nature biotechnology

LETTERS

https://doi.org/10.1038/s41587-019-0050-1

Genome-wide target specificity of CRISPR RNA-guided adenine base editors

Daesik Kim[1,2,4], Da-eun Kim[1,2,4], Gyeorae Lee[2,3], Sung-Ik Cho[1,2] and Jin-Soo Kim[1,2,3*]

Adenine base editors[1] enable efficient targeted adenine-to-guanine single nucleotide conversions to induce or correct point mutations in human cells, animals, and plants[1–4]. Here we present a modified version of Digenome-seq, an in vitro method for identifying CRISPR (clustered regularly interspaced short palindromic repeats)-induced double-strand breaks using whole-genome sequencing[5–8], to assess genome-wide target specificity of adenine base editors. To produce double-strand breaks at sites containing inosines, the products of adenine deamination, we treat human genomic DNA with an adenine base editor 7.10 protein–guide RNA complex and either endonuclease V or a combination of human alkyladenine DNA glycosylase and endonuclease VIII in vitro. Digenome-seq detects adenine base editor off-target sites with a substitution frequency of 0.1% or more. We show that adenine base editor 7.10, the cytosine base editor BE3, and unmodified CRISPR-associated protein 9 (Cas9) often recognize different off-target sites, highlighting the need for independent assessments of their genome-wide specificities[6]. Using targeted sequencing, we also show that use of preassembled adenine base editor ribonucleoproteins, modified guide RNAs[5,8–11], and Sniper/Cas9 (ref. [12]) reduces adenine base editor off-target activity in human cells.

CRISPR RNA-guided programmable deaminases, so-called base editors, enable clean and efficient single nucleotide conversions in cultured cells[1,13,14] and whole organisms[2–4,15] without inducing unwanted small insertions or deletions (indels) at target sites and without requiring donor DNA templates. They do not produce DNA double-strand breaks (DSBs) and, as a result, do not rely on error-prone nonhomologous end-joining or template-dependent homology-directed repair of DSBs.

Two types of programmable deaminases have been developed: cytosine base editors (CBEs) inducing C-to-T conversions[13,14] and adenine base editors (ABEs) inducing A-to-G conversions in a guide RNA-dependent manner[1]. To translate base editing to the clinic and biotechnological applications, it is essential to define the genome-wide target specificity of base editing in an unbiased manner. Recently, we were able to use a modified Digenome-seq protocol to detect CBE off-target sites in the human genome[6]. Here we sought to profile ABE off-target activity in the human genome.

We first investigated whether off-target activities of ABE7.10, BE3 (a CBE), and CRISPR-associated protein 9 (Cas9) are different from each other by testing a series of mismatched single guide RNAs (sgRNAs) targeted to endogenous genomic loci (Fig. 1, Supplementary Fig. 1, and Supplementary Table 1). To this end, we transfected HEK293T cells with plasmids encoding sgRNAs having one- to four-nucleotide mismatches and plasmids encoding human codon-optimized ABE7.10, BE3, or SpCas9 and measured frequencies of single nucleotide substitutions or small insertions and deletions (indels) using high-throughput DNA sequencing. As expected, base editors and Cas9 nucleases tolerated most of the single or double mismatches in the sgRNAs but were poorly active when combined with most of the sgRNAs having triple or quadruple mismatches. Notably, the tolerance of Cas9, ABE7.10, and BE3 for mismatched sgRNAs was often different from each other. For example, at the *HEK2* site, several sgRNAs with about two to four mismatches (indicated with asterisks in Fig. 1) were well tolerated by ABE7.10 but were poorly active or inactive with Cas9 or BE3. At the *HEK3* site, Cas9 complexed with an sgRNA having tandem mismatches at positions 3–4 (numbered 1–23 in the 5′ to 3′ direction) was highly active, showing a relative activity of 66% compared with the fully matched sgRNA, whereas ABE7.10 complexed with the same sgRNA was poorly active, showing a relative activity of 15% (Supplementary Fig. 1a). These results suggest that ABE7.10, BE3, and Cas9 could recognize separate sets of off-target sites in the human genome, calling for a method to determine genome-wide ABE specificity in an unbiased manner.

To identify genome-wide off-target sites of ABE7.10 via Digenome-seq, which relies on DNA DSBs produced in vitro[5], we first sought to produce DSBs at sites containing inosine, which result from adenine deamination. Note that we had previously used the recombinant BE3ΔUGI protein (BE3 with no UGI domain) and uracil-specific excision reagent (USER), a mixture of *Escherichia coli* uracil DNA glycosylase and DNA glycosylase-lyase endonuclease VIII (Endo VIII), to generate DSBs in vitro at sites containing uracil, a deamination product of cytosine, to identify genome-wide BE3 off-target sites comprehensively in an unbiased manner[6]. Likewise, we reasoned that *E. coli* endonuclease V (Endo V) or a combination of human alkyladenine DNA glycosylase (hAAG) and Endo VIII could be used to cleave a phosphodiester bond at inosine-containing sites (Fig. 2a and Supplementary Fig. 2). Endo V recognizes inosine in DNA and catalyzes the cleavage of the second phosphodiester bond 3′ to the inosine, whereas hAAG excises hypoxanthine to yield an AP site (apurinic–apyrimidinic site), which is then recognized and processed by Endo VIII to induce a single-strand break. To test this idea, we first treated a polymerase chain reaction (PCR) amplicon containing a target DNA sequence with the recombinant ABE7.10 protein expressed in and purified from *E. coli* (Supplementary Fig. 3) and its sgRNA to catalyze adenine deamination in one DNA strand and to produce a nick in the other DNA strand (Fig. 2a), and then with Endo V (Fig. 2a) or with hAAG/Endo VIII (Supplementary Fig. 2b) to yield a composite DSB. As expected, the PCR amplicon was cleaved by ABE7.10 plus Endo V (Fig. 2b) or by ABE7.10 plus hAAG/Endo VIII (Supplementary Fig. 4a).

[1]Department of Chemistry, Seoul National University, Seoul, Republic of Korea. [2]Center for Genome Engineering, Institute for Basic Science, Seoul, Republic of Korea. [3]Department of Functional Genomics, University of Science and Technology, Daejeon, Republic of Korea. [4]These authors contributed equally: Daesik Kim, Da-eun Kim. *e-mail: jskim01@snu.ac.kr

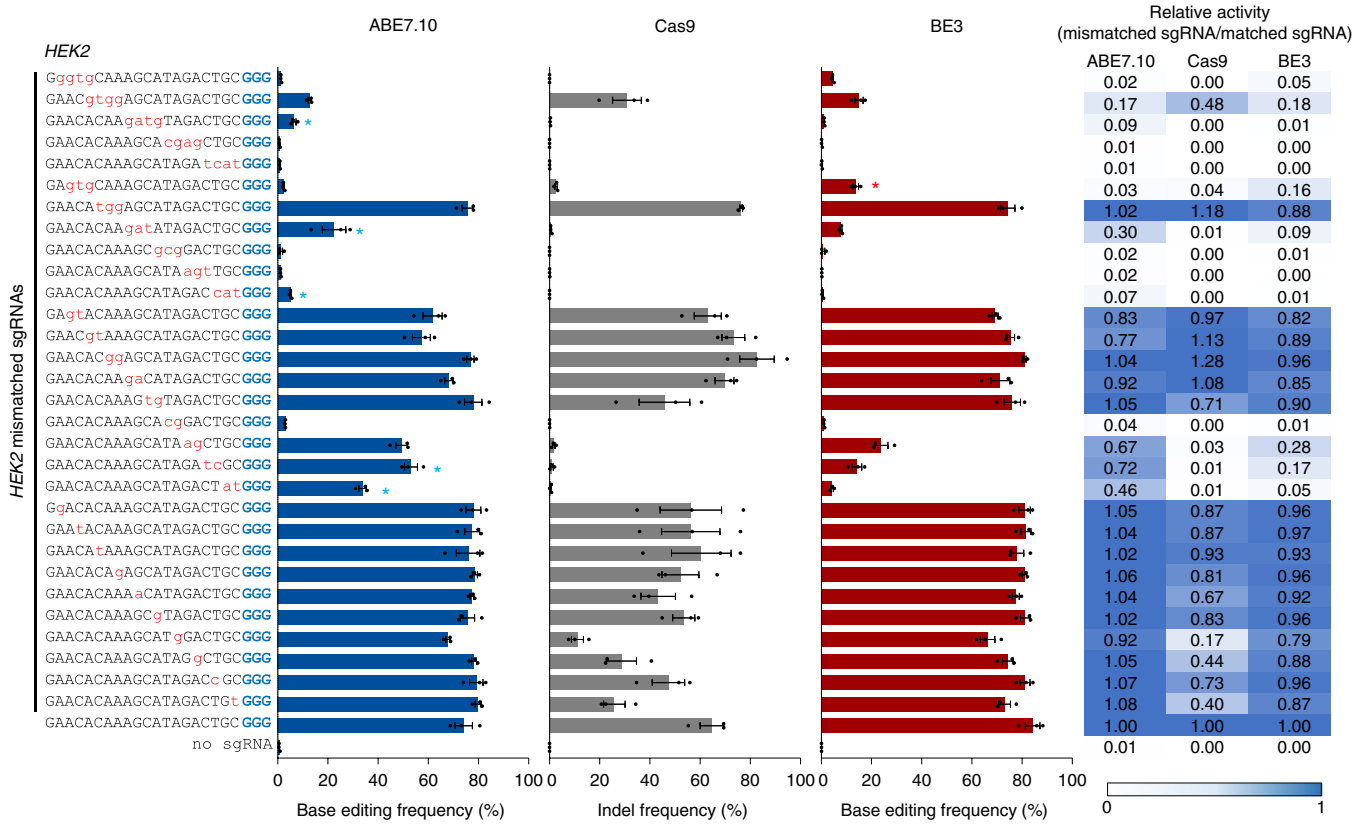


Fig. 1 | Tolerance of ABE7.10, BE3, and Cas9 for mismatched sgRNAs. Mismatched sgRNAs that differed from the *HEK2* site by one to four nucleotides were tested in HEK293T cells. Base editing or indel frequencies were measured using targeted deep sequencing. Mismatched nucleotides and the PAM sequence are shown in red and blue, respectively. Asterisks indicate mismatched sgRNAs whose relative activity (mismatched sgRNA/matched sgRNA) with one enzyme is more than three times higher than that with the other two enzymes. Means ± s.e.m. were from three independent experiments.

Having confirmed that DSBs can be produced at ABE-targeted sites containing inosine using inosine-specific DNA repair enzymes, we next investigated whether human genomic DNA isolated from HEK293T cells could be processed in vitro with ABE7.10 plus Endo V or with ABE7.10 plus hAAG/Endo VIII for Digenome-seq analysis to identify genome-wide ABE off-target sites. Human genomic DNA was incubated with ABE7.10 ribonucleoprotein (RNP) (300 nM ABE7.10 and 900 nM sgRNA) targeted to an endogenous chromosomal site (known as *HEK2*) for 8 hours and then with the repair enzymes for 2 hours. We used Sanger sequencing to confirm ABE-mediated adenine-to-inosine conversion (inosine is changed to guanine during PCR amplification) at the target site (the opposite strand with a nick cannot be sequenced) and Endo V-catalyzed DNA cleavage at the inosine-containing site (a nicked DNA strand cannot be amplified) (Fig. 2c). When genomic DNA is treated with both ABE7.10 and Endo V, the DNA strand containing inosine is cleaved by Endo V. Note that the DNA strand cleaved by Endo V cannot be amplified by PCR, whereas the DNA strand with no inosine is amplified. As a result, only adenine (and no guanine) is observed at the editing site by Sanger sequencing (Fig. 2c). We also used quantitative real-time PCR to verify targeted DNA cleavage using ABE7.10 plus Endo V (Fig. 2d). Similar results were obtained when hAAG and Endo VIII were used instead of Endo V (Supplementary Fig. 4b,c). The resulting digested genomic DNA samples were subjected to whole-genome sequencing (WGS). After mapping sequencing reads to the human reference genome (hg19), we monitored alignments of sequence reads at the on-target site using Integrative Genomics Viewer (IGV) and observed uniform rather than random alignments of sequencing reads, characteristic of a DSB (Fig. 2e and Supplementary Fig. 4d). No such alignments were seen with untreated genomic DNA or genomic DNA treated with ABE7.10 alone (Fig. 2e and Supplementary Fig. 4d). These results suggest that Digenome-seq can be used for comprehensively mapping genome-wide ABE on-target and off-target sites.

To determine ABE7.10 off-target sites in the human genome, we assigned a DNA cleavage score to each base pair position across the entire genome (Supplementary Fig. 5), using the WGS data, and listed all the sites with high scores (the initial cutoff score was set to 2.5 as done previously (Digenome 1.0)), corresponding to in vitro cleavage sites[8]. Note that the Digenome 1.0 program does not consider sequence homology. We obtained 17 in vitro cleavage sites using Endo V or 18 such sites using hAAG/Endo VIII (Fig. 2f–h). Importantly, 16 sites, including the on-target site, were commonly identified by the two different experimental methods, showing that Digenome-seq is highly reproducible and reliable (Fig. 2f,g). Sequence logos, obtained by comparing DNA sequences at these sites, showed that almost every nucleotide position contributed to the specificity of the ABE7.10 deaminase (Fig. 2h).

We next performed Digenome-seq with each of six additional ABE7.10 RNPs. A total of seven sgRNAs in complex with ABE7.10 plus Endo V cleaved human genomic DNA at 1–28 (8 ± 4 per sgRNA, on average) sites including on-target sites (Supplementary Fig. 6 and Supplementary Tables 2 and 3). Note that we had shown previously that Cas9 nucleases and BE3ΔUGI deaminases (plus USER) cleave genomic DNA at 70 ± 30 sites[8] and 8 ± 3 sites[6], respectively, in the human genome. Among the seven sgRNAs analyzed in this study, two sgRNAs had been combined with BE3ΔUGI and Cas9, and subjected to Digenome analysis previously[6,8], allowing direct comparisons between adenine and cytosine BEs and Cas9

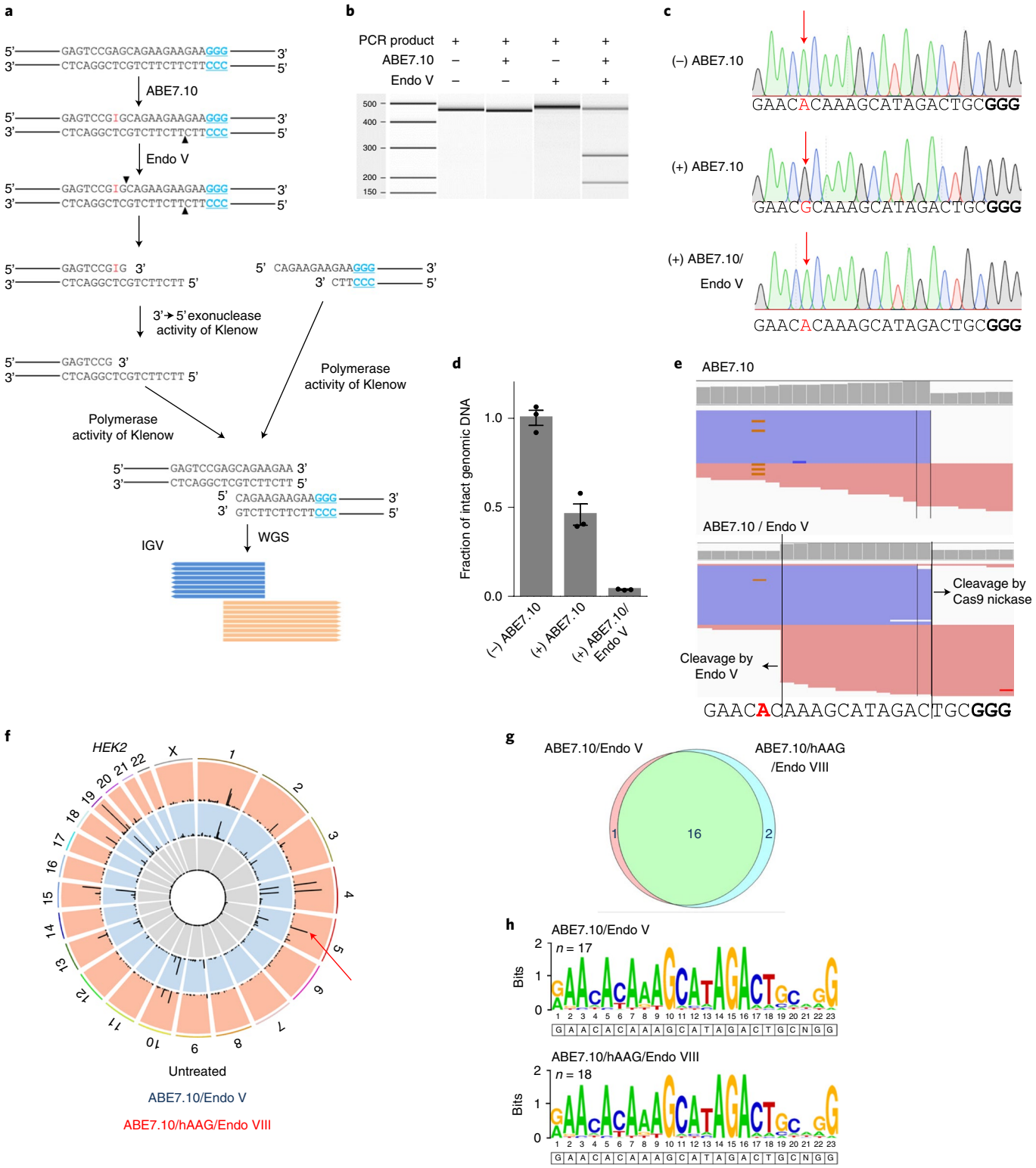


Fig. 2 | Digenome-seq workflow for the identification of genome-wide ABE7.10 off-target sites. a, Overview of Digenome-seq using ABE7.10. ABE7.10 catalyzes the conversion of guanine to inosine. Endo V recognizes inosine and cleaves the second phosphodiester bond 3′ to the inosine. Arrowheads indicate the positions of phosphodiester bonds cleaved by the ABE7.10 nickase and Endo V. **b**, The PCR product containing an ABE7.10 target sequence was cleaved when treated with both ABE7.10 and Endo V. We obtained similar results more than three times when the gel electrophoresis was repeated independently. **c**, Sanger sequencing results showing A-to-G conversion (red arrows) by ABE7.10 and DNA cleavage by Endo V. **d**, Quantitative real-time PCR with reverse transcription (qRT–PCR) results showing DNA cleavage by the ABE7.10 nickase and Endo V. Means ± s.e.m. were from three independent experiments. **e**, IGV images showing the alignments of sequence reads at the *HEK2* target site obtained using WGS data. **f**, Genome-wide Circos plots represent DNA cleavage scores obtained via WGS using intact genomic DNA (gray) and genomic DNA treated with ABE7.10 plus Endo V (blue) or with ABE7.10 plus hAAG and Endo VIII (red). The red arrow indicates the on-target site. **g**, A Venn diagram showing the number of in vitro cleavage sites in the human genome, identified using ABE7.10 plus Endo V or ABE7.10 plus hAAG and Endo VIII. **h**, Sequence logos were obtained using WebLogo by comparing DNA sequences at Digenome-captured sites.

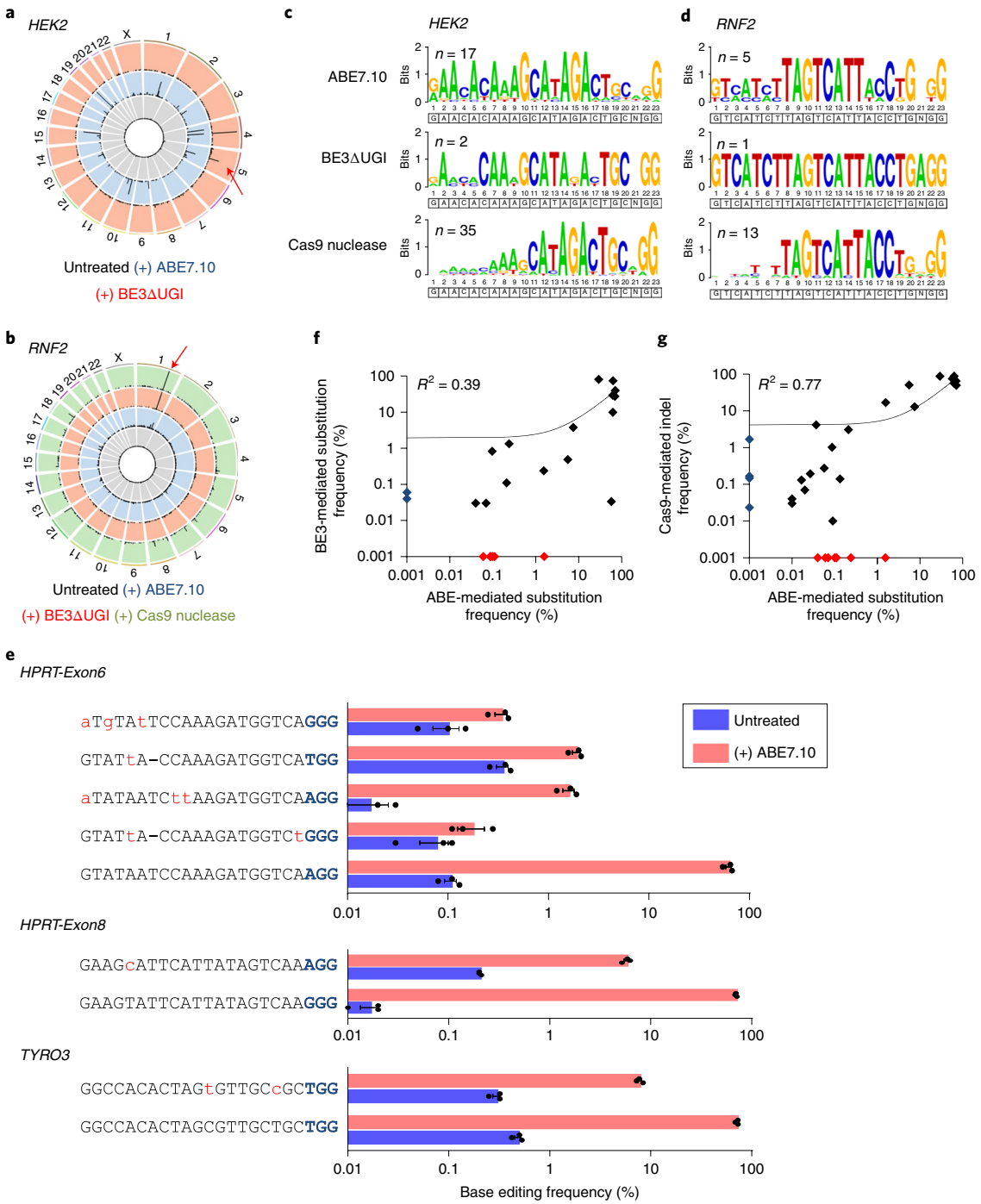


Fig. 3 | Genome-wide specificities of ABE7.10, BE3, and Cas9. a,**b**, Genome-wide Circos plots for *HEK2* (**a**) and *RNF2* (**b**) sites obtained with intact genomic DNA (gray) and genomic DNA treated with ABE7.10/Endo V (blue), BE3ΔUGI/USER (red), or Cas9 (green). Arrows indicate the two on-target sites. **c**,**d**, Sequence logos were obtained using WebLogo by comparing DNA sequences at Digenome-captured sites. **e**, Base editing efficiencies at ABE7.10 off-target sites measured using targeted deep sequencing. PAM sequences are shown in blue. Mismatched bases are shown in red. Dashes indicate RNA bulges. Means ± s.e.m. were from three independent experiments. **f**,**g**, Scatterplots showing correlations between ABE7.10-mediated substitution frequencies and BE3-mediated substitution frequencies ($n = 23$) (**f**), or between ABE7.10-mediated substitution frequencies and Cas9-mediated substitution frequencies ($n = 32$) (**g**). Pearson's correlation coefficient was used to determine the correlation. The blue and red dots indicate off-targets validated for Cas9 or BE3, but not for ABE7.10, and those validated for ABE7.10, but not for Cas9 or BE3, respectively.

(Fig. 3a,b). The *HEK2* sgRNA yielded 17, 2, and 35 in vitro cleavage sites when combined with ABE7.10, BE3, and Cas9, respectively (Fig. 3c). The *RNF2* sgRNA yielded 5, 1, and 13 sites with ABE7.10, BE3, and Cas9, respectively (Fig. 3d). All of the three in vitro cleavage sites identified using two BE3 deaminases were also identified using the respective ABE7.10 deaminases. By contrast, in vitro cleavage sites obtained with ABE7.10 and Cas9 were largely different. Thus, 10 out of 17 (59%) sites obtained with the *HEK2* ABE7.10 were not identified using Cas9. Two out of five (40%) sites obtained with the *RNF2* ABE7.10 were not identified using Cas9. These

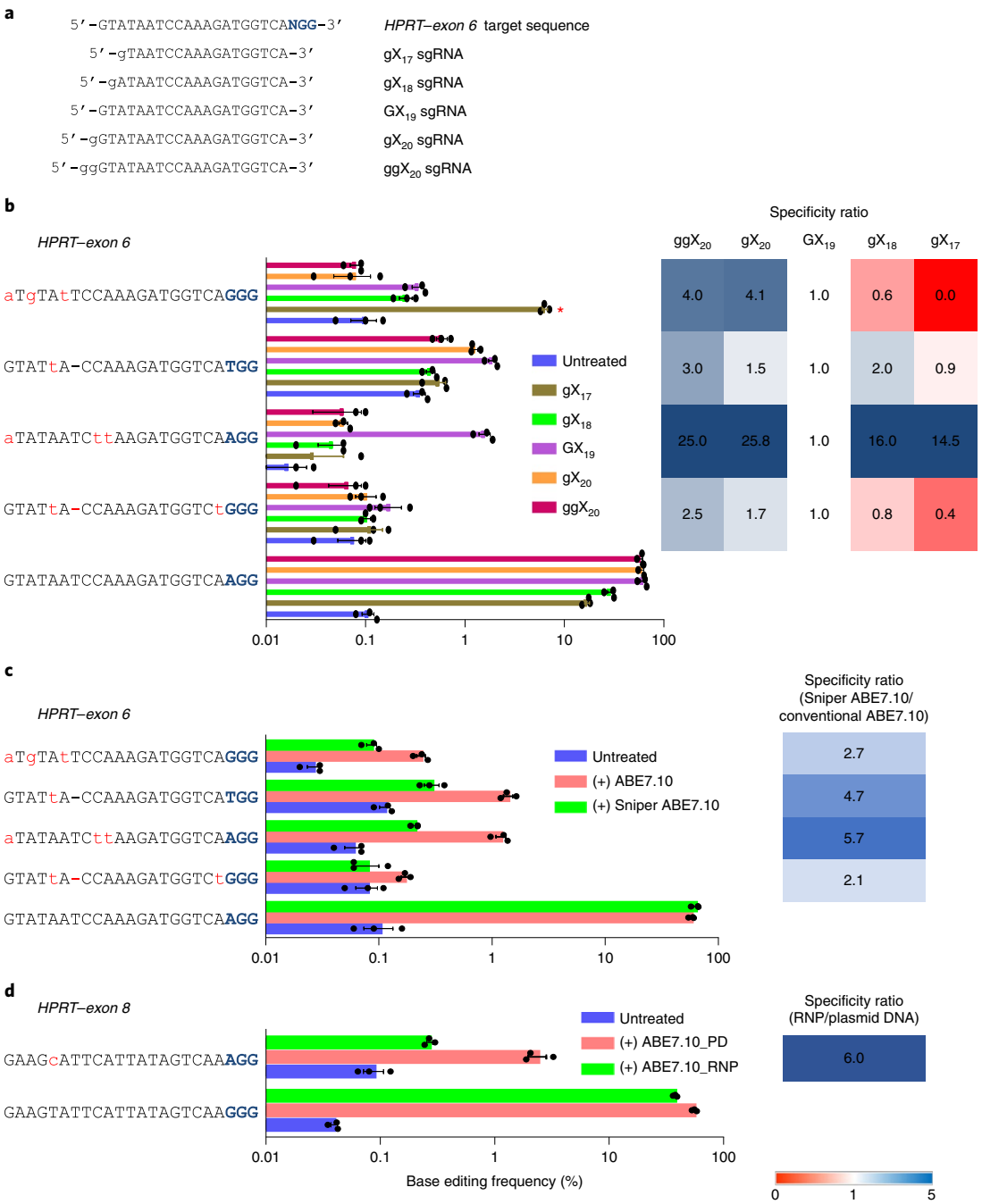


Fig. 4 | Reducing ABE7.10 off-target effects with modified sgRNAs, Sniper ABE7.10, and RNP delivery. a, Target DNA–complementary RNA sequences of conventional (GX_{19}), truncated (gX_{18} or gX_{17}), and extended (gX_{20} or ggX_{20}) sgRNAs. **b**, ABE7.10-mediated base editing frequencies determined by targeted deep sequencing at the *HPRT-exon 6* on- or off-target sites in HEK293T cells. The heatmap shows relative specificities of modified sgRNAs, compared with that of the GX_{19} sgRNA. Truncated sgRNAs reduced both ABE on-target and ABE off-target activity at most sites and exacerbated off-target effects at sites with mismatches at or near the 5′ terminus (asterisk). Specificity ratios were calculated by dividing the specificity of sgRNA of interest (on-target frequency/off-target frequency) by the specificity of the GX_{19} sgRNA (on-target frequency/off-target frequency). **c**,**d**, ABE7.10 specificity was improved by transfecting plasmid DNA encoding Sniper ABE7.10 rather than that encoding ABE7.10 (**c**), or by delivery of ABE7.10 RNP rather than plasmid DNA encoding ABE7.10 (**d**). The heatmap shows relative specificities of Sniper ABE7.10 (**c**) or ABE7.10 RNP (**d**), compared with that of plasmid DNA encoding ABE7.10. Means ± s.e.m. were from three independent experiments.

results suggest that ABEs are in general more specific than Cas9 but can recognize different sets of off-target sites.

We validated ABE off-target base editing at these 57 in vitro cleavage sites identified above using the seven sgRNAs via targeted amplicon sequencing. In addition, we used the Digenome 2.0 program to identify many additional candidate off-target sites with a cutoff score > 0.1, which contained PAM-like sequences (5′-NGN-3′ or 5′-NNG-3′) and had 10 or fewer mismatches, compared with the respective on-target sequences. A total of 12–152 such sites (60 ± 20, on average) were obtained with the seven sgRNAs complexed with ABE7.10, which included all of the 7 on-target sites and 57 in vitro cleavage sites obtained with Digenome 1.0 (Supplementary Tables 2

and 3). We measured base editing frequencies of the seven ABE7.10 deaminases at a total of 193 sites, including those with high sequence homology or with a cleavage score of at least 1.0, via deep sequencing (Fig. 3e and Supplementary Table 4). Single nucleotide substitutions were not measurably detected with frequencies above noise levels (typically in the range of ~0.1–1%) caused by sequencing errors at 173 sites. Among the other 20 validated sites, the seven on-target sites were base edited at frequencies that ranged from 29% to 72%. The other 13 validated off-target sites were edited much less efficiently, at frequencies that ranged from 0.1% to 7.8%. This result shows that Digenome-seq is sensitive enough to capture ABE off-target sites with very low editing frequencies that are near detection limits.

In parallel, we also measured Cas9-induced indel frequencies and BE3-induced substitution frequencies at the 193 sites to compare ABE7.10 with BE3 and Cas9. Unexpectedly, ABE7.10 activities were correlated more strongly with Cas9 activities ($R^2 = 0.77$) than with BE3 activities ($R^2 = 0.39$). Among six sites edited at relatively high frequencies with ABE7.10 but poorly with BE3 (base editing frequencies of 0.001% or lower) (shown as red dots in Fig. 3f), two sites had no cytosine within the BE3 editing window of several nucleotides. Seven sites were edited efficiently with ABE7.10, but not with Cas9 (indel frequencies of 0.001% or lower) (shown as red dots in Fig. 3g). Among these, three sites (43% = 3/7) had a single extra or missing nucleotide compared with their respective on-target sites, possibly forming a DNA or RNA bulge, respectively, when hybridized with their respective sgRNAs. We had also reported previously that such bulge-forming sites are common among BE3 off-target sites[6] but rare among Cas9 off-target sites[8]. Among the 193 sites identified via Digenome-seq using ABE7.10, two or four sites were edited in cells at relatively high frequencies with BE3 or Cas9, respectively, but poorly with ABE7.10 (substitution frequencies of 0.001% or lower) (shown as blue dots in Fig. 3f,g), showing again that ABE7.10, BE3, and Cas9 can recognize different sets of off-target sites in the human genome.

To minimize or avoid ABE off-target activity, we tested three different methods that have been proven useful for reducing Cas9 and BE3 off-target effects: sgRNA modification[5,8–11]; delivery of RNPs in lieu of plasmids[2,16,17]; and use of Sniper–Cas9[12], an engineered Cas9 variant obtained via directed evolution in *E. coli*. First, we replaced our standard GX_{19} sgRNAs with truncated sgRNAs (termed gX_{18} or gX_{17}) (g and G represent, respectively, a mismatched and matched guanine) or extended sgRNAs containing one or two extra guanines at the 5′ terminus (termed gX_{20} or ggX_{20}), and measured editing frequencies in HEK293T cells (Fig. 4a and Supplementary Fig. 7a). Extended sgRNAs reduced ABE off-target activity at almost every site without sacrificing its on-target activity (Fig. 4b, Supplementary Fig. 7b, and Supplementary Table 5). Thus, at the four validated off-target sites identified using a GX_{19} sgRNA targeted to *HPRT–exon 6*, gX_{20} or ggX_{20} sgRNAs showed 2- to 26-fold improvement in specificity ratios, compared with the conventional GX_{19} sgRNA (Fig. 4b). By contrast, truncated sgRNAs reduced both ABE on-target and ABE off-target activity at most sites and exacerbated off-target effects at sites with mismatches at or near the 5′ terminus (shown by asterisk and heatmaps in Fig. 4b and Supplementary Fig. 7b). We also found that use of Sniper ABE7.10, in which Sniper–Cas9 in lieu of the wild-type SpCas9 was fused to the adenine deaminase moiety (Fig. 4c, Supplementary Fig. 8, and Supplementary Table 6), or delivery of ABE7.10 RNPs rather than plasmids improved the specificity of base editing by a factor of up to 7 (Fig. 4d and Supplementary Fig. 9). Combining Sniper ABE7.10 with modified sgRNAs further reduced ABE off-target activity at many validated off-target sites (Supplementary Fig. 10). In this study, we showed, using mismatched sgRNAs, Digenome-seq, and targeted amplicon sequencing, that ABE7.10, BE3, and Cas9 can recognize different sets of off-target sites in the human genome, calling for independent assessments of each tool. The seven ABE7.10 deaminases were reasonably specific, catalyzing A-to-G conversions in vitro at a limited number of sites in the human genome. We also showed that use of modified sgRNAs, RNPs, and Sniper ABE7.10 can reduce or avoid ABE off-target activity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41587-019-0050-1.

Received: 20 September 2018; Accepted: 30 January 2019;
Published online: 04 March 2019

References

1. 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).
2. Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
3. Kang, B. C. et al. Precision genome engineering through adenine base editing in plants. *Nat. Plants* **4**, 427–431 (2018).
4. Liu, Z. et al. Highly efficient RNA-guided base editing in rabbit. *Nat. Commun.* **9**, 2717 (2018).
5. Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).
6. Kim, D. et al. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).
7. Kim, D. et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
8. Kim, D., Kim, S., Kim, S., Park, J. & Kim, J. S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res.* **26**, 406–415 (2016).
9. Cho, S. W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).
10. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).
11. 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).
12. Lee, J. K. et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**, 3048 (2018).
13. 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).
14. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
15. Liu, Z. et al. Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nat. Commun.* **9**, 2338 (2018).
16. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
17. Rees, H. A. et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).

Acknowledgements

This research was supported by grants from the Institute for Basic Science (no. IBS-R021-D1) to J.-S.K. and ToolGen, Inc. (no. 0409-20160107) to Daesik Kim.

Author contributions

J.-S.K. supervised the research. Daesik Kim, Da-eun Kim, G.L., and S.-I.C. performed the experiments. Daesik Kim and Da-eun Kim carried out bioinformatics analyses. J.-S.K. and Daesik Kim wrote the manuscript.

Competing interests

Daesik Kim and J.-S.K. have filed a patent application based on this work. J.-S.K. is a cofounder of, and holds stock in, ToolGen, Inc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-019-0050-1.

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

Correspondence and requests for materials should be addressed to J.-S.K.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Cell culture and transfection. HEK293T cells (ATCC CRL-11268) were cultured in DMEM supplemented with 10% FBS and 1% penicillin–streptomycin (Welgene) and verified using short tandem repeat profiling. HEK293T cells (1.5×10^5) were seeded into 24-well plates and transfected at 70–80% confluency with plasmids encoding ABE (1.5 μg of plasmid DNA), BE3 (1.5 μg), or Cas9 (0.5 μg) together with a plasmid encoding sgRNA (0.5 μg) using Lipofectamine 2000 (Invitrogen; 3 μl for ABEs and BE3 and 2 μl for Cas9). For ABE RNP-mediated genome editing, HEK293T cells (1×10^5) were electroporated with ABE7.10 protein (10 μg) and in vitro–transcribed sgRNA (6 μg) or plasmids encoding ABE (0.6 μg) and sgRNA (0.2 μg) via a Neon Transfection System. Genomic DNA was extracted using a DNeasy Blood & Tissue Kit (Qiagen) 72 h post-transfection.

Expression and purification of ABE7.10 protein. A plasmid encoding the human codon-optimized ABE7.10 with an N-terminal His purification tag was transformed into BL21 Star (DE3)-competent *E. coli* cells. Following transformation, a fresh single colony was grown overnight in Luria–Bertani broth containing 50 μg ml^{-1} kanamycin with shaking at 37 °C. Eight-milliliter precultures diluted at 1:50 were inoculated into individual flasks containing 400 ml of Luria–Bertani broth supplemented with 50 μg ml^{-1} kanamycin; these cultures were incubated with shaking at 37 °C until absorbance A_{600} = ~0.65–0.70; thereafter, the cultures were kept on ice. Protein expression was induced with 1 mM isopropyl-b-D-1-thiogalactopyranoside (IPTG; GoldBio), after which cultures were incubated for 16 hours with shaking at 18 °C.

The subsequent protein purification procedures were carried out at 4 °C. Cells were harvested by centrifugation at 6,000*g* for 10 min and resuspended in lysis buffer (50 mM sodium phosphate (Sigma-Aldrich), 500 mM NaCl (Sigma-Aldrich), 10 mM imidazole (Sigma-Aldrich), 1% Triton X-100 (Sigma-Aldrich), 20% glycerol, 1 mM 1,4-dithiothreitol (DTT; GoldBio), 1 mM phenylmethylsulfonyl fluoride (Sigma-Aldrich), concentration of 1 mg ml^{-1} lysozyme (Sigma-Aldrich), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0). The suspensions were then alternatively incubated in liquid nitrogen, and a water bath (37 °C) was repeated three times. Cells were lysed by sonication for 9 min (5 s (on), 10 s (off)), and the lysates were cleared by centrifugation at 15,000*g* for 20 min. The supernatant was incubated with nickel agarose beads (Ni-NTA; QIAGEN) for 60 min with shaking at 4 °C. The lysate–resin mixture was loaded into a polypropylene column, washed two times with three column volumes of wash buffer (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 35 mM imidazole (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0), and bound protein was eluted with elution buffer (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 250 mM imidazole (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0). The eluted protein fraction was next loaded into a polypropylene column that contained heparin agarose beads (Heparin Sepharose 6 Fast Flow; GE Healthcare) and washed two times with three column volumes of wash buffer (50 mM sodium phosphate (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0). Bound protein was eluted with elution buffer (50 mM sodium phosphate (Sigma-Aldrich), 750 mM NaCl (Sigma-Aldrich), 20% glycerol, 1 mM DTT (GoldBio), 10 μM $ZnCl_2$ (Sigma-Aldrich), pH 8.0). The resulting protein fractions ware concentrated with a centrifuge column (Amicon Ultra-4 Centrifugal Filter Devices; Millipore) at 6,000*g*.

ABE7.10-mediated in vitro digestion of a PCR amplicon. A PCR amplicon containing the *HEK2* site was incubated with the ABE7.10 protein and in vitro-transcribed *HEK2*-targeting sgRNA in a reaction volume of 100 μl for 1 h at 37 °C. The deaminased product was then purified using a PCR purification kit (MGmed). A total of 2 μg of deaminased product was incubated with Endo V (40 units) (New England Biolabs) or hAAG (10 units) (New England Biolabs) and Endo VIII (20 units) (New England Biolabs) in a reaction volume of 200 μl for 30 min at 37 °C, after which the DNA was purified again using a PCR purification kit (MGmed) and subjected to analysis with a Bioanalyzer (Agilent) to determine the results of the in vitro digestion.

ABE7.10-mediated in vitro digestion of genomic DNA. Genomic DNA was isolated from HEK293T cells using a DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. A total of 8 μg of genomic DNA was incubated with the purified ABE7.10 protein (300 nM) and an sgRNA (900 nM) in a reaction volume of 400 μl in reaction buffer (50 mM Tris-HCl (pH 8.0), 25 mM KCl, 2.5 mM $MgSO_4$, 0.1 mM EDTA, 10% glycerol, 2 mM DTT, 10 μM $ZnCl_2$) at 37 °C for 8 hours. The DNA was then purified with a DNeasy Blood & Tissue Kit (QIAGEN) after removal of sgRNA using RNase A (50 μg ml^{-1}). A total of 3 μg of purified DNA was incubated with Endo V (40 units) or hAAG (10 units) and Endo VIII (20 units) in a reaction volume of 200 μl at 37 °C for 2 hours, after which it was purified again with a DNeasy Blood & Tissue Kit (Qiagen). Purified DNA was incubated with KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) and analyzed by real-time quantitative PCR to check for ABE7.10- and Endo V- or ABE7.10-, hAAG-, and Endo VIII-mediated DNA DSBs. The fraction of intact genomic DNA was measured using the comparative C_T method[18].

Whole-genome and Digenome sequencing. A total of 1 μg of in vitro-digested DNA was fragmented to the 400- to 500-base-pair range using the Covaris system (Thermo Fisher Scientific) and incubated with End Repair Mix (Illumina) to generate blunt ends. Fragmented DNA was ligated with adapters to produce libraries and subjected to WGS using a HiSeq X Ten Sequencer (Illumina) at Macrogen. WGS was performed at a sequencing depth of 30–40×. Isaac aligner was used to map the genome sequence. DNA cleavage sites were identified using Digenome 1.0 and Digenome 2.0 programs. The source code of the version of Digenome 2.0 used in this article is available at https://github.com/chizksh/digenome-toolkit2.

Targeted deep sequencing. On-target and potential off-target sites were amplified using KAPA HiFi HotStart PCR polymerase (KK2502). Amplicons were again amplified using TruSeq HT Dual index-containing primers to generate deep sequencing libraries. The libraries were sequenced using Illumina MiniSeq with paired-end sequencing systems. Base editing frequencies indicate the frequencies of modified target sites with at least one edit within the editing window (position 4–7). The source code of the computer program to calculate substitution and indel frequencies is available at https://github.com/ibs-cge/maund.

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

Code availability

The source code used to generate the 2.0 version of Digenome can be accessed at https://github.com/chizksh/digenome-toolkit2. The source code used to calculate substitution and indel frequencies can be accessed at https://github.com/ibs-cge/maund.

Data availability

High-throughput sequencing data have been deposited in the NCBI Sequence Read Archive database with accession code PRJNA515331.

References

18. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108 (2008).

nature research

Corresponding author(s): Jin-Soo Kim

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Isaac aligner (01.15.02.08)
Data analysis	Digenome (https://github.com/chizksh/digenome-toolkit2),Excel, Graph Pad PRISM 7, Adobe Photoshop CC, the program to calculate substitutions and indel frequencies with source code at https://github.com/ibs-cge/maund

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The deep-sequencing datas are available upon request.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size for experiments.
Data exclusions	There are no data excluded.
Replication	The data we showed are replicated three times for cell line tests.
Randomization	Samples were not randomized
Blinding	The investigators were not blinded to group allocation

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T cell line was from ATCC
Authentication	No cell lines were authenticated
Mycoplasma contamination	Cells were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	Not mentioned