nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-017-0004-6

APOBEC3 induces mutations during repair of CRISPR–Cas9-generated DNA breaks

Liqun Lei[1,2,3], Hongquan Chen[4,5], Wei Xue[6], Bei Yang[7], Bian Hu[1,8], Jia Wei[6], Lijie Wang[1,2,3], Yiqiang Cui[4], Wei Li[4], Jianying Wang[4], Lei Yan[2,3,7], Wanjing Shang[1,2,3], Jimin Gao[5], Jiahao Sha[4], Min Zhuang[1], Xingxu Huang[1], Bin Shen[4*], Li Yang[6*] and Jia Chen[1*]

The APOBEC-AID family of cytidine deaminase prefers single-stranded nucleic acids for cytidine-to-uridine deamination. Single-stranded nucleic acids are commonly involved in the DNA repair system for breaks generated by CRISPR-Cas9. Here, we show in human cells that APOBEC3 can trigger cytidine deamination of single-stranded oligodeoxynucleotides, which ultimately results in base substitution mutations in genomic DNA through homology-directed repair (HDR) of Cas9-generated double-strand breaks. In addition, the APOBEC3-catalyzed deamination in genomic single-stranded DNA formed during the repair of Cas9 nickase-generated single-strand breaks in human cells can be further processed to yield mutations mainly involving insertions or deletions (indels). Both APOBEC3-mediated deamination and DNA-repair proteins play important roles in the generation of these indels. Therefore, optimizing conditions for the repair of CRISPR–Cas9-generated DNA breaks, such as using double-stranded donors in HDR or temporarily suppressing endogenous APOBEC3s, can repress these unwanted mutations in genomic DNA.

Members of the apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like or activation-induced cytidine deaminase (APOBEC-AID) family are single-strand-specific cytidine deaminases that are expressed ubiquitously in various cells and tissues and catalyze cytidine-to-uridine (C-to-U) base substitutions in RNA, viral DNA and genomic DNA[1–4] (Fig. 1a). Previously, we reported that the repair of preformed single-strand breaks (SSBs) can induce mutations in flanking DNA through a mechanism that involves the APOBEC3-induced cytidine deamination in single-stranded DNA (ssDNA)[5,6]. Moreover, APOBEC3-mediated mutational signatures were found in the genomic region around DNA double-strand breaks (DSBs)[7,8], which is correlated with the development of various cancers[9–13]. These phenomena indicated that ssDNA regions are formed near SSBs and DSBs and can be vulnerable to the deamination activity of APOBEC3s.

The CRISPR–Cas9 system achieves convenient genome editing by using guide RNAs to direct Cas9 or Cas9 nickase to generate DSBs or SSBs at targeted genomic loci[14–18]. Interestingly, single-stranded nucleic acids are prevalent in CRISPR–Cas9-mediated gene editing processes. For instance, single-guide RNA (sgRNA) is generally used to guide Cas9 to generate DNA breaks at target sites[19]; synthetic single-stranded oligodeoxynucleotides (ssODNs) can be used as homologous donors in HDR for correcting mutations[20,21], and genomic ssDNA regions are also generated during the repair of DSBs and SSBs[6,22,23]. Thus, it is intriguing to examine whether APOBEC3s can target these single-stranded nucleic acids to affect the repair outcome of CRISPR–Cas9-generated DNA breaks.

Here we show that APOBEC3s can target both ssODNs and genomic ssDNAs, which are involved in the repair of CRISPR–Cas9-generated DNA breaks, to trigger C-to-U deamination. The C-to-U deamination in ssODNs ultimately results in base substitution mutations in genomic DNA through HDR. Moreover, the uracil introduced in genomic ssDNA formed during the repair of Cas9 nickase-generated SSBs can be further processed to yield genomic indels. Because multiple DNA-repair proteins play important roles in the generation of these indels, inhibition of these proteins suppresses the formation of indels. Our results indicate that endogenous APOBEC3s can induce unintended mutations during the repair of CRISPR–Cas9-generated DNA breaks and that the mutation frequencies are correlated with APOBEC3 expression. These mutations can be reduced by optimizing repair conditions by, for example, using double-stranded donors in HDR or temporarily suppressing endogenous APOBEC3s.

Results

APOBEC3 induces base substitutions in ssODNs but not in sgRNAs. We first examine whether APOBEC can induce base substitutions in sgRNAs. The deep-sequencing results showed that no obvious base substitutions were determined after sgRNAs were transfected into 293FT cells (Supplementary Fig. 1d,e), in which various APOBEC3s were endogenously expressed (Supplementary Fig. 1a). In addition, overexpression of wild-type APOBEC3B (A3B) in 293FT cells (Supplementary Fig. 1b,c) could not induce base substitutions in sgRNA (Supplementary Fig. 1e). These results

[1]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [2]Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China. [3]University of Chinese Academy of Sciences, Beijing, China. [4]State Key Laboratory of Reproductive Medicine, Department of Histology and Embryology, Nanjing Medical University, Nanjing, China. [5]School of Laboratory Medicine and Life Science, Wenzhou Medical University, Wenzhou, China. [6]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [7]Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, China. [8]MOE Key Laboratory of Model Animal for Disease Study, Model Animal Research Center of Nanjing University, National Resource Center for Mutant Mice, Nanjing, China. Liqun Lei, Hongquan Chen, Wei Xue, Bei Yang and Bian Hu contributed equally to this work. *e-mail: binshen@njmu.edu.cn; liyang@picb.ac.cn; chenjia@shanghaitech.edu.cn

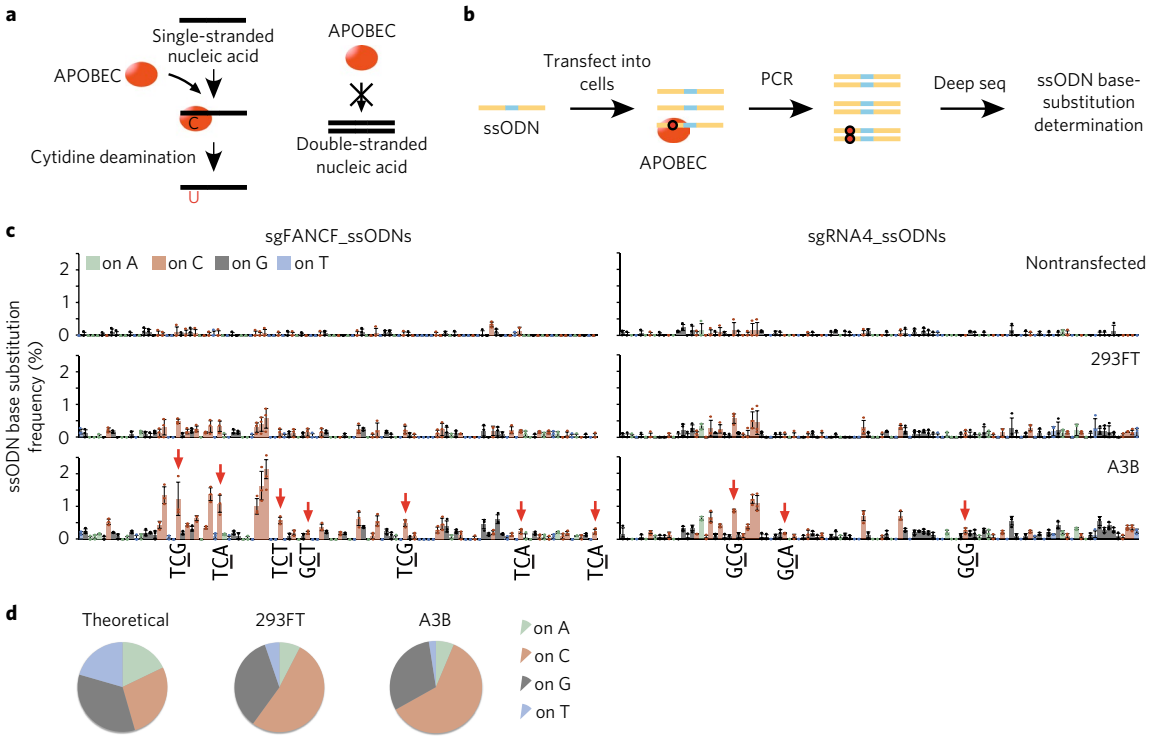

Fig. 1 | APOBEC3 can cause base substitution mutations in ssODNs. a, Schematic diagrams illustrating the hypothesis that APOBEC can cause base substitutions in ssODNs but not in dsODNs. **b**, Schematic diagrams illustrating the procedures to test whether APOBEC can cause base substitution in ssODNs in cells. **c**, Base substitution frequency of each base in ssODNs that are either nontransfected or transfected into 293FT cells or A3B cells. Red arrows highlight the same base substitutions as those observed in the ODN-cognate genomic regions in Fig. 2. Data shown are means ± s.d. from three independent experiments. **d**, Comparison of the theoretical base-substitution fractions calculated from the base content of ssODNs and the experimentally determined ones in 293FT and A3B cells. Source data files for **c**,**d** are available online.

suggest that APOBEC3s probably do not access sgRNA in vivo, possibly owing to the compact secondary structure of sgRNA.

In contrast, obvious base substitutions were detected in ssODNs that were transfected into 293FT cells when compared with untransfected ssODNs (Fig. 1b,c). Importantly, the majority of these detected base substitutions were on Cs (Fig. 1c,d), which suggests a role of APOBEC3 in triggering ssODN base substitution. To further prove this speculation, ssODNs were transfected into the A3B cell line or another 293FT cell line overexpressing a catalytically inactive A3B mutant (A3Bm)[24] (Supplementary Fig. 1b). As expected, the frequency of base substitutions in the ssODNs increased considerably in the A3B cell line with typical APOBEC3 signatures but not in the A3Bm cell line (Fig. 1c and Supplementary Fig. 1f,g). In contrast, no such base substitution was detected in double-stranded ODNs (dsODNs) after being transfected into A3B cells (Supplementary Fig. 1f). These results are consistent with the disfavor for dsODNs as a substrate by APOBEC3s[1] (Fig. 1a) and demonstrate that APOBEC3s prefer to mediate base substitutions in ssODNs, depending on their cytidine deaminase activity.

APOBEC3-induced base substitutions in ssODNs can be integrated into genomic DNA. Through HDR at CRISPR–Cas9-generated DSB sites, original genomic sequences are replaced by sequences of the exogenous ssODN donors[20,21]. In this case, the APOBEC3-mediated C-to-U substitutions in ssODNs (Fig. 1b,c) could be incorporated into the genome and eventually lead to unwanted mutations in genome (Fig. 2a). To test this possibility, ssODN-cognate genomic regions were specifically amplified and sequenced to examine the HDR outcome from 293FT, A3B and A3Bm cells (Supplementary Fig. 2a–c). The occurrences of APOBEC3-featured base substitutions in the examined ssODN-cognate genomic regions were prominent in A3B cells but barely detectable in A3Bm cells and 293FT cells (Fig. 2b, compare A3B_ssODNs to 293FT_ssODNs and A3Bm_ssODNs). A similar mutation pattern in cytosines was detected in the examined ssODN-cognate genomic regions (Fig. 2b, red arrows) as observed in the ssODN donors (Fig. 1c, red arrows). Because APOBEC3s disfavor double-stranded nucleic acids as substrate[1] (Fig. 1a and Supplementary Fig. 1f), we hypothesized that using double-stranded nucleic acids as the homologous donor could substantially inhibit APOBEC3-mediated base substitutions. Indeed, when dsODNs or double-stranded plasmids (ds-plasmids) were used as the homologous donor for Cas9-mediated HDR, APOBEC3-featured base substitutions in the cognate genomic regions were nearly undetectable, even in A3B cells (Fig. 2b, compare A3B_ssODNs to A3B_dsODNs and A3B_ds-plasmids).

APOBEC3s are retroviral restriction factors and are highly expressed in activated immune cells[2,25] (Supplementary Fig. 3a). To test the effects of endogenous APOBEC3s on Cas9-mediated HDR in more physiologically relevant conditions, we electroporated sgRNA–Cas9 ribonucleoprotein (RNP) complexes along with ssODNs or dsODNs into activated primary human T cells that were pretreated with control siRNAs or siRNAs against endogenous APOBEC3s (Supplementary Fig. 2d,e). When ssODNs were used in Cas9-mediated HDR, we identified four (two at C and two at G) prominent base-substitution hotspots (Fig. 2c, red arrows and triangles). The C substitutions identified in the cognate genomic DNA were the same ones detected in the ssODNs (compare Fig. 2c to Fig. 1c, red arrows), and these substitutions were clearly suppressed when dsODNs were used as donors or when the expression of endogenous APOBEC3s was inhibited (Fig. 2c, compare top left panel to the others, and Supplementary Fig. 3b). These

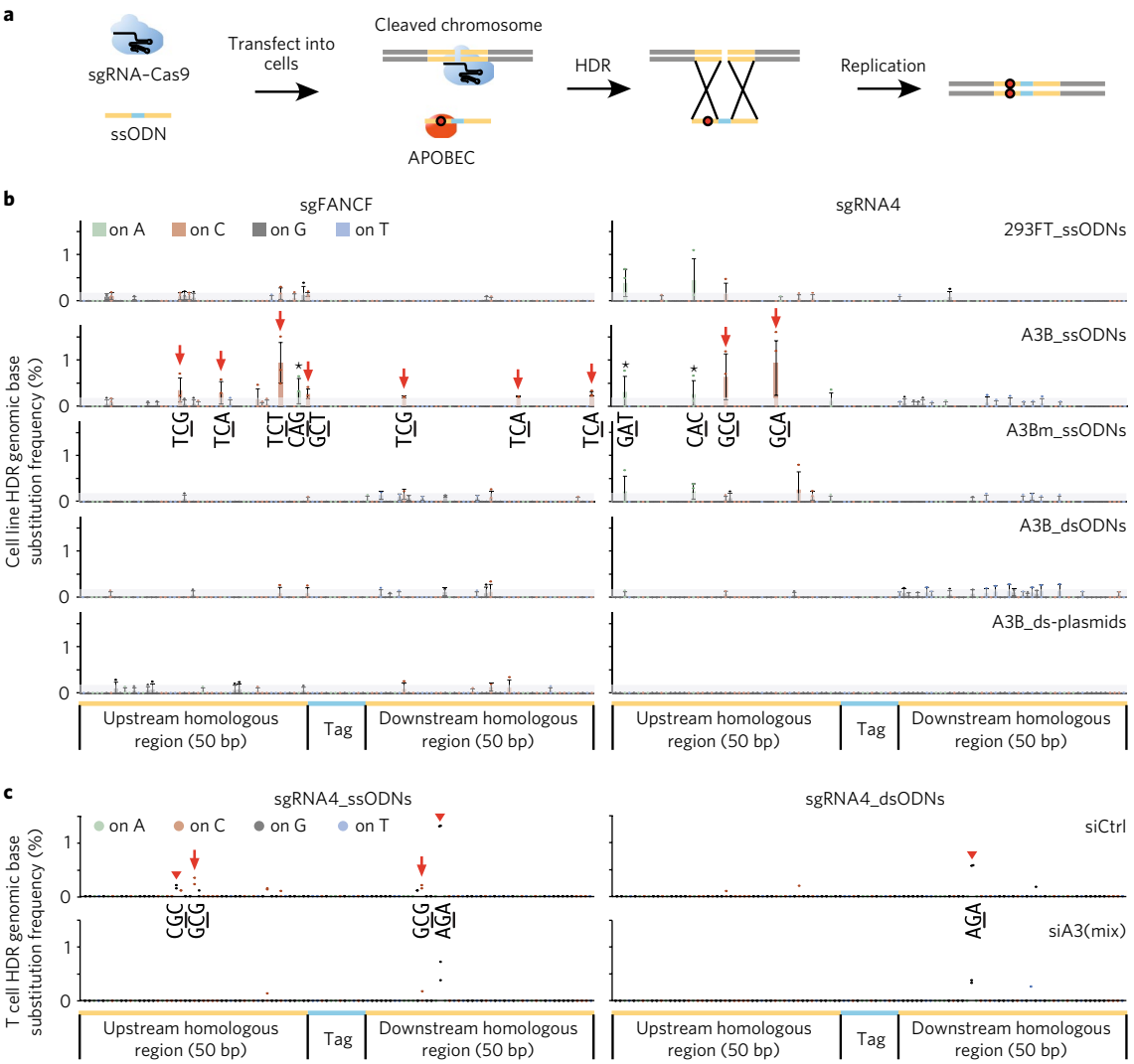

Fig. 2 | APOBEC3-mediated base substitution mutations in ssODNs can be integrated into genomic DNA during HDR. a, Schematic diagrams illustrating how APOBEC-generated base substitutions in ssODNs can be integrated into genomic DNA during HDR. **b**, Base substitution frequency of each base in the ODN-cognate genomic DNA regions. 293FT, A3B or A3Bm cells were cotransfected with Cas9 and the indicated sgRNAs in the presence of ssODNs, dsODNs or ds-plasmid donors, after which the base substitution frequencies were measured using deep sequencing. Background base substitution frequency: 0.17% (light-gray shadow, see Methods). Data shown are means ± s.d. from three independent experiments. **c**, Base substitution frequency of each base in the ODN-cognate genomic DNA regions in primary human T cells. Primary human T cells were pretreated with control siRNA (siCtrl) or siRNA against endogenous APOBEC3s (siA3(mix)) and then cotransfected with Cas9 and the indicated sgRNAs in the presence of ssODNs or dsODNs. Data shown are from two independent experiments. Red arrows indicate the same APOBEC3-mediated base substitutions as those observed in ssODNs in Fig. 1c. Red triangles indicate base substitutions mediated by APOBEC3s in complementary genomic ssDNA. Asterisk (*) indicates base substitution from an undetermined source. Source data files for **b**,**c** are available online.

results further confirm that base substitutions in ssODNs triggered by endogenous APOBEC3s can be integrated into genomic DNA during HDR. On the other hand, base substitutions at the Gs were also associated with the expression level of endogenous APOBEC3s (Fig. 2c and Supplementary Fig. 3b). APOBEC3s can trigger hypermutation of Gs in the HIV genome[26] and cancer genome[9–11] by deaminating Cs in the strand opposing the reference strand. Considering that ssDNA regions are generated during homology-dependent DSB repair[22], the two APOBEC3-correlated substitutions of Gs were therefore probably transformed from C deamination in the complementary genomic ssDNA formed during HDR[8] (Supplementary Fig. 3c). Together, our results indicate that APOBEC3-mediated deamination can introduce unexpected base substitutions in genomic DNA during HDR in both cultured 293FT and primary human T cells, whereas the application of dsODNs or ds-plasmids as HDR donors suppresses these APOBEC3-mediated base substitutions (Fig. 2b,c).

Notably, not all base substitutions in ssODNs could be integrated into genomic DNA (compare Fig. 2b,c with Fig. 1c). This finding could be explained by a recent study showing that homologous recombination is tolerant of some mismatches in the homology region, but excessive mismatches eliminate recombination[27]. Furthermore, several base substitutions at As were also observed in the genomic regions when ssODNs were used as the HDR donor (Fig. 2b, asterisks). How these base substitutions are triggered still awaits further investigation.

APOBEC3 mediates the formation of indels during the repair of SSBs generated by Cas9 nickase. Endogenous genomic ssDNAs can also be generated during the repair of SSBs, as

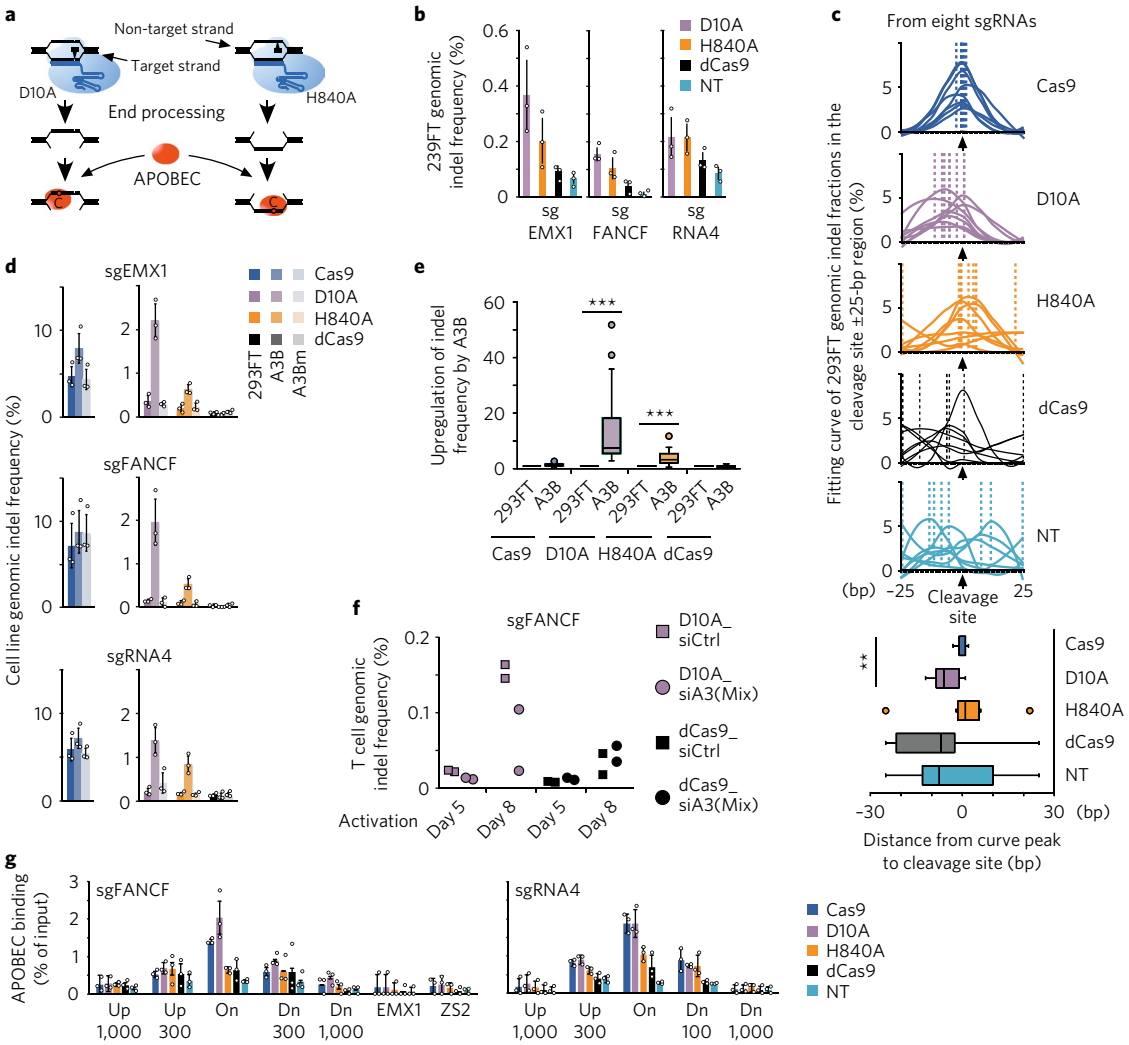

Fig. 3 | APOBEC3 mediates the formation of unexpected indel mutations near SSBs. a, Schematic diagrams illustrating the hypothesis that APOBEC can access and deaminate the genomic ssDNA formed during the repair of Cas9 nickase-generated SSBs. **b**, Genomic indel frequencies at sgRNA target sites in either nontransfected (NT) 293FT cells or 293FT cells cotransfected with the indicated sgRNAs and D10A, H840A or dCas9. Data shown are means ± s.d. from three independent experiments. **c**, Distribution curves of Cas9-variant-induced indels at a region 25 bp upstream and downstream of the cleavage site and statistical analysis of the distances from the cleavage site to curve peaks. Distribution curves from eight sgRNAs are shown in Supplementary Fig. 7a. **d**, Indel frequencies induced by Cas9-variant-generated breaks in 293FT, A3B and A3Bm cells. Data shown are means ± s.d. from three independent experiments. **e**, Statistical analysis of indel frequencies in different cells illustrating the upregulation of Cas9 nickase-induced indels by APOBEC3B overexpression. The indel frequencies induced by Cas9 variants in 293FT cells were set to 1. **f**, Indel frequencies induced by D10A-generated SSBs at indicated activation time points in primary human T cells pretreated with control siRNA (siCtrl) or siRNA against endogenous APOBEC (siA3(Mix)). Data shown are from two independent experiments. **g**, ChIP-qPCR assays determining the binding of APOBEC3B at indicated genomic loci in A3B cells treated with indicated Cas9 variants and sgRNAs. Data shown are means ± s.d. from three independent experiments. For box plots in **c**,**e**, midlines indicate medians, box edges show the interquartile range (IQR), whiskers show 1.5 × IQR, and circles indicate outliers. $^{**}P < 0.01$, $^{***}P < 0.001$, one-tailed Student's *t* test. Source data files for **b**–**g** are available online.

was recently suggested[5,23]. We therefore examined whether APOBEC3s could directly induce mutations in genomic ssDNAs that are formed during the repair of SSBs generated by the Cas9 nickases D10A and H840A (Fig. 3a); Cas9 and catalytically inactive Cas9 (dCas9) were included as controls. APOBEC3-featured base substitutions were first analyzed near a series of Cas9 nickase-generated SSB sites (Supplementary Fig. 4a). Surprisingly, one APOBEC3-featured base substitution was observed only near the SSB introduced by sgVEGFA-D10A (Supplementary Fig. 4b, red arrow) and not in the majority of the other SSBs examined (Supplementary Fig. 4b, compare D10A and H840A with Cas9 and dCas9). The base substitution detected near the sgVEGFA-D10A-generated SSB was further upregulated by the overexpression of catalytically active A3B (Supplementary Fig. 4c, red arrows).

Different from sparse base substitutions, indels were commonly detected near the examined Cas9 nickase-generated SSBs in both 293FT cells and HeLa cells (Fig. 3b and Supplementary Fig. 5a,b). Similar results were also obtained using an episomal shuttle vector system[5] followed by Sanger sequencing (Supplementary Fig. 6a–c). It is worth noting that single Cas9 nickase and D10A-APOBEC (AID) base editors (BEs) can both produce indels[28–36], although the underlying mechanism for these indel formations has not been extensively examined.

We then sought to understand the source of these unexpected indels. As indels are normally introduced when DSBs are resolved

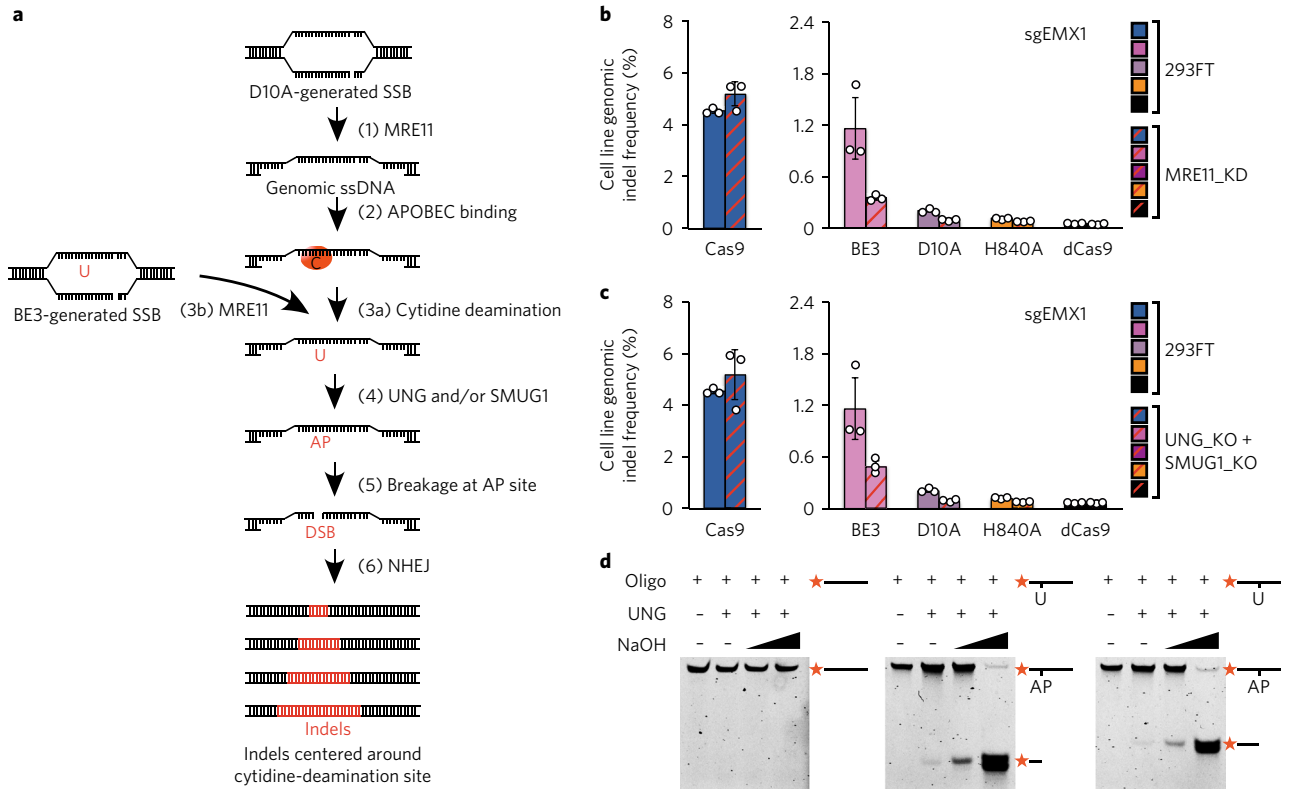

Fig. 4 | APOBEC3-catalyzed deamination of genomic ssDNA formed during SSB repair can be further processed to generate indel mutations. a, Schematic diagrams illustrating the potential pathway by which D10A- or BE3-generated SSBs can trigger indel formation. Steps 1–5 are supported by results in this study. **b**, Indel frequencies induced by Cas9 variants and BE3 in 293FT and MRE11-knockdown (MRE11_KD) cells. **c**, Indel frequencies induced by Cas9 variants and BE3 in 293FT and *UNG* and *SMUG1* double-knockout (UNG_KO + SMUG1_KO) cells. Data shown in **b**,**c** are means ± s.d. from three independent experiments. **d**, Representative gel images showing that the AP site generated from the UNG-catalyzed removal of U leads to spontaneous breakage of DNA oligos, with or without alkali treatment. Uncropped gel images for **d** are shown in Supplementary Dataset 1. Gels shown are representative of three experiments. Source data files for **b**,**c** are available online.

through nonhomologous end joining (NHEJ)[22], the distribution spectrum of indels can reflect the position of DSBs[28,37]. Cas9-generated indels were included as a control for these analyses (Fig. 3c and Supplementary Figs. 5c, 6d and 7). As expected, indels induced by Cas9 all centered on Cas9-generated DSB sites[28,37] (Fig. 3c and Supplementary Figs. 5c and 7a, Cas9), whereas indels induced by D10A or H840A did not center on the corresponding SSBs (Fig. 3c and Supplementary Figs. 5c and 7a, compare Cas9 with D10A or H840A). Similar results were also obtained with episomal shuttle vector systems (Supplementary Fig. 6d). Interestingly, indels originating from D10A-generated SSBs all peaked around Cs in the non-target strand, whereas indels originating from H840A-generated SSBs all peaked around Gs in the non-target strand (Cs in the target strand) (Supplementary Figs. 6d and 7a). These findings are similar to observation of strand-coordinated mutation clusters (C clusters and G clusters) mediated by APOBEC3s near chromosomal rearrangement breakpoints in cancer genomes[9,10]. Moreover, the frequency of SSB-induced indels was significantly upregulated by the overexpression of A3B but not by that of A3Bm (Fig. 3d,e and Supplementary Fig. 5d). In contrast, A3B overexpression did not induce indel formation in the absence of an SSB (Supplementary Fig. 5e), and the frequency of DSB-induced (Cas9) or background (dCas9) indels remained largely unchanged across 293FT, A3B and A3Bm cells (Fig. 3d,e and Supplementary Fig. 5d). Notably, when the frequencies of these indels were stimulated by A3B overexpression, their distribution still peaked around Cs but not SSBs (compare Fig. 3c and Supplementary Fig. 7b). Collectively, these data indicate a link between APOBEC3s and indel formation observed near the examined Cas9 nickase-generated SSBs.

The correlation between APOBEC3s and indels induced by Cas9 nickase-generated SSBs was further validated in primary human T cells. In conjunction with the physiologically upregulated expression of APOBEC3s during the activation of primary human T cells (Supplementary Fig. 3a), the indel frequency induced by D10A-generated SSBs was also upregulated (Fig. 3f and Supplementary Fig. 8b). Suppression of endogenous expression of APOBEC3s with siRNAs consistently and evidently repressed the indel frequencies induced by D10A-generated SSBs at each examined T cell differentiation time point (Fig. 3f and Supplementary Fig. 8b). As a control, indels induced by Cas9-generated DSBs were not suppressed by APOBEC3 knockdown (Supplementary Fig. 8a). Similar effects of endogenous APOBEC3 knockdown on Cas9 nickase-induced indel formation were also captured using episomal shuttle vectors systems in 293FT cells (Supplementary Fig. 8c). These results indicate that APOBEC3s can mediate indel formation in genomic DNA near SSB sites.

Next, we performed chromatin immunoprecipitation (ChIP) assays to confirm the binding of APOBEC3s to genomic ssDNA regions generated near Cas9 nickase-cleaved SSBs during the repair process in A3B cells. Compared with their low, albeit even, distribution in nontransfected cells, the A3B-binding signals were significantly enriched at the sgRNA target sites in the presence of Cas9 variants (Fig. 3g). A3B also bound at a modest strength to the region ~300 base pairs (bp) upstream or downstream of the sgRNA target site (Fig. 3g, Up300 and Dn300), which suggests that

the ssDNA region can extend a few hundred base pairs away from the SSB. As a control, the binding of A3B was barely detectable in the regions ~1,000 bp from the cleavage sites or nonrelevant sites (Fig. 3g). Additionally, the binding of histone H3 was universally detected at all examined sites, and there was no obvious variation (Supplementary Fig. 8d). These ChIP results thus confirm the direct binding of APOBEC3s to genomic ssDNA regions generated near SSBs.

Mechanism of APOBEC3-mediated indel formation. The above results indicate that APOBEC3s can mediate indel formation near Cas9 nickase-generated SSBs, probably by deaminating the single-stranded genomic DNA regions formed during the repair of these SSBs. How APOBEC3-dependent C-to-U deamination eventually leads to indels near Cas9 nickase-generated SSB sites was unknown. Mechanistically, Cas9 nickase-introduced SSBs could be processed by DNA exonucleases to generate genomic ssDNA[5,23] (Fig. 4a, step 1). APOBEC3s could then bind and deaminate Cs in these ssDNA regions to generate Us[1,2] (Fig. 4a, steps 2 and 3a). Alternatively, a U-containing genomic ssDNA could also be generated by DNA exonucleases in the case of the third generation of BE (BE3, D10A-APOBEC fusion protein) (Fig. 4a, step 3b). The unusual U in the genomic ssDNA would then be recognized and transformed to an apurinic or apyrimidinic (AP) site by various DNA glycosylases (Fig. 4a, step 4), such as UNG and SMUG1 (ref. [38]). Afterward, the AP-site-containing ssDNA may undergo AP endonuclease 1 (APE1)-catalyzed cleavage or spontaneous breakage[2], resulting in a DSB at the APOBEC3 deamination site (Fig. 4a, step 5). Finally, the DSB could be repaired by NHEJ to induce indels at the APOBEC3-mediated deamination site (Fig. 4a, step 6).

A series of enzymes that may function at each step during the indel formation were further examined to test this hypothesis. Two major DNA exonucleases, Exo1 and MRE11, are responsible for DNA end resection[39]. The inhibition of MRE11 (Supplementary Fig. 9a) significantly suppressed indel formation (Fig. 4b and Supplementary Fig. 9e) around both BE3- and Cas9 nickase-generated SSBs, whereas knocking out *Exo1* (Supplementary Fig. 9b) did not affect indel formation in either case (Supplementary Fig. 9f). These results therefore indicate that MRE11 is involved in the step of generating a genomic ssDNA region from an SSB in our system (Fig. 4a, step 1). A double knockout of *UNG* and *SMUG1* (Supplementary Fig. 9c) accordingly decreased the indel formation (Fig. 4c and Supplementary Fig. 9g), confirming that AP-site formation catalyzed by UNG and SMUG1 is involved in the indel formation process (Fig. 4a, step 4). Although APE1 can cleave AP-site-containing DNA[38], *APE1* knockout (Supplementary Fig. 9d) did not affect indel formation (Supplementary Fig. 9h), which suggests that other enzymes or conditions participate in this step. We thus examined the stability of AP-site-containing oligo DNAs in the absence or presence of alkalinity[40] and found that AP-site-containing oligo DNAs were spontaneously broken at the AP site, even without alkali treatment (Fig. 4d). These results indicate that AP-site-containing ssDNA itself is unstable and prone to breakage, thereby generating a DSB. Under the tested conditions, BE3 responded in a manner similar to that of D10A (Fig. 4b,c and Supplementary Fig. 9e–h), which suggests that BE3 and D10A share the same pathway for the induction of indel formation (Fig. 4a).

In most, if not all, of the examined cases, D10A induced higher indel frequencies than did H840A (Figs. 3 and 4 and Supplementary Fig. 5). Because the major difference between D10A and H840A is that D10A nicks the target strand, whereas H840A nicks the non-target strand[19] (Supplementary Fig. 10a,b), we speculated that the different indel frequencies could result from their different strand preferences (Supplementary Fig. 10e). In the case of D10A-mediated SSBs, APOBEC3s bound to the exposed non-target strand for C-to-U base substitutions, which ultimately resulted in DSBs and indels (Supplementary Fig. 10e, left). Instead, in the case of H840A-mediated SSBs, the non-target strand was nicked and resected, leaving the target strand single stranded. The single-stranded target strand could then be rebound by the sgRNA–H840A complex (Supplementary Fig. 10c,d) and therefore protected from APOBEC3-induced indel formation (Supplementary Fig. 10e, right).

Discussion

In this study, we showed that APOBEC3s can cause unexpected mutations (including both base substitutions and indels) during the repair of DNA breaks generated by CRISPR–Cas9, indicating a previously underappreciated crosstalk between APOBEC and CRISPR–Cas9-triggered DNA repair. Through catalyzing cytidine deamination, APOBEC3s cause base substitutions in ssODN donors (Fig. 1), which can be further incorporated into genomic DNA via HDR (Fig. 2). Although relatively low in somatic cells, the frequencies of APOBEC3-mediated base substitutions in genomic DNA are correlated with the activities of APOBEC3s and can be upregulated by APOBEC3 overexpression (Figs. 1 and 2). Our results also demonstrated that APOBEC-mediated cytidine deamination in the genomic ssDNA region formed around SSB can be further processed into unexpected indels (Fig. 3), in conjunction with a specific DNA repair process (Fig. 4). In this process, the base excision repair of uracils in genomic ssDNA results in a DSB, which is further processed into indels through NHEJ. Interestingly, Cas9 nickase-induced indel formation has been implicated in previous research[28–31], though the underlying mechanism was unclear. We show here that these unexpected indels are probably generated through the MRE11-APOBEC-UNG (SMUG1)-spontaneous breakage-NHEJ pathway (Fig. 4a). Correspondingly, suppressing endogenous APOBEC3s, MRE11 or UNG (SMUG1) can substantially decrease the formation of these unwanted indels (Figs. 3f and 4b,c).

In general, the frequency of APOBEC3-mediated mutations during the repair of CRISPR–Cas9-generated DNA breaks is low, and these unintended mutations introduced by APOBEC3s during genome editing are unlikely to cause side effects in the context of basic scientific research experiments. However, these APOBEC3-mediated mutations are correlated with the APOBEC3s' activities and may be problematic in cells or tissues with a high APOBEC3 expression level. Although it is practical to sequence the edited genomic loci and then select single clones without these unintended mutation byproducts in basic research experiments, it is usually difficult if not impossible to isolate single clones for postnatal gene correction of tissues or organs[21,41–43]. Thus, the unintended mutations mediated by APOBEC3s should be avoided as much as possible when performing genome editing in tissues or organs, especially those with high APOBEC3 activity[21,44]. Our results show that dsODN donors are largely resistant to the APOBEC-triggered deamination in Cas9-mediated HDR (Fig. 2 and Supplementary Fig. 1f), suggesting the utilization of dsODN instead of ssODN as homologous donors for optimized gene correction in postnatal tissues or organs.

Importantly, the APOBEC-mediated C-to-T base substitution has been harnessed to perform programmable base editing by conjugation with the CRISPR–Cas9 system[32–36,45–52]. However, as APOBEC is recruited to the proximity of SSB to perform C-to-U deamination in BE3-mediated base editing, unwanted indel byproducts are also induced[32–36] via a previously unclear mechanism[53,54]. In this study, we showed that BE3 induces unintended indels with the same MRE11-UNG (SMUG1)-spontaneous breakage-NHEJ pathway (Fig. 4a). During the submission of this paper, we and others have independently developed the enhanced base editors (eBEs) and BE4, both of which induce fewer indels and achieve higher C-to-T base-editing frequencies by using additional UNG inhibitor

to further inhibit the base excision repair pathway[55,56]. Such studies and our findings reported here together serve as proof of principle that suppressing the unintended mutagenesis mediated by APOBEC during the DNA repair process might help to improve gene editing fidelity.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41594-017-0004-6.

Received: 17 August 2017; Accepted: 2 November 2017;
Published online: 11 December 2017

References

1. Harris, R. S. & Liddament, M. T. Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* **4**, 868–877 (2004).
2. Henderson, S. & Fenton, T. APOBEC3 genes: retroviral restriction factors to cancer drivers. *Trends Mol. Med.* **21**, 274–284 (2015).
3. Salter, J. D., Bennett, R. P. & Smith, H. C. The APOBEC protein family: united by structure, divergent in function. *Trends Biochem. Sci.* **41**, 578–594 (2016).
4. Yang, B., Li, X., Lei, L. & Chen, J. APOBEC: From mutator to editor. *J. Genet. Genomics* **44**, 423–437 (2017).
5. Chen, J., Miller, B. F. & Furano, A. V. Repair of naturally occurring mismatches can induce mutations in flanking DNA. *eLife* **3**, e02001 (2014).
6. Chen, J. & Furano, A. V. Breaking bad: the mutagenic effect of DNA repair. *DNA Repair (Amst.)* **32**, 43–51 (2015).
7. Roberts, S. A. et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
8. Taylor, B. J. et al. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3 A in breast cancer kataegis. *eLife* **2**, e00534 (2013).
9. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
10. Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* **45**, 977–983 (2013).
11. Roberts, S. A. et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
12. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15**, 585–598 (2014).
13. Chan, K. & Gordenin, D. A. Clusters of multiple mutations: incidence and molecular mechanisms. *Annu. Rev. Genet.* **49**, 243–267 (2015).
14. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).
15. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
16. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).
17. Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).
18. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
19. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
20. Wu, Y. et al. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* **13**, 659–662 (2013).
21. Yin, H. et al. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* **32**, 551–553 (2014).
22. Symington, L. S. & Gautier, J. Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet.* **45**, 247–271 (2011).
23. Myler, L. R. et al. Single-molecule imaging reveals the mechanism of Exo1 regulation by single-stranded DNA binding proteins. *Proc. Natl. Acad. Sci. USA* **113**, E1170–E1179 (2016).
24. Burns, M. B. et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366–370 (2013).
25. Refsland, E. W. et al. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res.* **38**, 4274–4284 (2010).
26. Harris, R. S. et al. DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803–809 (2003).
27. Anand, R., Beach, A., Li, K. & Haber, J. Rad51-mediated double-strand break repair and mismatch correction of divergent substrates. *Nature* **544**, 377–380 (2017).
28. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
29. 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).
30. 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).
31. Tsai, S. Q. et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**, 569–576 (2014).
32. 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).
33. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
34. Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
35. Zhou, C. et al. Highly efficient base editing in human tripronuclear zygotes. *Protein Cell* (2017).
36. Li, G. et al. Highly efficient and precise base editing in discarded human tripronuclear embryos. *Protein Cell* **8**, 772–775 (2017).
37. Paquet, D. et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
38. Carter, R. J. & Parsons, J. L. Base excision repair, a pathway regulated by posttranslational modifications. *Mol. Cell. Biol.* **36**, 1426–1437 (2016).
39. Daley, J. M., Niu, H., Miller, A. S. & Sung, P. Biochemical mechanism of DSB end resection and its regulation. *DNA Repair (Amst.)* **32**, 66–74 (2015).
40. Starrett, G. J. et al. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat. Commun.* **7**, 12918 (2016).
41. Long, C. et al. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* **351**, 400–403 (2016).
42. Nelson, C. E. et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**, 403–407 (2016).
43. Tabebordbar, M. et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**, 407–411 (2016).
44. Bonvin, M. et al. Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology* **43**, 1364–1374 (2006).
45. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
46. Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol. Plant* **10**, 526–529 (2017).
47. Liang, P. et al. Correction of β-thalassemia mutant by base editor in human embryos. *Protein Cell* **8**, 811–822 (2017).
48. Lu, Y. & Zhu, J. K. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* **10**, 523–525 (2017).
49. 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).
50. Shimatani, Z. et al. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017).
51. Zhang, Y. et al. Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat. Commun.* **8**, 118 (2017).
52. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
53. Hess, G. T., Tycko, J., Yao, D. & Bassik, M. C. Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* **68**, 26–43 (2017).
54. Mitsunobu, H., Teramoto, J., Nishida, K. & Kondo, A. Beyond native Cas9: manipulating genomic information and function. *Trends Biotechnol.* **35**, 983–996 (2017).
55. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T: Abase editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
56. Wang, L. et al. Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res.* **27**, 1289–1292 (2017).

Acknowledgements

We are grateful to A. Furano, H. Lin and H. Wang for discussing and commenting on this manuscript, L.-L. Chen and N. Jing for technical support, X. Li and Y. Pan for participating in the examination of APOBEC expression, J. Wu for maintaining cell lines and H. Fang for participating in deep-sequencing library preparation. Next-generation deep sequencing was performed at the CAS-MPG PICB Omics Core, Shanghai, China. This work is supported by a MOST grant (2014CB910600 to L. Yang), NSFC grants (91540115 to L. Yang, 31571372 to B.S., 31471241 to L. Yang, 31600619 to B.Y. and 31600654 to J.C.), the Shanghai Pujiang program (16PJ1407000 to J.C. and 16PJ1407500

to B.Y.) and CAS Key Laboratory of Computational Biology grants (2015KLCB01 and 2016KLCB01 to L. Yang and J.C.).

Author contributions

J.C., L. Yang and B.S. conceived, designed and supervised the project. L.L., H.C., B.Y. and B.H. performed most of the experiments with the help of L.W., Y.C., W.L. and J. Wang on RT–qPCR, plasmid construction and in vitro transcription and W.S. and L. Yan on Cas9 protein purification. J. Wei prepared samples for deep sequencing, and W.X. performed the deep-sequencing data analyses and bioinformatics analysis, supervised by L. Yang. J.G., J.S., M.Z. and X.H. provided critical technical assistance. B.Y., J.C., L. Yang and B.S. wrote the paper with inputs from all authors. J.C. managed the project.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-017-0004-6.

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

Correspondence and requests for materials should be addressed to B.S. or L.Y. or J.C.

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

Methods

DNA constructs. Expression vectors of pST1374- N-NLS-flag-linker-Cas9 (Addgene 44758), pST1374-N-NLS-flag-linker-D10A (Addgene 51130) and pST1374 -N-NLS-flag-linker-H840A (Addgene 51129) were from X.H.'s lab[57]. To eliminate the activity of HNH nuclease domains[58], two addition mutations (D839A and N863A) were introduced into pST1374-N-NLS-flag-linker-H840A by using the In-Fusion PCR Cloning kit (Clontech, 638909). pST1374- N-NLS-flag-linker-dCas9 was constructed through combining pST1374-Cas9-N-NLS-flag-linker-D10A and pST1374-Cas9-N-NLS-flag-linker-H840A at BamHI and ApaI sites. For recombinant expression of Cas9 and its variants in *Escherichia coli*, genes were synthesized with codon optimized and cloned into pET28a vector. All sgRNA target sequences are listed in Supplementary Table 1.

The shuttle vector pSP189 containing the *SupF* report gene and the APOBEC3B-expressing plasmid pA3B were gifts from A. Furano (NIH). The kanamycin-resistant gene was amplified from pUC57-Kan by PCR (primers: pSP189-kan_F, 5′-TTCGGTCCTCCGATCGAATGGTTTC TTAGACGTCAGGTGGCACTTTTCGGGGA-3′; pSP189-kan_R, 5′ -ATTGTCAACAGAATTCTACGGGGTCTGACGCTCAGTGGAACGAA-3′) and subcloned into pSP189 at PvuI and EcoRI sites to produce kanamycin-resistant shuttle vector pSP189K. The catalytic deficient A3B (A3Bm)-expressing plasmid was produced by introducing mutations E68A, C97S, E255Q and C284S (refs [24,59]) with Agilent QuikChange II Kits in wild-type A3B. The BE3-expressing vector was purchased from Addgene (#73021).

Antibodies. Antibodies were purchased from the following sources: against alpha-tubulin (T6199), against HA-tag (HA-7, H3663) for western blot from Sigma; against 6×His-tag (M30111) for ChIP from Abmart; against HA-tag (ab9110) for ChIP, against Exo1 (ab95068), against MRE11 (ab214), against UNG (ab109214) for western blot from Abcam; against APE1 (#4128) for western blot, against Histone H3 for ChIP (#4620) from Cell Signaling Technology; against SMUG1 (A-1, sc-514343) for western blot from Santa Cruz Biotech.

siRNA. Control siRNA (no target in mammalian cells, A06001), siRNAs against *APOBEC3B* (5′-GUGAUUAAUUGGCUCCAUA-3′), *APOBEC3C* (5′-CAAUGUAUCCAGGCACAUU-3′), *APOBEC3F* (5′-GACGAUGAAGAAUUUGCAU-3′), *APOBEC3H* (5′-CUGUAAAUCCAGCCGGGUA-3′) and *MRE11* (5′ -CACUUCAAAGACAGAUCAA-3′) were synthesized by GenePharma (stable siRNA). siRNA against *APOBEC3D* (D-032279-24-0002) was purchased from Dharmacon RNAi Technologies (GE Healthcare Life Sciences).

RNA extraction, reverse transcription and quantitative PCR (RT–qPCR). Total RNAs were extracted with the TransZol Up Plus RNA Kit (TransGen) and reversely transcribed with cDNA Synthesis SuperMix (TransGen) according to the manufacturer's instructions. Total RNA (2 µl at 500 ng/µl), 1 µl random hexamer DNA primers (0.1 µg/µl) and 5 µl nuclease-free water were heated at 65 °C for 10 min and immediately put on ice, and then the reverse transcription mixture containing 10 µl 2×TS Reaction Mix, 1 µl gDNA remover, 1 µl TransScript Reverse Transcriptase and RNase Inhibitor Enzyme Mix was added. The reactions were incubated at 25 °C for 10 min, 42 °C for 30 min and finally 85 °C for 5 s to inactivate the enzymes. The reactions were diluted with 80 µl of nuclease-free water and qPCR was performed as described previously[5].

Cell line culture and transfection. 293FT and HeLa cells were from ATCC and have been tested for mycoplasma contamination by PCR methods. Both cell lines were maintained in DMEM (10566, Gibco, Thermo Fisher Scientific) + 10% FBS (16000-044, Gibco, Thermo Fisher Scientific). To establish stable cell lines of APOBEC3B (A3B) or mutant APOBEC3B (A3Bm), 293FT cells were seeded into a 60-mm plate at a density of 4×10^5 per well and cultured for 24 h. Cells were transfected with 1 µg APOBEC-expressing plasmid (wild-type A3B or its mutant A3Bm), 1 µl PLUS and 2 µl Lipofectamine LTX (Thermo Fisher Scientific) according to the manufacturer's instructions. After 48 h, 10 µg/ml puromycin (ant-pr-1, InvivoGen) was added to the media in the following two weeks, and stable cell lines were maintained with 1 µg/ml puromycin. The overexpression of A3B or mutant A3Bm in stable cell lines was validated by western blots.

A dual-sgRNA vector expressing two sgRNAs against *Exo1* and a dual-sgRNA vector expressing two sgRNAs against *APE1* was used for establishing *Exo1*-knockout and *APE1*-knockout cells, respectively. A tetra-sgRNA vector expressing two sgRNAs against *UNG* and the other two against *SMUG1* was used for establishing *UNG* and *SMUG1* double-knockout cells. Briefly, 293FT cells were seeded into a six-well plate at a density of 2.5×10^5 cells per well. After 24 h, cells were transfected with 4 µg Cas9-expressing plasmid, 2 µg sgRNA-expressing plasmid, 8 µl PLUS and 10 µl Lipofectamine LTX (Thermo Fisher Scientific) according to the manufacturer's instructions. After the treatment of 10 µg/ml blasticidin and 5 µg/ml puromycin in three days, transfected cells were trypsinized and transferred into a 100-mm dish for the selection of single-cell colonies. The genomic DNA of *UNG*, *SMUG1*, *Exo1* and *APE1* genes from the single cell colonies were sequenced, and knockdown effects were further validated by western blots.

To determine base substitution frequencies in ODNs, cells were transfected with 100 pmol ssODNs, dsODNs or ds-plasmid by electroporation (program Q-001) with Cell Line Nucleofector Kit V (VCA-1003) on Lonza Nucleofector-2b. Electroporated cells were then plated into a 60-mm dish. 72 h after electroporation, cells were lysed with QuickExtract DNA Extraction Solution and PCR amplified (primers: ODN_F, 5′-TGGTGTAGTGGTGTGAGAG-3′; ODN_R, 5′ -ACACCTACCCACCACACTT-3′). PCR products were gel purified with a DNA gel extraction kit (AP-GX-50, Axygen) and subjected to deep sequencing. dsODNs were produced by PCR amplification of ssDNAs according to published methods[37] and ds-plasmids were produced by inserting dsODNs into a cloning vector (pEASY-Blunt Cloning Kit, Transgen, CB101-02). ODN sequences are listed in Supplementary Table 2.

To determine base substitution frequencies in the genomic DNA during HDR, cells were cotransfected with 3 µg Cas9-expressing plasmid, 1.5 µg sgRNA-expressing plasmid and 100 pmol ssODNs, dsODNs or ds-plasmid donors by electroporation (program Q-001) with Cell Line Nucleofector Kit V (VCA-1003) on Lonza Nucleofector-2b. Electroporated cells were then plated into a 60-mm dish. 72 h after electroporation, cells were lysed with QuickExtract DNA Extraction Solution, and ODN-cognate regions were amplified by using PCR with PrimeSTAR HS DNA polymerase and 3′ end phosphorothioate-modified HDR-tag-specific primers listed in Supplementary Table 3. PCR products were gel purified with a DNA gel extraction kit (AP-GX-50, Axygen) and subjected to deep sequencing.

To determine the indels in genomic DNA, cells were seeded in a 24-well plate at a density of 1×10^5 cells per well and transfected with 1 µg Cas9-variant-expressing plasmid, 0.5 µg single sgRNA-expressing plasmid and 5.4 µl Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. After 24 h, blasticidin was added (ant-bl-1, InvivoGen) to the media at a final concentration of 10 µg/ml. After another 48 h, the genomic DNA was extracted with QuickExtract DNA Extraction Solution (QE09050, Epicentre) according to the manufacturer's instructions. Genomic DNA sequences at sgRNA target sites were individually amplified by PrimeSTAR HS DNA polymerase (Takara, R010A) with primers listed in Supplementary Table 3 and gel purified with DNA gel extraction kit (AP-GX-50, Axygen).

Screening *SupF* mutations on shuttle vector were performed as previously reported[5]. For APOBEC3 knockdown, we transfected 20 pmol siRNA with 6 µl Lipofectamine RNAiMAX (Thermo Fisher Scientific) into cells immediately after the cells were plated. All of the *SupF* mutations are listed in the Supplementary Note.

sgRNA sequencing. Total small RNAs were extracted with miRNeasy Mini kit (Qiagen, 217004) 48 h after sgRNA transfection and then reverse transcribed to cDNAs using NEBNext Small RNA Library Prep Set for Illumina (NEB, E7330S). Briefly, 1 µg of total small RNAs were ligated with the 3′ SR adaptor, hybridized with the reverse transcription primer and then ligated with the 5′ SR adaptor. After reverse transcription, the cDNAs were amplified 30 cycles with PrimeSTAR HS DNA polymerase and the primer set (5′-SR-Adaptor-F: 5′-GTTCAGAGTTCTACAGTCCGACGAT-3′; sgRNA-Seq-R: 5′ -ACCGACTCGGTGCCACTTTTTC-3′). The 122-bp PCR products were gel purified with Monarch DNA Gel Extraction Kit (NEB, T1020S) and then applied to deep sequencing.

DNA library preparation and sequencing. Indexed DNA libraries for next-generation deep sequencing were prepared by using the TruSeq ChIP Sample Preparation Kit (Illumina) according to the manufacturer's instructions. Briefly, the PCR products were individually fragmented with Covaris S220 and then amplified using the TruSeq ChIP Sample Preparation Kit (Illumina). After being quantitated with the Qubit High-Sensitivity DNA kit (Invitrogen), PCR products with different barcodes were pooled together and sequenced by using the Illumina Hiseq 2500 (1×100) or Hiseq X-10 (2×150) at CAS-MPG Partner Institute for Computational Biology Omics Core, Shanghai, China.

Chromatin immunoprecipitation qPCR. ChIP was performed with SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling Technology, #9003) following the provided instruction. Briefly, A3B cells were seeded into a 10-cm dish plate at a density of 1.6×10^6, and after 24 h, 1 ml serum-free DMEM containing 16 µg sgRNA-expressing plasmids, 32 µg Cas9-expressing plasmids, 64 µl Lipofectamine LTX and 48 µl PLUS was added. After another 48 h, the cells were washed once with PBS, and 270 µl of 37% formaldehyde was added to 10 ml of PBS for cross-linking (final formaldehyde concentration is 1%). After 30-min of cross-linking, the cells were harvested for the ChIP assay. Subsequent qPCRs were performed with primers listed in Supplementary Table 4 to detect the binding capacities of APOBEC3B to the sgRNA target regions.

Cas9 protein expression and purification. Cas9, Cas9 nickases and dCas9 were recombinantly expressed as C-terminal 6×His fusion proteins in *E. coli* Rosetta2 (DE3) at 16 °C overnight in LB medium. Harvested cells were lysed with cell disruptor (Microfluidics M-110P) in lysis buffer (20 mM Tris, 500 mM NaCl, pH 8.0, 2 mM PMSF and 5 mM β-mercaptoethanol) and affinity purified using His-Talon resin (Clontech). Eluted fractions from the His-Talon column were then

dialyzed against buffer SPA (20 mM HEPES, pH 7.5, 100 mM KCl, 1 mM DTT, 10% glycerol) overnight, loaded onto a 5-ml HiTrap SP-Sepharose column (GE Healthcare) and eluted with a gradient of 0–100% buffer SPB (20 mM HEPES pH 7.5, 1 M KCl, 1 mM DTT, 10% glycerol) in 12 column volumes. Eluted target fractions were further purified by size-exclusion chromatography on Superdex G200 column (GE Healthcare). The peak fractions containing Cas9 or its variants were then pooled, filter sterilized, concentrated and stored in buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 10% glycerol, 1 mM DTT) at −80 °C.

sgRNA transcription in vitro. pUC57-sgRNA expression vectors were linearized by DraI and in vitro transcribed using the MEGAshortscript Kit (Ambion, AM1354). sgRNAs were then purified using the MEGAclear Kit (Ambion, AM1908). The target sequences of sgRNAs are listed in Supplementary Table 1.

Oligonucleotide DNA cleavage and electrophoresis mobility shift assays (EMSA). The DNA oligonucleotides 5′ labeled with fluorescent dye (50 nM) were annealed with equal molar amounts of unlabeled complementary oligonucleotides at 95 °C for 3 min and then slowly cooled to room temperature to generate double-stranded DNA substrate containing the sgeGFP target site. The sequences of oligonucleotides used were listed in Supplementary Table 2.

For cleavage assays, Cas9, Cas9 nickase or dCas9 (50 nM final concentration) was preincubated with equimolar amounts of sgRNA in cleavage assay buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol) in a total volume of 9 µl. Reactions were started with the addition of 1 µl target DNA (50 nM) and incubated at 37 °C for 30 min. Reactions were quenched by adding 10 µl of 2× loading dye (5 mM EDTA, 0.025% SDS, 5% glycerol in formamide) and then heated at 95 °C for 5 min. Cleaved 5′-fluorescent-dye-labeled DNA oligonucleotides were resolved on 12% denaturing polyacrylamide gels containing 8 M urea and visualized by fluorescence imaging (Typhoon FLA 9500, GE Healthcare Life Sciences).

For EMSA, Cas9 nickase or dCas9 (30 or 300 nM final concentration) was preincubated with equal molar amounts of sgRNA in EMSA buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT and 10% glycerol) in a total volume of 9 µl as indicated in figures, and then 1 µl target DNA (30 nM) was added. The 10 µl EMSA reaction mix was incubated for 30 min at 37 °C and resolved at 4 °C on an 8% native polyacrylamide gel containing 1× TBE and 5 mM $MgCl_2$.

CD8⁺ T lymphocytes isolation and electroporation. Cryopreserved human peripheral blood mononuclear cells (PBMCs) were purchased from AllCells, LLC (PB003F). $CD8^+$ T cells were pre-enriched from the PBMCs with Easysep Human $CD8^+$ T cell enrichment kit (19053, Stemcell Technologies) according to the manufacturer's protocol. The cells were then activated with Dynabeads Human T-Activator CD3/CD28 (LifeTech, 11131D) for 3 d in AIM V medium (Gibco, 0870112DK) supplemented with 10% FBS. The stimulated $CD8^+$ T cells were then cultured in AIM V medium with 10% FBS and 100 IU/ml hIL2 (200-02-100, PeproTech).

To determine base substitutions during HDR, 4×10^6 cells were electroporated with 30 pmol of siRNA (program EO-115) by using P3 Primary Cell 4D-Nucleofector X Kit L (Lonza, V4XP-3024) on day 4. After 24 h, RNP mixture with or without 100 pmol of ssODN or dsODN donor and 2×10^5 transfected cells were combined in a Lonza 4D strip nucleocuvette and electroporated (program EO-115) by using P3 Primary Cell 4D-Nucleofector X Kit. After another 48 h, cells were lysed with QuickExtract DNA Extraction Solution for HDR taq-specific PCR amplification.

To determine indel formation, 4×10^6 cells were electroporated with 30 pmol of siRNA (program EO-115) by using P3 Primary Cell 4D-Nucleofector X Kit L (Lonza, V4XP-3024) on day 4 or day 7. After 24 h, 8 µg of Cas9 protein (or its variants) and 2 µg of sgRNA were mixed and incubated for 10 min to allow RNP formation. RNP mixture and 2×10^5 cells were combined in a Lonza 4D strip nucleocuvette and electroporated (program EO-115) by using P3 Primary Cell 4D-Nucleofector X Kit (V4XP-3032) on day 5 and day 8. After another 48 h, Cells were collected and lysed with QuickExtract DNA Extraction Solution for PCR amplification.

AP-site-containing ODN-breakage assay. The method was modified from the assay for APOBEC DNA deaminase activity[40]. Briefly, 10 nM ssODNs or uracil-containing ssODNs 5′ labeled with fluorescent dye were incubated with 0.1 unit of UNG (NEB, M0280) for 1 h at 37 °C, then treated with or without 10 mM and 100 mM NaOH for 20 min at 37 °C. The DNA fragments were resolved on 12% denaturing polyacrylamide gels containing 8 M urea. The sequences of oligonucleotides used are listed in Supplementary Table 2.

Identification of base substitution. On average, about 1,000,000 of 1 × 100 reads were obtained for each sample. These 1 × 100 reads were trimmed to remove the first five nucleotides at both sides and then mapped to specific human sequences with the BWA-MEM algorithm by multiple threads (BWA v0.7.9a, parameter: bwa mem reference_index.fa reads.fq -t 12 > bwa_mem_SE.sam). After alignment, unmapped reads were removed from the output sam files, which were further sorted to bam files using SAMtools (v0.1.18, parameter: samtools view -bS -F 4 bwa_mem_SE.sam |samtools sort - bwa_mem_SE). Mapped deep-sequencing reads were piped up at their aligned genomic sites with Perl scripts based on SAMtools (parameter: samtools mpileup -d 1000 -OI -f reference_index.fa bwa_mem_SE.bam). Base substitutions were selected with the following strict criteria: (i) sites were mapped with at least 1,000 independent reads (Read Depth ≥ 1,000); (ii) read qualities on the examined base substitution sites should be greater than 30 when evaluated with Phred quality (Q score, Base Quality ≥ 30); (iii) the base substitution rate on each selected site should be greater than 0.1% (Variant Ratio ≥ 0.1%). To determine the base substitutions in the ODN-cognate genomic region for HDR outcome, false positive sites that were detected in genomic regions from untreated cell samples were removed. Because the base substitution frequencies in the ODN-cognate region examined from the dsODN-HDR and ds-plasmid-HDR were all less than 0.17%, the base substitution frequencies lower than this value were filtered as background for the APOBEC3-caused base substitutions in ssODNs during HDR (Fig. 2b). All the base substitutions determined in this study are listed in Supplementary Dataset 2.

Indel frequency calculation. Indels were estimated with sorted bam files (bwa_mem_SE.bam) and only calculated within the aligned regions spanning 25 bp upstream and downstream of Cas9-cleavage sites (50 bp total) with SAMtools. Reads with Q scores greater than 30 were selected, and indel frequencies were subsequently calculated by dividing reads containing inserted or deleted nucleotides (>2 bp) by all of the mapped reads in the same region by the following equation:

$$\text{Indel Frequency}(\%) = \frac{\sum_{i=1}^{n} (I_i + D_i)}{\sum_{i=1}^{n} R_i} \times 100\%$$

where I_i, D_i and R_i represent total insertion reads, total deletions reads and total mapped reads, respectively, in the aligned regions spanning 25 bp upstream and downstream of Cas9-cleavage sites. All of the indel frequencies determined in this study are listed in Supplementary Dataset 3.

The indel distributions were analyzed on both flanking sides of the Cas9-cleavage sites by plotting the indels counts at each position against the total indel counts within 25 bp upstream and downstream of the cleavage site. For each indel-containing read, insertion was counted with the total bases inserted at each inserted position, and deletions were counted once at each deleted position.

$$\text{Indel Fraction}(\%) = \frac{(\text{InsertionCounts} + \text{DeletionCounts})_x}{\sum_{-25}^{25} (\text{InsertionCounts} + \text{DeletionCounts})_i} \times 100\% (x \subseteq [-25, 25])$$

Statistical analysis. *P* values were calculated using a one-tailed Student's *t* test in this study. The number of technical replicates or independent cell culture experiments is indicated in the relevant figure legends.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. Deep-sequencing data can be accessed at the NCBI Gene Expression Omnibus under accession code GSE105146. Source data for Figs. 1c,d, 2b,c, 3b–g and 4b,c are available online. Other data are available from the corresponding author upon reasonable request.

References

57. Shen, B. et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods.* **11**, 399–402 (2014).
58. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
59. Bogerd, H. P., Wiegand, H. L., Doehle, B. P. & Cullen, B. R. The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. *Virology* **364**, 486–493 (2007).

nature research

Corresponding author(s): Bin Shen, Li Yang, Jia Chen

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined. not applicable

2. Data exclusions

Describe any data exclusions. not applicable

3. Replication

Describe whether the experimental findings were reliably reproduced. The experimental findings in all figures were reproduced successfully.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups. not applicable

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis. not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	KaleidaGraph and Microsoft Excel

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Materials in this study are available for distribution following MTA.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	We stated antibody information in the method section.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	All cell lines were from ATCC.
b. Describe the method of cell line authentication used.	not applicable
c. Report whether the cell lines were tested for mycoplasma contamination.	All cell lines have been tested for mycoplasma contamination free by PCR methods.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	not applicable

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	not applicable

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	not applicable