



3'-end ligation sequencing is a sensitive method to detect DNA nicks at potential sites of off-target activity induced by prime editors

Jacob Stewart-Ornstein, Matthew J. Irby, Marina K. Lilieholm, et al.

Genome Res. published online August 5, 2025

Access the most recent version at doi:[10.1101/gr.280164.124](https://doi.org/10.1101/gr.280164.124)

P<P Published online August 5, 2025 in advance of the print journal.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

© 2025 Stewart-Ornstein et al.; Published by Cold Spring Harbor Laboratory Press

Method

3'-end ligation sequencing is a sensitive method to detect DNA nicks at potential sites of off-target activity induced by prime editors

Jacob Stewart-Ornstein, Matthew J. Irby, Marina K. Lilieholm, Dylan Laprise, Maria D. Collier, Thomas Aunins, Dewi Harjanto, Aaron N. Chang, Deepak Reyon, and Jeremy S. Duffield

Prime Medicine Incorporated, Cambridge, Massachusetts 02141, USA

Gene editing makes precise changes in DNA to restore normal function or expression of genes; however, the advancement of gene editing to the clinic is limited by the potential genotoxicity of off-target editing. To comprehensively identify potential sites in the genome that may be recognized by gene editing agents, *in vitro* approaches, in which the editor is combined with human genomic DNA and sites where editing may occur are identified biochemically, are important tools. Existing biochemical approaches for off-target discovery recognize double-stranded breaks generated by nuclease-based gene editors such as SpCas9, but novel approaches are needed for new editing modalities, such as prime editing, that nick one strand of DNA. To fill this gap, we have developed 3'-end ligation sequencing (PEG-seq), which can identify prime editor-induced nicks throughout the genome on *in vitro* digested human genomic DNA to identify potential off-target sites. Here we show that PEG-seq is an important addition to the off-target detection toolkit, enabling off-target discovery for DNA nicking gene editors such as prime editors.

[Supplemental material is available for this article.]

Biochemical methods for identifying sites in genomic DNA that are recognized or acted on by a protein or protein–RNA complex form the backbone of strategies to evaluate the specificity of gene editing molecules that may be used therapeutically. The CRISPR family of enzymes are bacterial proteins that complex with RNA molecules consisting of structural motifs and a “spacer” element that can be used to search the genome for a complementary sequence. When such a target DNA sequence is identified, the enzyme becomes active and can cleave the DNA backbone at that genomic location. The search and recognition mechanism for CRISPR enzymes is highly specific to the spacer sequence of the provided gRNA, but incomplete base pairing between the guide sequence and the genomic DNA is tolerated to the point that, in rare circumstances, the enzyme can bind tightly and cut DNA where the sequence is similar, but not identical, to the RNA-encoded spacer sequence (Jinek et al. 2012). Identifying sequences in the genome that are bound and enzymatically acted on by a specific CRISPR–RNA complex relies on molecular biology protocols that can recognize and isolate the product of the CRISPR–DNA reaction—typically a double-stranded (DSB) or single-stranded break (SSB, or “nick”) in the DNA—using high-throughput sequencing to determine the sequence of those sites and map them in the genome. Since 2015, when the first of such protocols was developed (Kim et al. 2015; Tsai et al. 2015), many methods have been applied to map DSBs produced by *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9), and other CRISPR enzymes, including SITE-seq (Cameron et al. 2017), CIRCLE-seq (Tsai et al. 2017), and CHANGE-seq (Lazzarotto et al. 2020).

These methods have allowed for broad analysis of the specificity of SpCas9 gene editors and have been used to identify off-target risks of gene editing therapies.

In addition to the natural CRISPR proteins, new genome editing modalities have been engineered by converting SpCas9 into a nickase (a protein that generates SSBs by cutting either the top or bottom strand of DNA, not both) and adding additional enzymatic activities, such as deaminases (Komor et al. 2016; Nishida et al. 2016) to generate “base editors” (BEs) and reverse transcriptase (Anzalone et al. 2019) to construct “prime editors” (PEs). As with natural CRISPR enzymes, if these BEs and PEs are to be used therapeutically, it is necessary to determine their specificity. However, traditional biochemical approaches, referenced above, that recognize DSBs do not work in a straightforward way with these newer genome editors. When DNA is edited using a BE, in addition to an SSB generated by an active Cas9 HNH endonuclease domain, an altered base is introduced in the opposite strand (inosine for adenine BE [ABE] or uracil for cytosine BE [CBE]), which can be recognized by DNA repair enzymes *in vitro* and converted to DSBs at the site of BE activity. Therefore, conventional biochemical off-target recognition strategies can be used to evaluate BE off-targets at sites where both nicking and base editing activity occur (Lazzarotto et al. 2024), although these analyses potentially miss sites where only one activity occurs. PE, however, only acts on one strand of DNA through the activity of Cas9’s RuvC endonuclease domain and prevents similar strategies from being used. Therefore, new strategies are required to biochemically identify where PE may act in the genome. Recently, methods for mapping PE off-targets have been published that utilize the integration of

Corresponding author: jstewart-ornstein@primemedicine.com

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280164.124>. Freely available online through the *Genome Research* Open Access option.

© 2025 Stewart-Ornstein et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

DNA tags into the genome through the PE reverse transcriptase domain and a modified pegRNA sequence (Kwon et al. 2022; Yu et al. 2022; Liang et al. 2023). These methods are highly specific and are applicable to both biochemical and cell-based systems; however, they require custom pegRNA sequences, have potential biases in the reverse transcriptase step owing to sequence-specific targets and mismatches between the primer binding sequence and the off-target, and do not directly detect nicking events, which may be the most common form of off-target activity by PEs.

To identify where in the genome PE may produce off-targets, we set out to identify where Cas9^{H840A} nickase-induced SSBs occur in the genome, as these SSBs are the essential first step in prime editing. In general, two difficulties arise when attempting to identify SSBs: (1) biochemical ligation or isolation of a single broken end of DNA is less efficient than the capture of DSBs, and (2) SSBs are extremely common in DNA, occurring at the rate of ten to 20,000 a day in cells (Ciccia and Elledge 2010), and thus, any attempt to capture SSBs is confounded by the significant background. One effective strategy for identifying SSBs in the genome uses high-depth whole-genome sequencing of digested gDNA to identify locations that have higher rates of SSBs (Kim et al. 2020). Other approaches that take advantage of enzymatic activities, such as nick translation, are potentially efficient but have additional challenges and have not been widely used for off-target identification (Cao et al. 2020; Elacqua et al. 2021). We therefore sought to develop a direct deep sequencing strategy that would utilize single-strand ligation of sequencing adapters to efficiently capture SSBs induced by nicking endonucleases.

Results

To develop our protocol, we extracted common elements from existing SSB capture and ligation protocols such as GLOE-Seq (Sriramachandran et al. 2020), DENT-seq (Elacqua et al. 2021), Nick-seq (Cao et al. 2020), and TrAEL-seq (Kara et al. 2021). As SSBs are common in DNA, owing to oxidative damage and DNA handling, we implemented a method to block pre-existing SSBs by treating the genomic DNA with a polymerase and chain-terminating ddNTPs prior to digestion with an enzyme; this ensures that most SSBs captured will come from the tested enzyme. We then developed a ligation strategy in which a reverse-complemented Illumina read 1 adapter was directly ligated to the 3' end of a SSB using a randomized splint sequence with a 5' biotin modification (Supplemental Table S4). This ligation was followed by fragmentation, streptavidin purification, end repair, ligation of an Illumina read 2 adapter, and two rounds of PCR amplification (Supplemental Table S2). Additionally, we streamlined the protocol to reduce clean-up steps and retain input genomic complexity and added a UMI tag to the Illumina read 1 adapter that enabled deconvolution of PCR duplicates (Fig. 1A). We named this optimized protocol 3'-end ligation-seq (PEG-seq).

We initially tested PEG-seq on genomic DNA digested with the nicking endonuclease Nb.BsrDI, which generates a SSB 3' to the "GCAATG" recognition motif on the bottom strand. Analysis of PEG-seq data from human genomic DNA treated with saturating amounts of Nb.BsrDI showed sharp peaks of DNA enrichment across the genome with 1,099,701 sites receiving more than seven unique reads in at least one replicate. Examination of individual sites showed a clear Nb.BsrDI recognition site 3' to the start of the peak of signal (Fig. 1B). Overall, this protocol captured 72.01% (Rep.1) (Supplemental Table S7) or 74.08% (Rep.2) (Supplemental Table S8) of the expected Nb.BsrDI sites in the ge-

nome, with coverage appearing to be limited by sequencing depth (Fig. 1C). Further genome-wide analysis of the 6 bp immediately 3' to each detected "peak" revealed a strong motif that was identical to the Nb.BsrDI recognition sequence (Fig. 1D). Examining the position-weight matrix, we note that the bases immediately 5' to the recognition motif where the adapter is ligated during the experiment are nearly randomly distributed, suggesting the PEG-seq protocol end capture step shows low sequence bias (Fig. 1D). To further explore and quantify any ligation bias, we carried out a pooled random end-capture experiment and similarly found little or no difference in ligation efficiency based on the first or second bases adjacent to the nick (Supplemental Fig. S1; Supplemental Table S9).

As shown above, the Nb.BsrDI enzyme generates about 10⁶ SSBs per genome. We expect SpCas9 nicking endonucleases to generate far fewer, on the order of 10¹–10³ of SSBs per genome, based on the number of off-target sites observed by different methods with comparable intensity to the on-target site (Kim et al. 2015; Tsai et al. 2017). To test if our protocol could identify SSBs induced by SpCas9 variants, we digested genomic DNA with SpCas9, SpCas9^{D10A}, SpCas9^{H840A}, and PE (which also contains the SpCas9^{H840A} domain) proteins complexed with *FANCF* sgRNA or *RNF2* sgRNA and performed the PEG-seq protocol (Supplemental Table S1). The *FANCF* and *RNF2* spacer sequences were selected because they have been widely used as "models" for on- and off-target SpCas9 experiments, and substantial data are available on their off-target activity. Examining the expected on-target site, we readily identified a strong signal of read accumulation on the nontarget strand for the SpCas9^{H840A} and PE protein, the target strand for the SpCas9^{D10A} protein, and both strands for SpCas9 (Fig. 1E,F). We note that compared with SpCas9^{D10A}, SpCas9^{H840A} and PE show a site of nicking 5' to the expected site, the exact features of which appear to be spacer dependent, suggesting staggered cutting or chew back by the RuvC nuclease domain. We observe, supportive of processive removal of the 3' end during digestion (chew back), that when a digestion time course is run, the position of the nick identified by PEG-seq at the on-target site moves further 5' in the spacer sequence (Supplemental Fig. S2). The absolute signal also varied between the SpCas9^{D10A} and SpCas9^{H840A} variants, suggesting potential differences in activity that may be spacer dependent. Critically, the SpCas9 condition (which contains both nickase activities) appears to be a combination of the SpCas9^{D10A} and SpCas9^{H840A} conditions. Overall, these results show that PEG-seq robustly detects SSBs induced by either SpCas9 nicking domain with strand and base resolution, allowing for differential activity analysis of SpCas9 mutants.

In addition to the on-target site, we expect to observe PEG-seq signal at off-target sites where the SpCas9 nickase recognizes sites in the genome, including at sites with a close sequence match to the spacer sequence. We therefore developed an analysis pipeline (Supplemental Code) that (1) aligns read 1 sequences to the genome (note each read 1 sequence starts with the base that was ligated to the capture adapter at the SSB site and is typically 75 bp long), (2) removes reads in regions with inaccurate sequencing (blacklist regions) and deduplicates reads based on UMI analysis, (3) scans the genome and identifies regions with a disproportionate number of read 1 start sites on one strand, and (4) thresholds, filters, and merges this list of peaks to identify likely sites where SSBs disproportionately occur and assigns a PEG-seq score by computing the local enrichment of the observed peak. We analyzed the above *FANCF* and *RNF2* data from the PEG-seq of SpCas9, SpCas9^{D10A}, and PE proteins (SpCas9^{H840A}). For all enzymes, we identify multiple peaks across the genome. When applying a

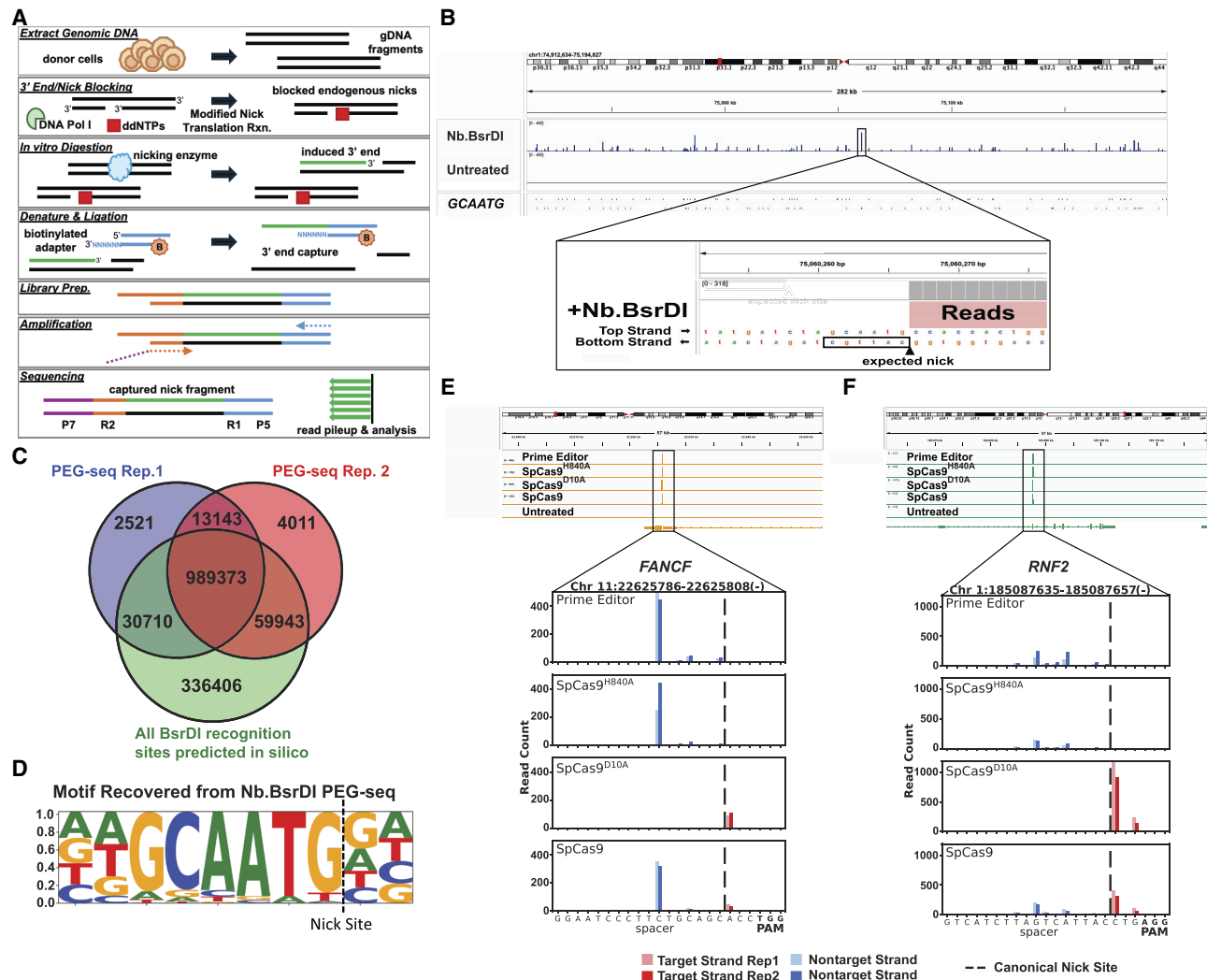


Figure 1. Development and validation of PEG-seq as a genome-wide nick detection method. (A) Diagram of PEG-seq protocol for capture of single-strand breaks (SSBs) in DNA induced by an enzyme. (B) PEG-seq detection of nicks from Nb.BsrDI-treated DNA visualized using IGV (Robinson et al. 2011) with a ~280 kb region view and zoom-in (52 bp) to one peak showing a pileup of reads next to the Nb.BsrDI enzyme recognition site. (C) Overlap (dark brown) between Nb.BsrDI PEG-seq replicates and sites identified by in silico analysis across the genome. (D) The motif identified from PEG-seq sites is identical to that of the Nb.BsrDI recognition site (GCAATG). (E, F) Identification of peaks in PEG-seq data at the on-target site for the prime editor, SpCas9^{H840A}, SpCas9^{D10A}, and SpCas9 complexed with sgRNA targeting *FANCF* (E) or *RNF2* (F) sequences. (Top) Genomic region around the target site in IGV indicating one significant peak in all treated conditions. (Bottom) Zoomed-in view of signal around the spacer sequence of the on-target site; reads from two replicates are shown with base pair resolution (indicated by dark and light shading). As expected, reads from the SpCas9^{D10A} data set are on the target strand, and prime editor or SpCas9^{H840A} shows accumulation of reads on the nontarget strand; SpCas9 shows reads on both strands as it maintains both enzymatic activities. The expected site of canonical nicking is indicated with a dashed line and aligns with the target strand (SpCas9^{D10A}) signal, whereas the nontarget strand signal (prime editor, SpCas9^{H840A}, SpCas9) is offset by 4–6 nucleotides.

threshold of P -value 10^{-30} , SpCas9^{D10A} recovers 398 and 2599 sites for *FANCF* and *RNF2*, respectively, whereas PE recovers 2599 and 8985 sites for *FANCF* and *RNF2*, respectively, implying that in vitro SpCas9 generates many SSBs in addition to the classical DSBs (Fig. 2A,D). Zooming into these data in detail, we observe sites in the genome associated with both RuvC (SpCas9^{H840A}) and HNH (SpCas9^{D10A}) enzymatic activities, and sites associated with only RuvC or HNH activity (Fig. 2B,E). Critically, and consistent with a naive model, we find a strong linear correlation between the PEG-seq signal identified for SpCas9 and the sum of SpCas9^{D10A} and SpCas9^{H840A} activities (Fig. 2C,F).

To gain insight into the nature of the selectivity difference in RuvC (nontarget strand) and HNH (target strand) nicking activity,

we visualized the proportion of matched and unmatched nucleotides across the spacer sequence. Consistent with the observations above and literature on SpCas9, the PAM-proximal region is highly conserved across off-targets whereas the 5' PAM distal region is more weakly constrained (Fig. 2G). We further observe differences between the RuvC and HNH activities by this measure, with somewhat less constraint being observed for the RuvC (SpCas9^{H840A}) in the 8–10 nt region, consistent with the larger number of potential off-target sites identified in the PE and SpCas9 nickase data sets (Fig. 2G). For all enzymes, we observe that as the peak signal increases (PEG-seq score), the tendency for those sites to have low mismatches to the sgRNA spacer sequence also increases. However, because of the relatively small number of sites identified

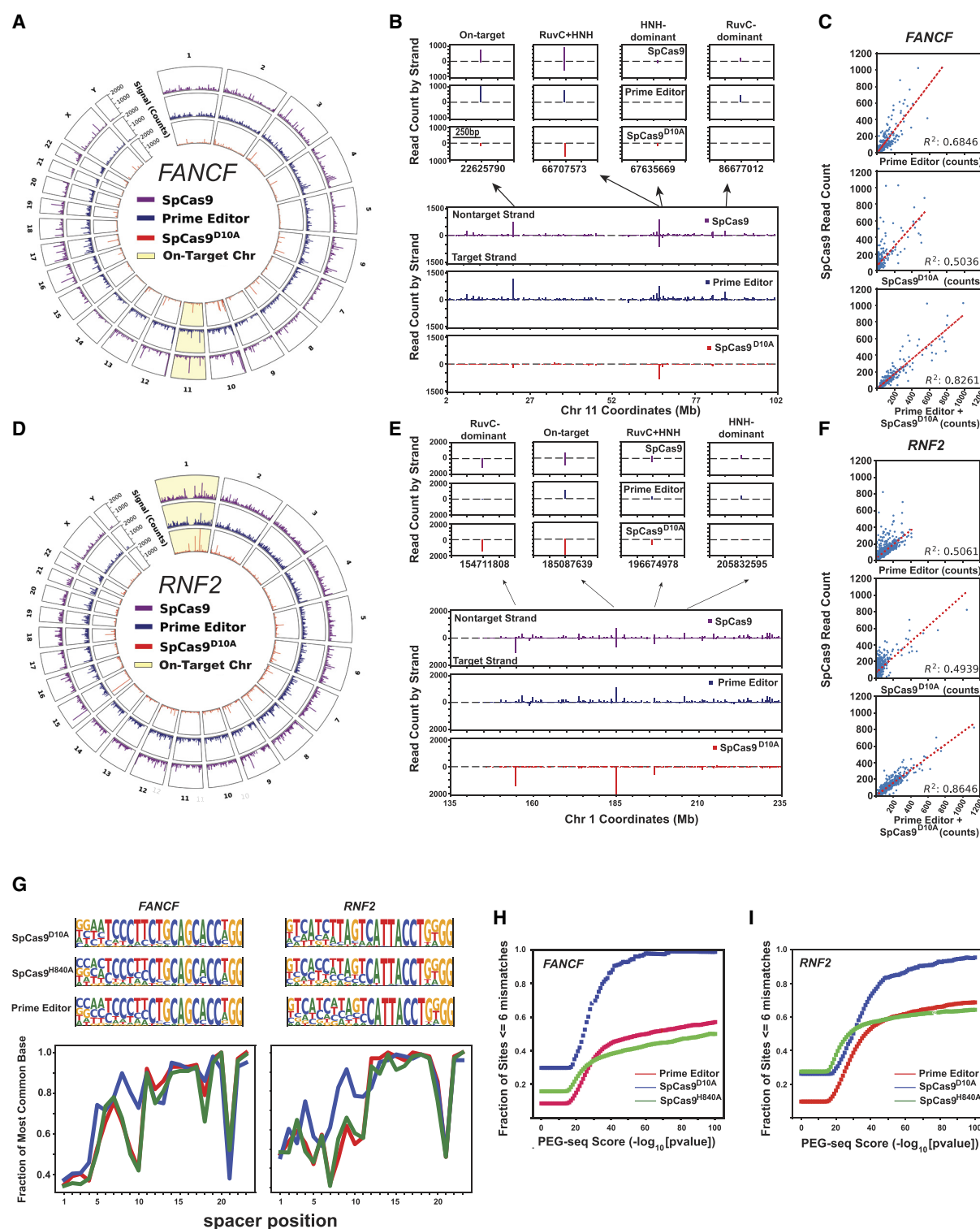


Figure 2. PEG-seq protocol detects off-target SSBs generated by SpCas9, prime editor, and SpCas9^{D10A} genome-wide. (A,D) Circle plot of genome-wide PEG-seq signal for each indicated enzyme for the *FANCF* (A) or *RNF2* (D) guide RNA. (B,E) Zoom-in on 100 Mb of Chromosome 11 showing the on-target site for *FANCF* guide (B) or 100 Mb on Chromosome 1 for the *RNF2* guide (E) and the off-target signal that is apparent in enzymes with only the RuvC (prime editor) or only HNH activity (SpCas9^{D10A}) or both (SpCas9). (C,F) Analysis of SpCas9 signal plotted against either prime editor, SpCas9^{D10A} counts, or the sum of the PE and SpCas9^{D10A} counts for *FANCF* (C) or *RNF2* (F) guides. As expected the SpCas9 signal is strongly predicted by the sum of the signals of the two individual enzymatic activities. (G, top) Position-weight matrices for all sites identified with PEG-seq using the *FANCF* or *RNF2* sgRNA and prime editor, SpCas9^{D10A}, or SpCas9^{H840A} proteins. (Bottom) Line plots of the most frequent bases at each position in the spacer and PAM for the top 100 sites as identified by PEG-seq for *FANCF* and *RNF2*. (H,I) Plots of fraction of sites identified by PEG-seq at each score threshold that are within six mismatches or gaps of the indicated spacer sequence.

at high scores, this increase is not monotonic and shows variation as individual sites drop out (Fig. 2H,I).

From these experiments we conclude that (1) PEG-seq identifies potential off-target sites with high sensitivity and fidelity, (2) the SpCas9^{H840A} used in PE has a similar pattern of off-targets compared with the wild-type SpCas9 protein DSB activity but also recognizes additional sites owing to a more relaxed recognition of specific positions in the spacer sequence, and (3) SpCas9 can generate nicks in the genome that do not form DSBs, potentially expanding the spectrum of off-target sites SpCas9 might cause beyond those typically considered.

We then focused on the PE data and set out to determine PEG-seq's ability to identify potential off-target sites. Examining data from PEG-seq performed using a PE complexed with a *FANCF* or *RNF2* sgRNA (Fig. 3A,B) using a thresholding strategy based on *P*-value peak enrichments (see Methods), we identify 5368 and 8981 potential off-target sites, respectively, dozens of which show comparable signal to the on-target site (Fig. 3C,D). We note that at relatively high in vitro RNP concentrations, the on-target site is not always the strongest signal, as has been previously observed for other off-target methods (Cameron et al. 2017; Lazzarotto et al. 2020). Additionally, we found sites identified by PEG-seq for *FANCF* and *RNF2* spacer sequences often overlapped with published off-target data sets generated using traditional DSB-based SpCas9 methods such as CIRCLE-seq, GUIDE-seq, and in silico prediction (Fig. 3A–D; for SpCas9, SpCas9^{D10A} and SpCas9^{H840A}, see Supplemental Fig. S3A,B). We note that in addition to the absolute PEG-seq signal, the position of the peak identified in the spacer varies with substantial “chew back” or staggered cutting observed at the on-target site and the majority of the signal falling on the expected nick site for off-target sites. This suggests that the position of nicking observed with PEG-seq could be a measure of the RNP's affinity for the site. Although we generally observe a trend of increased 3' signal at lower mismatch sites, this behavior is heterogeneous, and we lack an orthogonal RNP affinity measurement that would let us generalize this observation.

To validate the ability of PE to nick at these target sites, we selected the on-target and the top nine identified off-target sites from the PEG-seq data set for *FANCF* and *RNF2* spacers (10 total sites each), cloned those sequences into plasmids, and digested those plasmids with a PE–sgRNA complex (Supplemental Table S5). We could observe in vitro nicking in the majority of target plasmids in these conditions using an assay that measures loss of supercoiling when assessed by agarose gel electrophoresis (seven of 10 for *FANCF*, eight of 10 *RNF2*) (Supplemental Fig. S4).

To optimize reaction conditions, we compared PEG-seq signal at 10, 100, and 500 nM PE RNP concentrations and found that 100 nM appears to optimize on-target signal (Supplemental Fig. S5A, B). Lower or higher doses of RNP resulted in less or more identified sites, respectively (Supplemental Fig. S5C,D). We note, however, that these results depend on the activity of the RNP complex and higher RNP concentrations can be helpful in conditions with lower enzymatic activity or with sgRNA derivatives, such as pegRNAs, that may not fold or complex as well.

To extend and generalize this analysis, we analyzed 10 additional sgRNA spacer sequences with the PE protein. In each case, we observed a strong on-target signal that, in seven of 10 guides tested, was in the top 25 identified sites and in all (10) cases was above our threshold for identification (Fig. 4A; Supplemental Fig. S6). Examining the PEG-seq top 25 site list for a selection of guides (*EMX1*, *HEK3*, *VEGFA*), we identified strong potential off-target sites (low mismatches to the spacer sequence), some of which

have previously been identified as having off-target activity by SpCas9-based assays (Fig. 4B–D; for the remaining spacer plots, see Supplemental Fig. S6; note the *VEGFA* spacer used here does not have CIRCLE- or GUIDE-seq data available). To verify the ability of PEG-seq to identify “protospacer-like” sequences in the genome that are substrates for nickase activity, we quantified the fraction of identified sites that had 6 or less mismatches at a range of PEG-seq score thresholds (Fig. 4E). In all cases, we observe a trend of an increasing fraction of sites with a close match to the spacer, although the degree of enrichment varies across the spacers examined. Examining this, we assigned a cutoff PEG-seq score of 30 to identify high-likelihood true-nicking at off-targets, although, we note that this threshold may be modified depending on the spacer activity, read depth, and desired specificity (see Methods) (Supplemental Fig. S7). At this threshold, on-target sites of all 12 spacer sequences we examined were recovered.

Prime editing utilizes pegRNAs, which consist of a standard sgRNA with a 3' extension encoding a potential genomic edit. We evaluated if PEG-seq could be used to assess PE complexed with pegRNAs for *HEK3* (encoding a +CTT edit), *VEGFA* (G-to-T edit), or *EMX1* (G-to-T edit) and compared the signal for the pegRNAs to the corresponding sgRNAs (Supplemental Fig. S8A). PEG-seq identified the on-target and several low-mismatch off-target sites for all three pegRNAs (Supplemental Fig. S8B–D). The signal obtained with *HEK3* and *VEGFA* spacers was highly correlated between the sgRNA and pegRNA data sets ($R^2 > 0.8$), whereas the *EMX1* spacer correction is somewhat weaker (R^2 of 0.6), likely owing to the lower numbers of sites detected overall in the *EMX1* data set (Supplemental Fig. S8E–G). These data support the notion that the 3' sequence extension in the pegRNA does not greatly alter the specificity of the RNP in the human genome.

To determine if PEG-seq can identify potential off-target sites in the human genome for prime editing or nuclease SpCas9 editing, we used our prime editing data sets from the 12 guides examined (Supplemental Table S6) to select the top 250 sites from each guide and designed a hybrid capture pool to detect potential off-target activity at each of these sites. We edited HEK293T cells in duplicate using SpCas9 or PE mRNA and pegRNAs designed to insert a 3 bp sequence at each site using a fixed RTT and PBS length. On-target editing for SpCas9 varied from 41% to 96% (average, 83%) and 11% to 73% (average, 33%) for PE (Supplemental Fig. S9A,B). Looking across all sites sequenced, we observed two off-targets for PE, one each in *VEGFA* and *HEK4*, each of which also had SpCas9 off-target editing at those sites (Fig. 5A, Supplemental Fig. 9C). Additional off-target sites for *VEGFA* and *HEK4* were observed with only SpCas9 and not PE (Fig. 5A). These results suggest that (1) prime editing and SpCas9 off-targets are identified by PEG-seq, and (2) prime editing off-targets appear to be less frequent than SpCas9 off-targets. Both PE off-target sites identified show signal in both the PE and SpCas9 samples and would therefore be captured by other off-target methods designed for SpCas9 as well. Indeed, the *HEK4* PE off-target site has been previously observed (Anzalone et al. 2019) and has been previously identified by GUIDE-seq (Tsai et al. 2015).

To further investigate if PEG-seq could be used to further characterize on- and off-target editing, we focused on the edits observed for *VEGFA* and *HEK4* pegRNAs. Second, we analyzed edited cell populations by applying the targeted PEG-seq protocol (Fig. 5B), in which a target-specific primer was used to analyze a single locus for nicks and extensions at both on- and off-target sites of *VEGFA* and *HEK4*. Using this approach, we observe robust on-target signal for both SpCas9 mRNA-treated and PE mRNA-treated

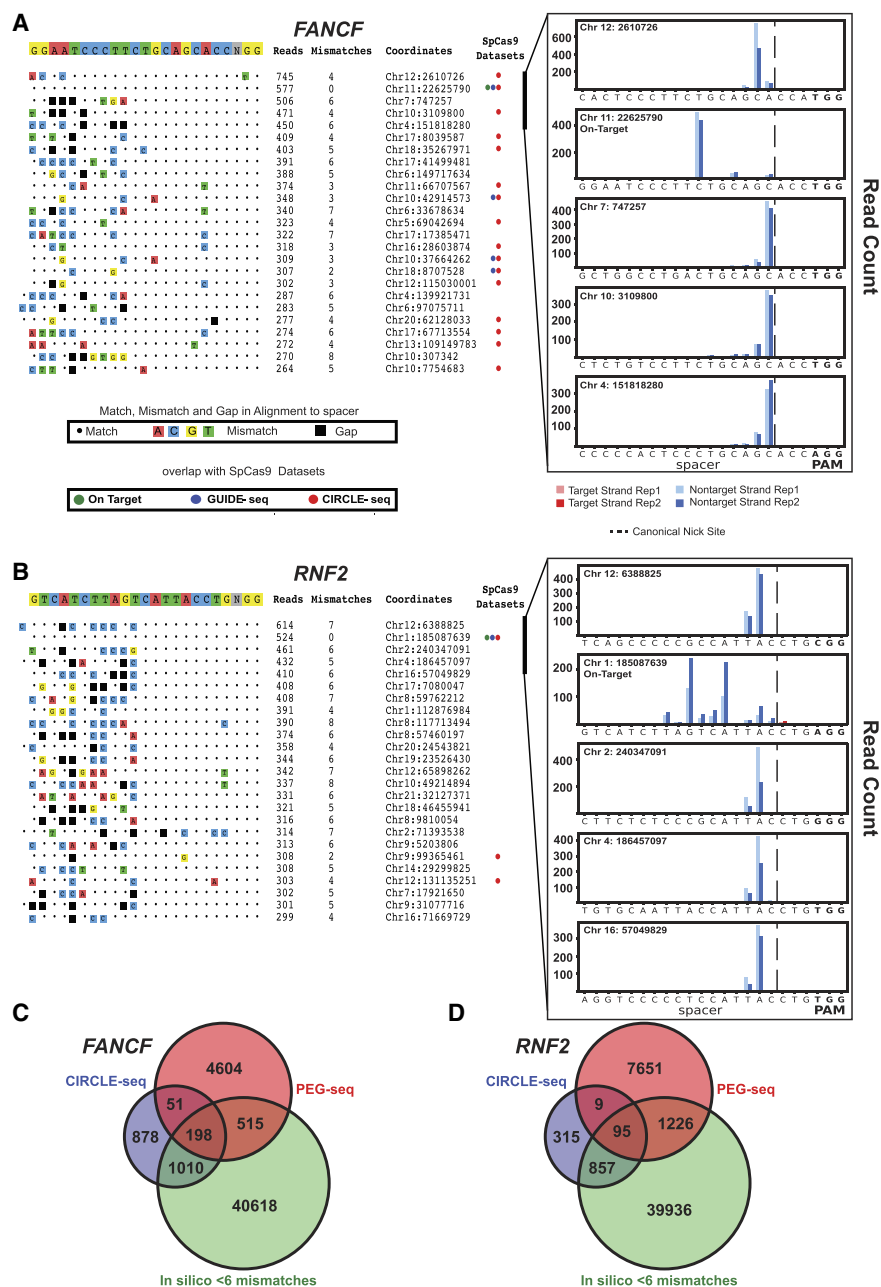


Figure 3. PEG-seq can detect in vitro off-target nicking by the prime editor. (A, B, left) Sorted list of top 25 highest signal peaks in PEG-seq data for *FANCF* (A) or *RNF2* (B) sgRNA spacer sequences. For each site, the alignment to the spacer sequence is indicated with any mismatches (colored boxes) or gaps (black boxes), average read number, genomic location, and overlap with SpCas9 off-target data sets for CIRCLE-seq or GUIDE-seq shown for each potential off-target site. (Right) Plots of the top five sites for each spacer are shown for two replicates (indicated by shading) displaying the accumulation of DNA breaks across the spacer sequence with strand and base pair resolution. It is notable that the precise position of the reads relative to the expected canonical nick site (gapped line) varies across sites and is notably left-shifted at the on-target site. (C, D) Venn diagrams of genome-wide sites identified by PEG-seq, CIRCLE-seq, and in silico analysis (sites six or fewer mismatches or gaps in the human genome) for the *FANCF* (C) and *RNF2* (D) spacers.

cells with the expected “+CTT” extension being observed for the PE samples (Supplemental Table S10). At the off-target sites, a robust nick site signal is observed in the SpCas9-treated samples and a weak signal including a reverse-transcribed sequence in

the PE-treated samples (Fig. 5C,D). Overall, these results confirm that PEG-seq can be used to detect both nuclease and SSB editing sites in cells and suggests that these editing modalities have very different repair kinetics and availability in cells.

Discussion

We have developed PEG-seq, a sensitive biochemical approach for detecting nicks in genomic DNA generated by enzymatic activity in isolated genomic DNA. This method fills a gap in the armamentarium to detect off-targets by identifying nicking sites of SpCas9 gene editing agents with strand- and base-specific resolution. Using this tool, we identified differences in the activities of the HNH and RuvC domains using mutants with one or the other activity compromised and using SpCas9 protein when both activities are present. We find that SpCas9 generates many more detectable nicks on the nontarget strand compared with the target strand throughout the genome and that this activity is recapitulated with SpCas9^{H840A} and SpCas9^{D10A} proteins that have only nontarget strand (RuvC) or target strand (HNH) activity. This result suggests that in addition to the DSBs identified by traditional SpCas9 off-target analysis, SSBs may also occur in the genome of cells treated with SpCas9, perhaps at a frequency several times higher than DSBs, although validation of these in vitro findings will require sensitive in-cell methods to identify these events and their consequences.

We apply the PEG-seq approach to identify off-target sites recognized by the PE protein and show that this strategy identifies potential off-target sites efficiently in the genome for many different spacer sequences. These off-target sites partially overlap with SpCas9 sites identified by DSB off-target methods. Further, analysis of these sites in edited cell populations identifies off-targets generated by both SpCas9 and PE, although PE off-targets are relatively rare. More generally, we note that, like DSBs induced by SpCas9, the majority of SSB sites detected for PE have strong matches to the spacer sequence, especially in positions 11–20. These results suggest that, as for SpCas9, in silico methods are useful

to identify potent potential off-target sites, such as those sequences with three or fewer mismatches to the spacer sequence, but have limitations in sorting more divergent sites that nonetheless can interact efficiently with the enzyme. Collection of larger data sets

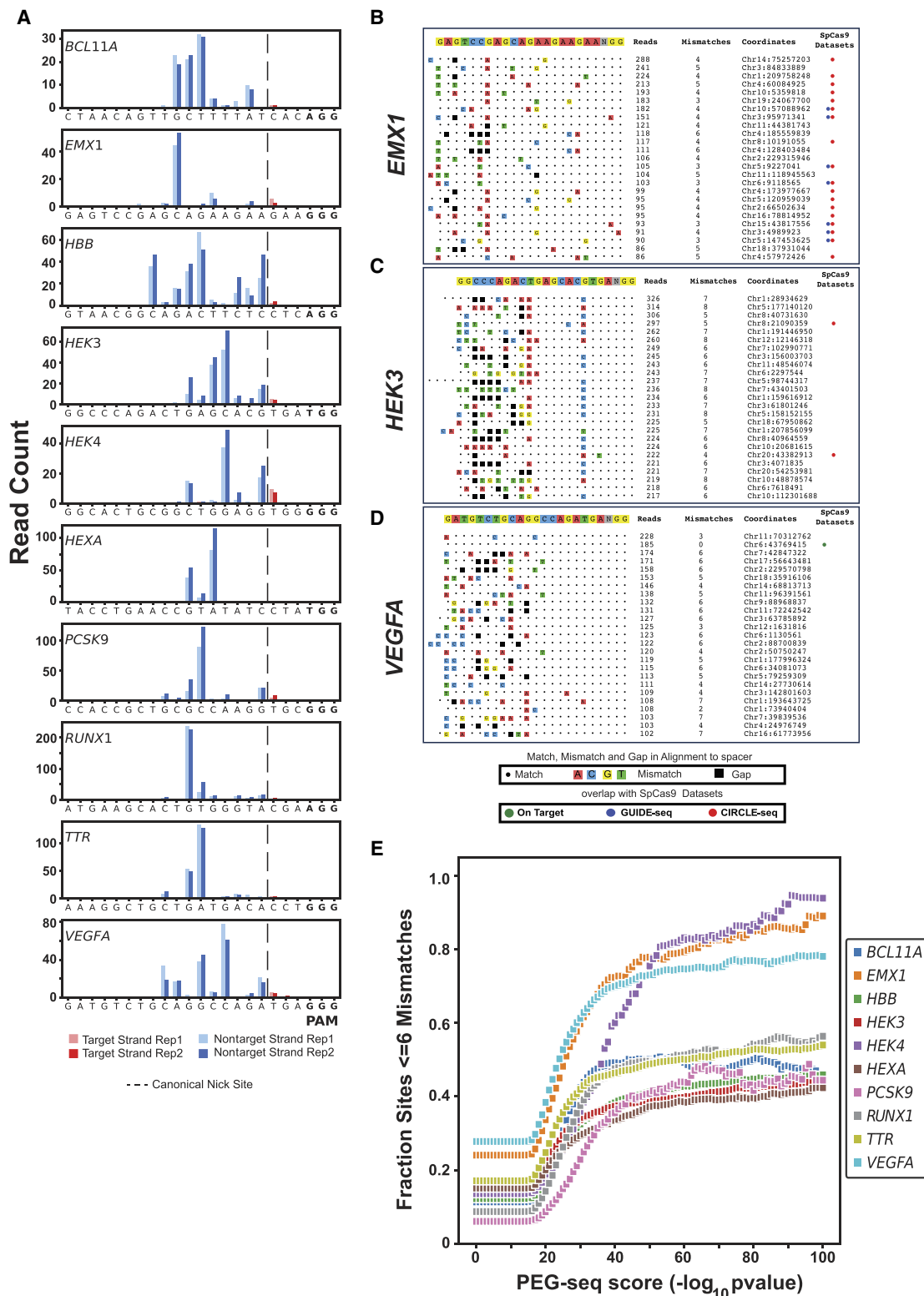


Figure 4. PEG-seq detects off-target sites for a variety of spacer sequences. (A) Distribution of reads at the on-target for each of the indicated spacer sequences. Note that although the majority of signal comes from the nontarget strand (blue), a small amount of signal is observed on the target strand (red) for some spacers (e.g., *HEK4*). Data from replicate experiments are plotted for each site indicated by bar shading. (B–D) Sorted list of top 25 highest signal peaks in PEG-seq data for *EMX1* (B), *HEK3* (C), and *VEGFA* (D). For each site, the alignment to the spacer sequence is indicated with any mismatches (colored boxes) or gaps (black boxes), average read number, genomic location, and overlap with SpCas9 off-target data sets for CIRCLE- or GUIDE-seq shown for each potential off-target site. (E) Plot of the fraction of sites identified by PEG-seq at each score threshold that are within six mismatches or gaps of the indicated spacer sequence.

may enable improvement of *in silico* algorithms to capture the unique signatures of SSBs induced by the RuvC and HNH domains of SpCas9. More generally, as described in the introduction, there are many potential methods to detect off-target activity of PE, including PE-tag (Liang et al. 2023), and biochemical approaches for identifying SSBs, such as nick-translation methods like DENT-seq

(Elacqua et al. 2021). Developing approaches to comprehensively benchmark these tools will be important for the field moving forward.

The ability to rapidly profile potential off-target sites for nick-dependent editors opens the potential to robustly analyze prime editing and base editing and to further separate activities of these

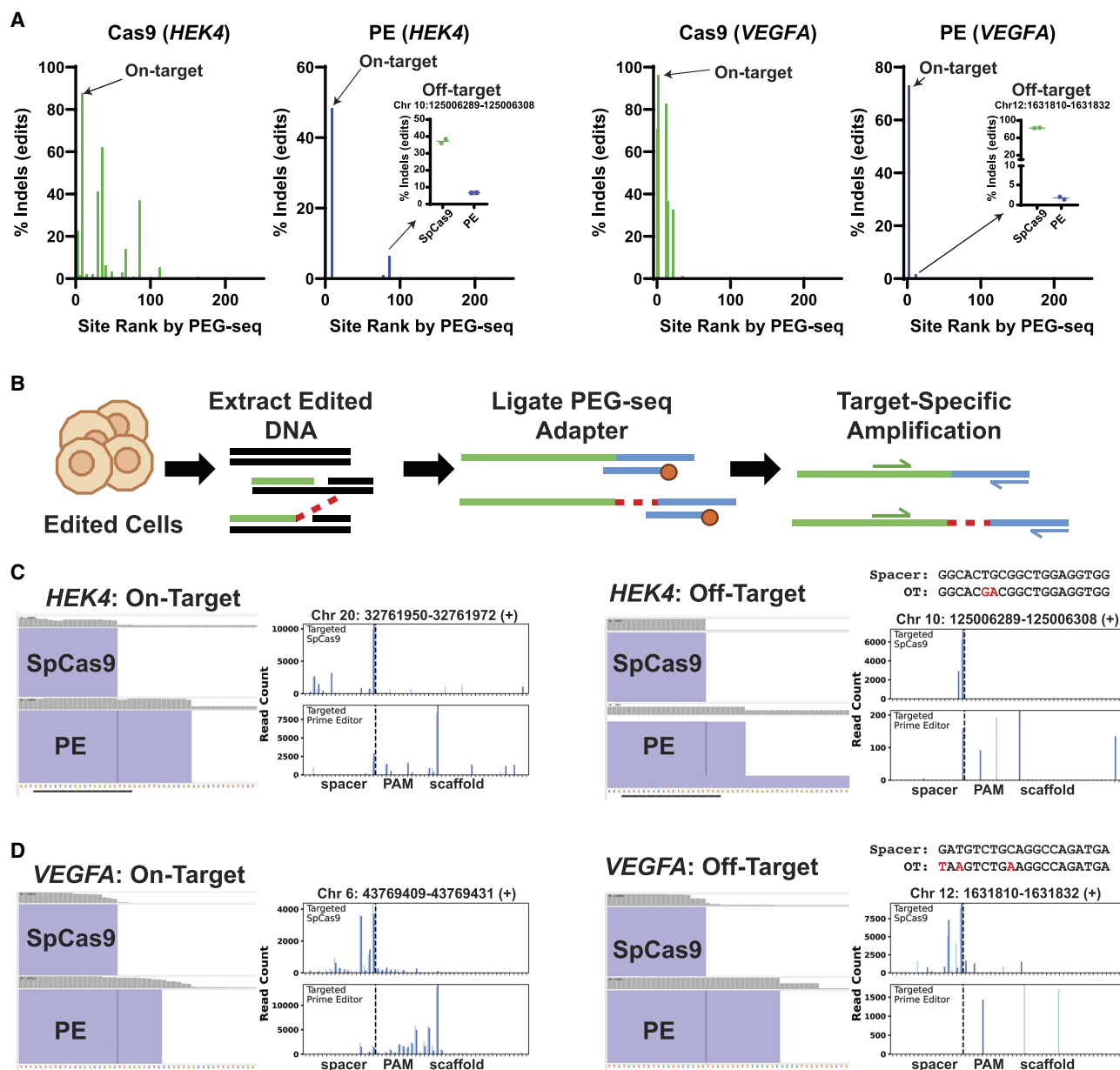


Figure 5. Analysis of PEG-seq identified off-targets in edited cells for SpCas9, and prime editor validates the ability of PEG-seq to identify editing sites in cells and shows RT extension at the on-target and infrequently at the identified off-target sites. (A) Hybrid capture analysis of SpCas9 and PE edited cells for the HEK4 (left) and VEGFA (right) pegRNAs. The top 250 sites identified by PEG-seq were analyzed, and the indel rate minus mock edited sample is shown averaged across two replicates. Each on-target is identified, and the prime editor off-target is shown in the *inset* plot (N = 2). (B) Overview of the targeted PEG-seq method used to analyze in cellulose generated DNA nicks and flaps at each corresponding on- and off-target site. (C,D) For each on- and off-target site, a representative IGV plot displays the mapped sequencing reads (purple blocks) and the read coverage at each position (gray bars) for prime editor and SpCas9 samples. (C,D) A PEG-seq signal plot is included with each IGV plot to display the count of reads that start at a given position. In all samples, the SpCas9 signal is apparent with a strong accumulation at the predicted nick site or 1–5 bp 5' to that site (nick site is indicated by a broken line). A “scaffold” region is included to depict reads demonstrating RTT extension (seen as an accumulation of read start sites 3' to the predicted nick location) that would not align to the reference sequence. The PE signal, including RTT extension, is apparent as the dominant signal at each on-target site (left) and is measurable, albeit at low frequency, in the off-target sites (right).

editors. For example, in the context of BE, differentiating nicking from base modification may be important as, in principle, both events are not required for DNA editing, whereas, for PE, robust SSB detection can potentially be used to both profile nicking activity and reverse transcriptase extension.

Methods

Extracting genomic DNA and blocking background nicks

The input substrate of PEG-seq is high-molecular-weight genomic DNA (HMW gDNA). Healthy single-donor bulk leukocytes from a mobilized leukopak (AllCells) are prepared and treated with ammonium chloride solution (STEMCELL) to lyse the remaining red blood cells according to the manufacturer's instructions. HMW gDNA is then extracted from the remaining leukocytes with a Puregene cell kit (Qiagen) according to a scaled-up version of the manufacturer's instructions. The average size of purified gDNA fragments is confirmed to be >50 kb using a 0.8% agarose gel (Lonza) with a 1 kb extend ladder (New England Biolabs [NEB]) and is quantified by Qubit fluorimetry (Invitrogen).

The extracted genomic fragments contain both free-ends and random nicks that contribute to significant background signal. Prior to *in vitro* digestion, HMW gDNA is treated with a modified nick translation reaction to incorporate dideoxynucleosides (ddNTPs; TriLink) at available 3' ends, inhibiting capture in subsequent steps. Replicate blocking reactions are prepared in a 96-well plate (Bio-Rad) by combining 62.5 μ M of each ddNTP base (TriLink), ~10 μ g of extracted HMW gDNA, and 15 units of DNA Pol I (NEB) in 1 $\times$ CutSmart buffer (NEB) to a final volume of 20 μ L per well. The blocking sample plate is incubated for 90 min at 15°C, 60 min at 37°C, and 20 min at 75°C. Following incubation, all blocked samples are pooled and allowed to gently mix on a tube rotator for ~30 min. Blocked HMW gDNA is precipitated, washed with 70% ethanol, and resuspended in IDTE buffer (at pH 7.5; IDT) to remove residual ddNTPs. Each PEG-seq sample requires ~5 μ g of blocked gDNA (~500 ng/ μ L).

Digestion of blocked gDNA

All digestion reactions (excluding Nb.BsrDI samples) are performed for 1 h at 37°C in 1 $\times$ digestion buffer (20 mM HEPES at pH 7.4, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, 5% glycerol). The relative concentration of the RNA guide of interest and protein ribonucleoprotein (RNP) complex will vary by sample, but the molar ratio of RNA guide to protein will be 2:1.

Nb.BsrDI samples are digested for 1 h at 65°C in 1 $\times$ CutSmart buffer (NEB).

First, RNA guides (Supplemental Table S3) are diluted to the appropriate concentration, denatured for 2 min at 95°C, and cooled to 20°C at a ramp rate of 0.1°C/sec. Protein mixes are prepared for the corresponding concentrations in stock digestion buffer. RNP complexes are formed by combining 10 μ L each of the corresponding denatured guide and protein mix followed by incubating for 10 min at 37°C. Approximately 5 μ g of blocked gDNA in 10 μ L is added to the complexed RNP and digested for 1 h.

To terminate the digestion reaction, 1 μ L each of thermolabile Proteinase K (NEB) and RNase A (NEB) is added to the digested samples and incubated for 30 min at 37°C followed by 15 min at 55°C.

Sample denaturation and nick capture ligation

Following termination of the digestion reaction, samples are heat-denatured by incubating for 2 min at 95°C to generate single-

stranded DNA ("ssDNA") then immediately transferred to slushy ice for 2 min. The ssDNA is then incubated with 200 units of salt-T4 DNA ligase (NEB) and 2 μ L of 15 μ M P5 capture adapter in 1 $\times$ T4 ligase buffer (NEB). The ligation reaction has a total volume of 42 μ L and is incubated for 30 min at 22°C followed by overnight at 16°C. Following the ligation reaction, the T4 ligase is heat inactivated by incubating for 20 min at 65°C to prevent subsequent activity in downstream steps.

The P5 capture adapter consists of an Illumina read 1 sequence, a unique molecular identifier (UMI), i5 index sequence, and a P5 sequence. It is annealed to a short oligo that is complementary to part of the read 1 sequence. This short oligo has a randomized 3' end for capture, a 3' amino group for blocking, and a 5' biotin. Bulk stocks of P5 capture adapter are prepared by combining equimolar inputs of *P5_adapter_top* and *p5_adapter_bottom* (Supplemental Table S4) in duplex buffer (IDT) to a working concentration of 15 μ M. The oligos are then annealed by incubating for 2 min at 95°C followed by cooling to 20°C at a rate of 0.1°C/sec.

Library preparation

Following the ligation reaction, samples are enzymatically fragmented to the appropriate size for Illumina sequencing (~200–800 bp) by adding 2 μ L of NEBNext dsDNA fragmentase in 1 $\times$ fragmentation buffer (NEB) to final volume of 50 μ L. The fragmentation reaction is incubated for 1 h at 37°C, terminated by adding 10 μ L of 0.5 M EDTA, and then purified with SPRIselect beads (Beckman Coulter) at a 1.2 $\times$ bead:sample ratio according to the manufacturer's instructions. The fragmented samples are eluted from the SPRIselect beads in 25 μ L of molecular biology grade water. Note that fragmentation results may vary based on the relative size of the input HMW gDNA, and trial reactions may be necessary to optimize for each HMW gDNA sample.

Following the first SPRIselect cleanup, Dynabeads MyOne streptavidin C1 beads (Invitrogen) are prepared according to the manufacturer's instructions. An equal volume (25 μ L) of prepared Dynabeads is added to each sample and incubated for 30 min at room temperature on a tube rotator. Following incubation of the Dynabeads, samples are washed according to the manufacturer's instructions and resuspended in 10 μ L of IDTE. Bound fragments are then simultaneously eluted and converted to double-stranded DNA (dsDNA) with a single-cycle polymerase chain reaction (PCR) reaction consisting of 10 μ L of washed Dynabead-bound sample, 12 μ L of 2X Q5 high-fidelity DNA polymerase (NEB), and 2 μ L of *P5_primer_short* (Supplemental Table S4). Samples are incubated for 2 min at 98°C followed by 5 min at 65°C. The dsDNA products are purified with SPRIselect beads (1.2 $\times$) according to the manufacturer's instructions and eluted in 25 μ L of molecular biology-grade water.

dsDNA products are then end-repaired and 3'-adenylated by adding 1.5 μ L of the NEBNext Ultra II end repair/dA-tailing enzyme mix and 3.5 μ L of the corresponding NEBNext Ultra II end repair/dA-tailing buffer mix (NEB). Samples are incubated for 30 min at 20°C followed by 30 min at 65°C.

A second ligation mix consisting of 15 μ L of NEBNext Ultra II ligation master mix, 0.5 μ L of NEBNext ligation enhancer (NEB), and 1.5 μ L of 15 μ M annealed *p7_adapter_loop* (Supplemental Table S4) is prepared; added to each sample; and incubated for 15 min at 20°C. Following the second ligation, 1.5 μ L of USER enzyme (NEB) is added to each sample and incubated for 15 min at 37°C. The USER enzyme will excise an uracil base from the *p7_adapter_loop* to convert to the y-shaped adapter necessary for downstream amplification and binding the Illumina flow cell. Following USER enzyme treatment, samples are purified with

SPRIselect beads (1.2×) according to the manufacturer's instructions and eluted in 20 µL of molecular biology grade water for amplification.

The looped P7 adapter consists of a 3'-T overhang that allows for ligation at the previously repaired and dA-tailed ends, a partial Illumina R2 sequence that allows for subsequent amplification with an i7 index, and a dU base that will allow for the conversion of the adapter from a loop to a splint structure. Bulk stocks were prepared by diluting *p7_adapter_loop* in duplex buffer to working a concentration of 15 µM. The oligo was then annealed by incubating for 2 min at 95°C followed by cooling to 20°C at a rate of 0.1°C/sec.

Amplification, library QC, and sequencing

Functional products containing both the P5 and P7 adapters are amplified in a 50 µL reaction consisting of 20 µL of sample, 25 µL of 2× Q5 high-fidelity DNA polymerase (NEB), 2.5 µL of 10 µM *P5_primer_short*, and 2.5 µL of 10 µM *i7_primer* (referred to as PCR 2) (Supplemental Table S4). Cycling conditions consist of 30 sec at 98°C, 13 cycles of 10 sec at 98°C and 75 sec at 65°C, followed by 5 min at 65°C and a hold at 4°C.

The PCR 2 product is then purified with a double-sided SPRIselect cleanup (0.5× + 0.8×) following the manufacturer's instructions to narrow the library size range. An additional left-sided SPRI cleanup (0.9×) is performed according to the manufacturer's instructions to remove any residual primer dimer. Purified PCR 2 product is quantified with a Qubit fluorometer and normalized for an additional round of amplification to generate substantial product for sequencing (referred to as PCR 3).

In a 25 µL reaction, 5 ng of purified PCR 2 product is combined with 12.5 µL of 2× Q5 high-fidelity DNA polymerase 10 µM 1.25 µL of *p5_primer_long* (Supplemental Table S4) and 1.25 µL 10 µM of *p7_primer* (Supplemental Table S4). Cycling conditions consist 30 sec at 98°C, eight cycles of 10 sec at 98°C and 75 sec at 65°C, followed by 5 min at 65°C and a hold at 4°C.

PCR 3 product is purified with two rounds of SPRIselect cleanup (0.9×) according to the manufacturer's instruction to remove residual primer dimer. Purified PCR 3 product is then quantified by Qubit fluorimetry and run on a D5000 TapeStation kit (Agilent) to determine the average fragment size for molarity calculations.

Dually indexed PEG-seq libraries were normalized according to the manufacturer's instructions and paired-end sequenced (75 × 25) on a NextSeq 2000 using 100-cycle P3 kits (Illumina).

Protein synthesis and purification

The PE protein was expressed from a pET21b (Novagen) vector in *Escherichia coli*. Cultures were grown in Terrific Broth (Boston Bioproducts) at 37°C to the mid-log phase, at which time the cultures were transitioned to 18°C for induction with 0.5 mM IPTG (Teknova). Cells were left overnight and harvested the following day. Cells were resuspended in 50 mM HEPES (pH 7.4), 1 M NaCl, 10% glycerol, 1 mM TCEP, 0.5% Triton X-100, complete EDTA-free protease inhibitors (MilliporeSigma), benzamide (MilliporeSigma), and lysozyme (MilliporeSigma) before lysis occurred from mechanical disruption. Resulting lysate was clarified and proceeded immediately to a three-step purification process using standard liquid chromatography protein purification techniques. Final purification step included Superdex 200 size exclusion (Cytiva) in 20 mM HEPES (pH 7.4), 400 mM NaCl, 10% glycerol, and 1 mM TCEP to ensure material was free of insoluble aggregate (Supplemental Table S2). Protein was aliquoted and stored at −80°C until use.

Analysis of PEG-seq data and calculating PEG-seq score

To analyze PEG-seq data, the algorithm (1) aligns reads to the genome using BWA (Li and Durbin 2009), (2) removes UMI duplicates using the UMI-tool kit (Smith et al 2017) and reads in ENCODE blacklisted regions (Amemiya et al. 2019), (3) trawls the genome identifying sites with accumulations of more than seven read starts on one strand in a 10 bp window (the window is used because of the chew back identified in PE-treated DNA), (4) compute PEG-seq score value using local (10 kb) read coverage and a Poisson model, and (5) merge nearby peaks with BEDTools (Quinlan and Hall 2010) in a strand specific manner. This list of sites can then be thresholded by PEG-seq score to identify off-target sites. For nicking enzymes, accumulations of reads are strand specific. For SpCas9 nuclease editors reads are accumulated on both strands. Code for this analysis is provided in the manuscript's Supplemental Material, and similar results are easily obtained using standard peak-calling software.

Thresholding

To evaluate an appropriate PEG-seq score threshold for site filtering we took two approaches: (1) using replicate data to define a threshold at which replicates maintain agreement and (2) quantifying the fraction of "spacer-like" sites (six or fewer mismatches from the on-target spacer sequence) across a range of PEG-seq thresholds. Data from *FANCF*, *RNF2*, and the additional 10 sgRNAs were used to calibrate these values (Supplemental Fig. S7).

Based on these analyses, we defined a PEG-seq score threshold of 30 for the analyses in this paper, with a reasonable range for these data between 25 and 35. The threshold may be modified depending on the spacer activity, read depth, and desired specificity, with lower thresholding increasing sensitivity and the number of false positives.

Site filtering with PEG-seq score

Samples are first processed with the PEG-seq pipeline as described above. For each sample and set negative controls, replicate outputs are then concatenated, sorted, and merged with BEDTools to form a union of all unique sites. For sites found in both replicates, the read count is summed and the PEG-seq score is averaged. The union of negative controls sites is the subtracted from each sample list using BEDTools, and sites with a PEG-seq score <30 are filtered out.

In silico analysis

Potential off-target sites were predicted from the hg38 human genome using Calitas (Fennell et al 2021) using the following parameters: up to six mismatches, up to three gaps, and one PAM mismatch.

Ligation bias analysis

An ultramer oligo *LigBias_oligo* (IDT) (see Supplemental Table S4) was designed with Illumina P5 and partial P7 sequences, as well as two sets of randomized Nmers (8N and 10N) flanking a Nb.BsrDI recognition motif sequence. The oligo was first amplified with *LigBias_amp_rev* and *P7_stub_fwd* (Supplemental Table S4) to produce excess double-stranded product and then digested with the Nb.BsrDI restriction enzyme to produce free randomized ends (10N). The PEG-seq ligation reaction was then replicated, and the ligated product was amplified with a *P5_primer_short* and *i7_primer* (PCR 2) to produce viable sequencing libraries. In parallel, "nonnicked" controls were denatured, ligated, and amplified

to provide an “expected” base composition of the randomized Nmers.

FANCF digestion time course

Bulk 100 nM RNP complexes were prepared using *FANCF* sgRNA and each corresponding protein (SpCas9, SpCas9^{H840A} and SpCas9^{D10A}), following the previously described methods. For each protein and time point (5, 15, 30, and 60 min), digestion reactions were performed in duplicate using 5 µg of blocked genomic DNA. At each time point, corresponding digestion reactions were quenched with a mix of thermolabile Proteinase K and RNase A and then processed through the subsequent steps of denaturation, capture ligation, and library preparation, as previously described.

FANCF/RNF2 top PEG-seq sites plasmid digestion

For both *FANCF* and *RNF2*, replicates for the “100 nM sgRNA PE” samples were extended (using BEDTools slop), sorted, and merged with BEDTools. Sites were ranked based on average PEG-seq score, and the top 10 sites for each sample were selected. Sites from repetitive regions were removed and replaced with the next ranked site.

FANCF and *RNF2* “100 nM sgRNA PE” RNPs were prepared as previously described. Prior to digestion, ~500 ng of each target plasmid stock was treated with 15 units of T5 exonuclease (NEB) to remove nonsupercoiled contaminant species. Ten microliters of corresponding 100 nM RNP was added to 5 µL of T5 exonuclease-treated target plasmid (~25 ng total). Approximately 50 ng of heparin was added to the digestion reaction to limit background nicking of the excess RNP (O’Connell et al. 2014). Samples were incubated for 30 min at 37°C and terminated as previously described. Samples were then purified with SPRIselect beads (1.2×) according to the manufacturer’s instructions, eluted in 20 µL of molecular biology grade water, and run on a 1% E-Gel (Thermo Fisher Scientific).

Cell editing experiments with SpCas9 and PE

Lenti-X 293T (Takara) cells were grown in DMEM (Gibco) supplemented with 10% FBS (heat-inactivated; Gibco). Cells were plated in 24-well plates and 1 day later (~70% confluence) were edited using the manufacturer’s protocol with Lipofectamine MessengerMAX (Thermo Fisher Scientific) with mRNA SpCas9 or PE (325 ng/well) and pegRNA (IDT, 125 ng/well). After 72 h, cells were harvested and DNA-prepped using a Monarch genomic DNA extraction kit (NEB) according to the manufacturer’s protocol. The pegRNA’s for the 12 selected spacers were designed by selecting a PBS length of 13 bp, a +CTT insertion, and a total RTT of 16 bp and were ordered from IDT as desalted guides.

Hybrid capture and targeted sequencing

A hybrid capture library was designed by IDT to cover the top 250 sites identified by PEG-seq for each spacer. Each DNA sample was prepped using the NEB Ultra II library prep kit and NEB UMI adapters (NEB). The libraries were captured using the IDT xGen hybrid capture V2 protocol as recommended by the manufacturer (IDT). Libraries were pooled and sequenced on an Illumina NextSeq2000 using XLEAP chemistry (Illumina). FASTQ files were aligned using BWA and UMI corrected using Gencore before calling for indels at the potential off-target sites using a pysam-based strategy highly similar to that described by Chaudhari et al. (2020). The indel rate at each site (12 bp surrounding the predicted nick site) was evaluated and, if at least 1000 reads (average, 2852 reads/site for rep1 and 2623/site for rep2), were recovered and subtracted for background from the mock controls. Significant off-

targets were identified with a 1% threshold averaged across the two replicates. All sites passing these thresholds were manually evaluated by viewing in the Integrative Genomics Viewer (IGV; Robinson et al. 2011) and to determine if they represent genuine off-target sites with an expected indel (SpCas9) or insertion pattern (PE).

Targeted PEG-seq

Lenti-X 293T cells were edited using Lipofectamine MessengerMAX with SpCas9 or PE mRNA and *HEK4* or *VEGFA* pegRNA as previously described. After 24 h, cells were harvested and HMW gDNA was extracted using the Puregene cell kit, heat-denatured, and ligated with the PEG-seq capture adapter as previously described.

Target-specific primers for *HEK4* and *VEGFA* (Supplemental Table S4) were designed upstream (5′) of each identified on- and off-target site and oriented within the same strand as the expected nick site. Seminested PCR was performed using 200 ng of genomic DNA as input for the initial amplification. The first PCR consisted of 15 cycles using a target-specific primer and the Illumina P5 primer, followed by a 1.0× SPRIselect bead cleanup. The second PCR used a nested target-specific forward primer incorporating a 5′ Illumina Read 2 sequence, along with the P5 primer, for an additional 15 cycles, followed by a second 1.0× SPRIselect bead cleanup. A final i7-indexing PCR was performed for 12 cycles, and the resulting libraries were purified with a 0.9X SPRIselect bead cleanup. Sample replicates were then pooled and cleaned up with an additional 0.5×/0.9× double-sided SPRIselect cleanup, according to the manufacturer’s instructions. Libraries were pooled and sequenced on an Illumina MiSeq.

FASTQ files were aligned using BWA, UMI-corrected using UMI-Tools, and subsampled to 100,000 reads per sample.

Data access

The PEG-seq and capture data generated in this study have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA1123405. All source data for figures are available in Supplemental Table S10, and all other relevant data are available in the paper or supplement. Analysis of PEGseq data can be reproduced using many peak-calling approaches. Code for PEGseq analysis is also available as Supplemental Code.

Competing interest statement

J.S.-O., M.J.I., M.K.L., D.L., M.D.C., D.R., A.N.C., D.H., T.A., and J.S.D. are current or former employees of Prime Medicine and hold equity in Prime Medicine. Prime Medicine has filed a patent application related to this work (WO2025090524A1).

Acknowledgments

We thank Matt Ranaghan for initial support with protein expression and design and Hari Prasanna Subramanian, Chris Wilson, Andrew Anzalone, and Jeff Hussman for discussions around methods for off-target detection and analysis.

Author contributions: M.J.I., M.K.L., and J.S.-O. collected data, analyzed data, and developed the method. T.A. and D.H. contributed to PEGseq analysis strategy. D.L. and M.D.C. provided experimental support and advice on strategy. A.N.C., D.R., J.S.D., and J.S.-O. provided scientific advice and supervision. J.S.-O., M.J.I., and J.S.D. prepared the manuscript.

References

- Amemiya HM, Kundaje A, Boyle AP. 2019. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep* **9**: 9354. doi:10.1038/s41598-019-45839-z
- Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, et al. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**: 149–157. doi:10.1038/s41586-019-1711-4
- Cameron P, Fuller C, Donohoue PD, Jones BN, Thompson MS, Carter MM, Gradia S, Vidal B, Garner E, Slorach EM, et al. 2017. Mapping the genomic landscape of CRISPR–Cas9 cleavage. *Nat Methods* **14**: 600–606. doi:10.1038/nmeth.4284
- Cao B, Wu X, Zhou J, Wu H, Liu L, Zhang Q, DeMott MS, Gu C, Wang L, You D, et al. 2020. Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage. *Nucleic Acids Res* **48**: 6715–6725. doi:10.1093/nar/gkaa473
- Chaudhari HG, Penterman J, Whitton HJ, Spencer SJ, Flanagan N, Lei Zhang MC, Huang E, Khedkar AS, Toomey JM, Shearer CA, et al. 2020. Evaluation of homology-independent CRISPR–Cas9 off-target assessment methods. *CRISPR J* **3**: 440–453. doi:10.1089/crispr.2020.0053
- Ciccia A, Elledge SJ. 2010. The DNA damage response: making it safe to play with knives. *Mol Cell* **40**: 179–204. doi:10.1016/j.molcel.2010.09.019
- Elacqua JJ, Ranu N, Dilorio SE, Blainey PC. 2021. DENT-seq for genome-wide strand-specific identification of DNA single-strand break sites with single-nucleotide resolution. *Genome Res* **31**: 75–87. doi:10.1101/gr.265223.120
- Fennell T, Zhang D, Isik M, Wang T, Gotta G, Wilson CJ, Marco E. 2021. CALITAS: a CRISPR–Cas-aware ALigner for in silico off-TArget search. *CRISPR J* **4**: 264–274. doi:10.1089/crispr.2020.0036
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–821. doi:10.1126/science.1225829
- Kara N, Krueger F, Rugg-Gunn P, Houseley J. 2021. Genome-wide analysis of DNA replication and DNA double-strand breaks using TrAEL-seq. *PLoS Biol* **19**: e3000886. doi:10.1371/journal.pbio.3000886
- Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, Hwang J, Kim J-I, Kim J-S. 2015. Digenome-seq: genome-wide profiling of CRISPR–Cas9 off-target effects in human cells. *Nat Methods* **12**: 237–243. doi:10.1038/nmeth.3284
- Kim DY, Moon SB, Ko J-H, Kim Y-S, Kim D. 2020. Unbiased investigation of specificities of prime editing systems in human cells. *Nucleic Acids Res* **48**: 10576–10589. doi:10.1093/nar/gkaa764
- Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**: 420–424. doi:10.1038/nature17946
- Kwon J, Kim M, Bae S, Jo A, Kim Y, Lee JK. 2022. TAPE-seq is a cell-based method for predicting genome-wide off-target effects of prime editor. *Nat Commun* **13**: 7975. doi:10.1038/s41467-022-35743-y
- Lazzarotto CR, Malinin NL, Li Y, Zhang R, Yang Y, Lee GH, Cowley E, He Y, Lan X, Jividen K, et al. 2020. CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nat Biotechnol* **38**: 1317–1327. doi:10.1038/s41587-020-0555-7
- Lazzarotto CR, Katta V, Li Y, Urbina E, Lee G, Tsai SQ. 2024. CHANGE-seq enables simultaneously sensitive and unbiased in vitro profiling of base editor genome-wide activity. *bioRxiv* doi:10.1101/2024.03.28.586621
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760. doi:10.1093/bioinformatics/btp324
- Liang SQ, Liu P, Ponnieselvan K, Suresh S, Chen Z, Kramme C, Chatterjee P, Zhu LJ, Sontheimer EJ, Xue W, et al. 2023. Genome-wide profiling of prime editor off-target sites in vitro and in vivo using PE-tag. *Nat Methods* **20**: 898–907. doi:10.1038/s41592-023-01859-2
- Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, et al. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**: aaf8729. doi:10.1126/science.aaf8729
- O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA. 2014. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**: 263–266. doi:10.1038/nature13769
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26. doi:10.1038/nbt.1754
- Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res* **27**: 491–499. doi:10.1101/gr.209601.116
- Sriramachandran AM, Petrosino G, Méndez-Lago M, Schäfer AJ, Batista-Nascimento LS, Zilio N, Ulrich HD. 2020. Genome-wide nucleotide-resolution mapping of DNA replication patterns, single-strand breaks, and lesions by GLOE-Seq. *Mol Cell* **78**: 975–985.e7. doi:10.1016/j.molcel.2020.03.027
- Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP, et al. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR–Cas nucleases. *Nat Biotechnol* **33**: 187–197. doi:10.1038/nbt.3117
- Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK. 2017. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nat Methods* **14**: 607–614. doi:10.1038/nmeth.4278
- Yu Z, Lu Z, Li J, Wang Y, Wu P, Li Y, Zhou Y, Li B, Zhang H, Liu Y, et al. 2022. PEAC-seq adopts prime editor to detect CRISPR off-target and DNA translocation. *Nat Commun* **13**: 7545. doi:10.1038/s41467-022-35086-8

Received October 29, 2024; accepted in revised form July 3, 2025.