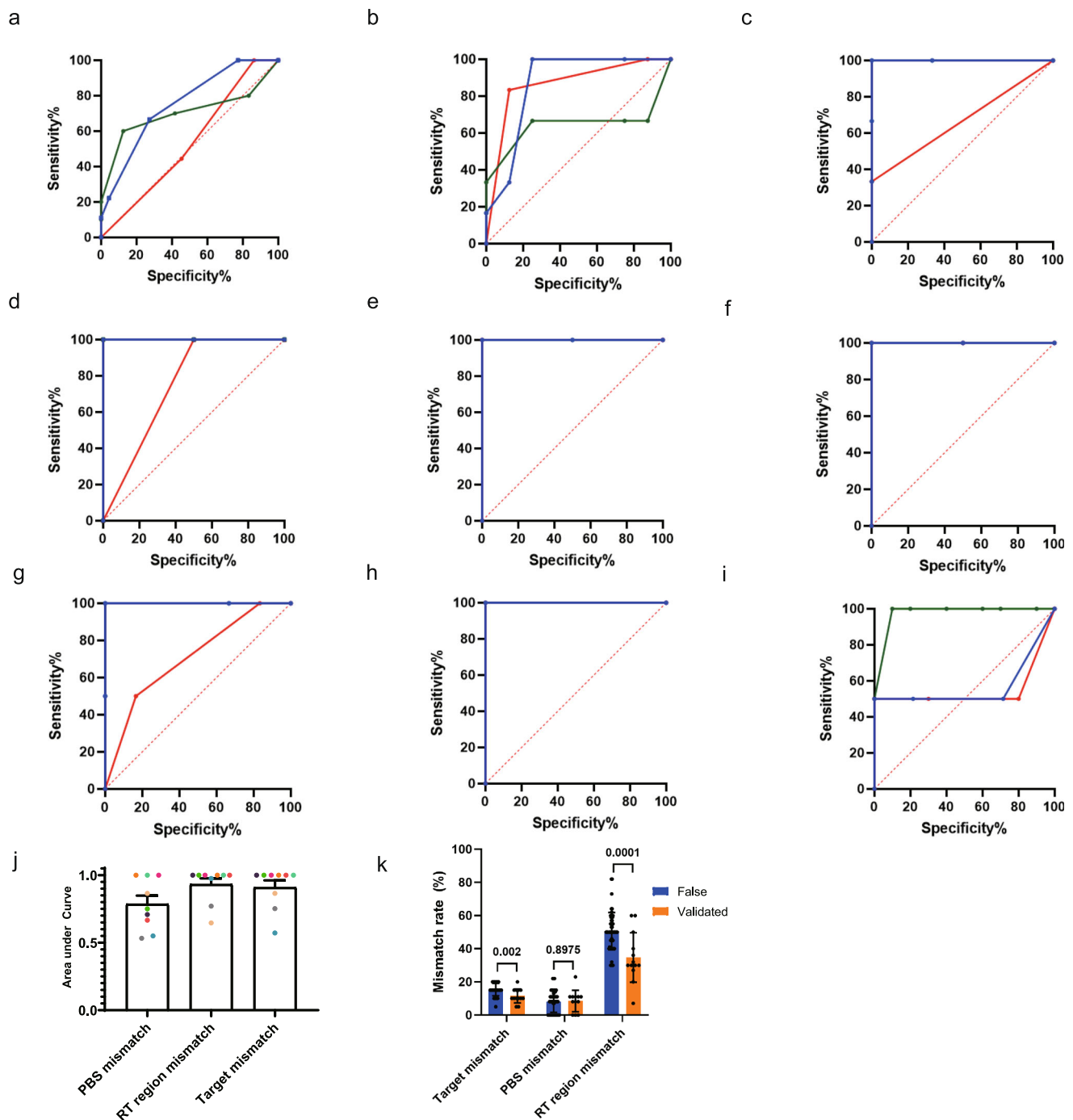


Fig. 7 | Mismatch analysis by region. **a–i** ROC curves for mismatches in the PBS (red), RT template (green), and target (blue) regions for the **a** *HEK4* (+2 G to T), **b** *HEK4* (+3 TAA ins), **c** *HBB* (+4 A to T), **d** *HEK3* (+1 CTT ins), **e** *FANCF* (+6 G to C), **f** *EMX1* (+5 G to T), **g** *DNMT1* (+6 G to C), **h** *RUNX1* (+6 G to C), and **i** *VEGFA* (+5G to T) pегRNAs. **j** Area under the ROC curve for the nine different pегRNAs ($n = 10$ independent experiments each represented by different color). **k** Mismatch rates for sites predicted by TAPE-seq, false positive sites ($n = 54$ independent target loci

for Target mismatch and RT region mismatch and $n = 47$ target loci for PBS mismatch; 7 loci with DNA/RNA bulge for PBS mismatch were excluded from the analysis), and validated sites ($n = 13$ independent target loci for Target mismatch and RT region mismatch and $n = 12$ target loci for PBS mismatch; 1 locus with DNA/RNA bulge for PBS mismatch was excluded from the analysis); two-sided unpaired student t test. The bars represent the mean. Error bars indicate standard deviation.

[Dulbecco's Modified Eagle Medium (DMEM) for HEK293T and HeLa cell lines, Roswell Park Memorial Institute 1640 Medium (RPMI 1640) for the K562 cell line] containing 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37°C in the presence of 5% CO₂. 1×10^5 HEK293T cells or 4×10^4 HeLa cells were seeded in a 24-well plate in preparation for transfection. One day after seeding, cells were transfected with an adequate amount of plasmid (see below) and 2 μ l Lipofectamine 2000 (Thermo Fisher

Scientific). [For transient PE2 expression, 500 ng piggy-PE2 and 500 ng pRG2-pegRNA; for stable PE2 expression, 850 ng pAllin1-PE2 and 150 ng piggyBac Transposase Expression Vector (System Biosciences); for stable PE4 expression, 880 ng pAllin1-PE4 and 120 ng piggyBac Transposase Expression Vector; for stable PE2-EGFP expression, 865 ng pAllin1-PE2-EGFP and 135 ng piggyBac Transposase Expression Vector; for transient PE2-nuclease expression, 1000 ng pAllin1-PE2-nuclease; and for transient

PEmax-nuclease and epegRNA expression, 1000 ng pAllin1-PEmax-nuclease-epegRNA were used.] The transposon and piggybac plasmids were used at about a 2.5:1 transposon:transposase plasmid molar ratio. 1×10^5 K562 cells were electroporated with the above-mentioned quantities of plasmid via a Neon transfection system (electroporation conditions: 1450 V, 10 ms, 3 pulses). One day after transfection (or electroporation), antibiotic selection was conducted using puromycin (InvivoGen) at a concentration of 2 mg/ml. Puromycin selection was continued for 2 weeks [for TAPE-seq and fluorescence-activated cell sorting (FACS)], 4 weeks (for targeted deep sequencing), or 2 days (for TAPE-seq using PE2-nuclease or PEmax-nuclease; after puromycin selection, cells were cultured for an additional 4 days in normal media). Genomic DNA was purified with a Blood Genomic DNA Extraction Mini Kit (Favorgen) following the manufacturer's instructions.

TAPE-seq

A full description of the TAPE-seq method can be found in Supplementary Note 1. Genomic DNA was sheared with a Covaris M220 instrument to an average length of 325 bp and isolated with $1 \times$ AMPure XP beads (Beckman coulter). Using an NEBNext[®] Ultra[™] II DNA Library Prep Kit for Illumina (NEB), a next-generation sequencing (NGS) library was prepared according to the manufacturer's protocol, with slight modifications to certain reaction times (adaptor ligation, 1 h; treatment with Uracil-Specific Excision Reagent, 30 min). Using tag- and adaptor-specific primers, tag-specific library amplification was performed according to previously described GUIDE-seq methods^{1,2}. The amplified library was analyzed with a MiSeq or HiSeq platform (Illumina).

Paired-end FASTQ files were processed using the following steps: 1. Sequences including the tag were collected using the BBDuk program (Tag sequences for sense library(+), 5'-GTTTAATTGAGTTGTCATATGT-3' and 5'-ACATATGACAACCTCAATTAAAC; Tag sequences for antisense library(-), 5'-TTGAGTTGTCATATGTTAATAACGGTA-3' and 5'-TACCGTTATTAACATATGACAACCTCAA-3'). 2. Filtered FASTQ files were mapped to the reference genome (hg19) and the read depth was calculated using BWA, Picard tools, and SAMtools programs. 3. Off-target candidates (containing up to 4 mismatches and/or 2 bulges relative to the on-target site) were identified using Cas-OFFinder³ (<http://www.rgenome.net>). 4. The read depths at the sites identified by Cas-OFFinder were extracted from the region spanning -150 bp to +150 bp around the site using an in-house script. 5. Short-mapped sequences (less than 30 bp in length) and false tagmentation sequences (in which tagmentation occurred outside of the PE nick site) were excluded.

Targeted deep sequencing and validation of off-target sites

Following expression of PE2 and the pegRNA, target sites were analyzed by targeted deep sequencing. Deep sequencing libraries were generated by PCR. TruSeq HT Dual Index primers were used to label each sample. Pooled libraries were subjected to paired-end sequencing using MiSeq (Illumina). Paired-end FASTQ files were analyzed with PE-Analyzer (<http://www.rgenome.net>). Candidates that satisfied the following two conditions were designated as 'validated off-targets': 1. The frequency of at least one of the following events (mutation, insertion, deletion, substitution, or major editing) was higher than that in the wild-type sample. 2. A mutated sequence that could only be generated by prime editing (the major edited sequence) was present. To overcome the detection limits of NGS and problems created by PCR error, validation experiments were conducted using cells in which PE2 had been stably expressed for 4 weeks, and were performed in triplicate using biologically independent genomic DNA. The validation rate was

calculated by dividing the number of validated targets by the sum of (the number of validated targets + the number of false positive targets). Targets that were not analyzed were excluded from the validation rate calculation.

Prime editing tagmentation analysis

The presence of the tag sequence (34-bp full-length tag: GTTTAATTGAGTTGTCATATGTTAATAACGGTAT, 29-bp tag: GTTAAATTGAGTTGTCATATGTTAATAAC, 24-bp tag: GTTTAATTGAGTTGTCATATGTTA, 19-bp tag GTTTAATTGAGTTGTCATA) was defined as tagmentation. PE-Analyzer (<http://www.rgenome.net>) was used to identify reads in which tagmentation occurred⁴⁰. Tagmentation Case 1 and Case 2 were distinguished by sequence analysis. After TAPE-seq reads were analyzed by NGS, only the reads that contained full-length tag sequences were selected. The tag sequences were then removed from the sequences for analysis, and the remaining sequences were compared with the NGS reads from targeted deep sequencing of the cells that had undergone prime editing with pegRNAs without the tag sequence. Case 1 means that the editing pattern after the tag sequence is removed is the same as the editing pattern generated with the pegRNA lacking the tag sequence. If that pattern could not be found, the sequence was classified as Case 2.

PiggyBac copy number analysis

To quantify the average copy number of integrated piggyBac transposons, we used a set of primers directed at the 5' inverted repeat (IR) of the piggyBac vector. The sequences of the forward and reverse primers used to amplify the 5' IR are 5'-CTAAA-TAGCGGAATCCGTC-3' and 5'-TCATTTGACTACGCGG-3', respectively. Copy numbers were calculated using standard curves generated using a mixture of untransfected HEK293T genomic DNA and the serially diluted piggyBac plasmid with a known copy number. Real-time PCR was performed using a QuantStudio 3 Real-Time PCR System (Applied Biosystems) with PowerUp SYBR Green Master Mix (Applied Biosystems).

FACS of GFP-expressing cells

Two weeks after puromycin selection, cells were washed with phosphate-buffered saline and detached from the plate with trypsin-EDTA. Cells were centrifuged at $500 \times g$ for 5 min at room temperature and resuspended in phosphate-buffered saline with 2% FBS. GFP-positive cells were isolated using an Attune NxT Acoustic Focusing Cytometer (Thermo Scientific). Attune NxT software v4.2.0 was used to analyze the raw data.

Statistics and reproducibility

10 sample sites, which were studied in the previous nDigenome-seq paper³¹, were analyzed. No data were excluded from the analyses. Results from the two-sided unpaired student t-test calculated by Prism (version 9.4.1) are shown.

Reporting summary

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

Data availability

Deep sequencing data that support the findings of this study have been deposited in NCBI Bioproject with the accession code [PRJNA802977](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA802977). Source Data is available as a Source Data file. Source data are provided with this paper.

Code availability

Codes supporting the findings of this study have been archived online (https://github.com/PhyzenInc/TAPE-seq_flanking_depth).

References

- 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).
- Liang, S. Q. et al. Genome-wide detection of CRISPR editing in vivo using GUIDE-tag. *Nat. Commun.* **13**, 437 (2022).
- Yan, W. X. et al. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).
- Crosetto, N. et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).
- Wienert, B. et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286–289 (2019).
- Wang, X. et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* **33**, 175–178 (2015).
- Chiarle, R. et al. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119 (2011).
- Kim, H. S. et al. CReVIS-Seq: A highly accurate and multiplexable method for genome-wide mapping of lentiviral integration sites. *Mol. Ther. Methods Clin. Dev.* **20**, 792–800 (2021).
- Breton, C., Clark, P. M., Wang, L., Greig, J. A. & Wilson, J. M. ITR-Seq, a next-generation sequencing assay, identifies genome-wide DNA editing sites in vivo following adeno-associated viral vector-mediated genome editing. *BMC Genomics* **21**, 239 (2020).
- Huang, H. et al. Tag-seq: a convenient and scalable method for genome-wide specificity assessment of CRISPR/Cas nucleases. *Commun. Biol.* **4**, 830 (2021).
- Dobbs, F. M. et al. Precision digital mapping of endogenous and induced genomic DNA breaks by INDUCE-seq. *Nat. Commun.* **13**, 3989 (2022).
- Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).
- Kim, D. & Kim, J. S. DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. *Genome Res.* **28**, 1894–1900 (2018).
- Lazzarotto, C. R. et al. CHANGE-seq reveals genetic and epigenetic effects on CRISPR-Cas9 genome-wide activity. *Nat. Biotechnol.* **38**, 1317–1327 (2020).
- Tsai, S. Q. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).
- Cameron, P. et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods* **14**, 600–606 (2017).
- Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
- Concordet, J. P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).
- Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407 (2014).
- Gillmore, J. D. et al. CRISPR-Cas9 in vivo gene editing for transthyretin amyloidosis. *N. Engl. J. Med.* **385**, 493–502 (2021).
- Maeder, M. L. et al. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat. Med.* **25**, 229–233 (2019).
- Frangoul, H. et al. CRISPR-Cas9 gene editing for sickle cell disease and beta-thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2020).
- 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).
- 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).
- Petri, K. et al. Global-scale CRISPR gene editor specificity profiling by ONE-seq identifies population-specific, variant off-target effects. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2021.04.05.438458v1> (2021).
- Lei, Z. et al. Detect-seq reveals out-of-protospacer editing and target-strand editing by cytosine base editors. *Nat. Methods* **18**, 643–651 (2021).
- Kim, D. et al. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).
- Kim, D., Kim, D. E., Lee, G., Cho, S. I. & Kim, J. S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).
- Liang, P. et al. Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat. Commun.* **10**, 67 (2019).
- Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
- 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).
- Jin, S. et al. Genome-wide specificity of prime editors in plants. *Nat. Biotechnol.* **39**, 1292–1299 (2021).
- Kim, D., Kang, B. C. & Kim, J. S. Identifying genome-wide off-target sites of CRISPR RNA-guided nucleases and deaminases with Digenome-seq. *Nat. Protoc.* **16**, 1170–1192 (2021).
- Li, X. et al. piggyBac transposase tools for genome engineering. *Proc. Natl Acad. Sci. USA* **110**, E2279–E2287 (2013).
- Malinin, N. L. et al. Defining genome-wide CRISPR-Cas genome-editing nuclease activity with GUIDE-seq. *Nat. Protoc.* **16**, 5592–5615 (2021).
- Zheng, Z. et al. Anchored multiplex PCR for targeted next-generation sequencing. *Nat. Med.* **20**, 1479–1484 (2014).
- lafrate, A. J., Le, L. P. & Zheng, Z., Vol. US 9,487,828 B2 (The General Hospital Corporation, Boston, MA, USA; 2016).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
- Hwang, G. H. et al. PE-Designer and PE-Analyzer: web-based design and analysis tools for CRISPR prime editing. *Nucleic Acids Res.* **49**, W499–W504 (2021).
- Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652 e5629 (2021).
- Adikusuma, F. et al. Optimized nickase- and nuclease-based prime editing in human and mouse cells. *Nucleic Acids Res.* **49**, 10785–10795 (2021).
- Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2021).
- Liu, P. et al. Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat. Commun.* **12**, 2121 (2021).
- Choi, J. et al. Precise genomic deletions using paired prime editing. *Nat. Biotechnol.* **40**, 218–226 (2021).
- Lin, Q. et al. High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nat. Biotechnol.* **39**, 923–927 (2021).
- Song, M. et al. Generation of a more efficient prime editor 2 by addition of the Rad51 DNA-binding domain. *Nat. Commun.* **12**, 5617 (2021).
- Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
- Jang, H. et al. Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. *Nat. Biomed. Eng.* **6**, 181–194 (2021).

50. Kim, Y. et al. Adenine base editing and prime editing of chemically derived hepatic progenitors rescue genetic liver disease. *Cell Stem Cell* **28**, 1614–1624 e1615 (2021).
51. Schene, I. F. et al. Prime editing for functional repair in patient-derived disease models. *Nat. Commun.* **11**, 5352 (2020).
52. Petri, K. et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **40**, 189–193 (2021).

Acknowledgements

We would like to thank Drs. Nam-Hoon Kim and Hyun Oh Lee for assisting with bioinformatics analysis. This work was supported by grants from the Ministry of Food and Drug Safety (19172MFDS167) and ‘Alchemist Project’ funded by the Ministry of Trade, Industry, and Energy (grant no. 20012443) to J.K.L.

Author contributions

J.K., M.K., S.B., and A.J. performed experiments and Y. K. and J.K.L. supervised the research.

Competing interests

A patent application has been filed based on this work: Toolgen filed 10-2022-0161819 (Status: Docketed New Case, Inventor: J.K.L., J.K., M.K., A.J., Y.K.) covering TAPE-seq. S.B. declare no competing interests. J.K.L., J.K., M.K., A.J., S.B., and Y.K. are employees of ToolGen, Inc.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-35743-y>.

Correspondence and requests for materials should be addressed to Jungjoon K. Lee.

Peer review information *Nature Communications* thanks Yong-Sam Kim and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions 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