RESEARCH

Adenine base editors induce off-target structure variations in mouse embryos and primary human T cells

Leilei Wu[1,2†], Shutan Jiang[2†], Meisong Shi[2†], Tanglong Yuan[3], Yaqin Li[2], Pinzheng Huang[2], Yingqi Li[2], Erwei Zuo[3*], Changyang Zhou[1*] and Yidi Sun[1,4*]

†Leilei Wu, Shutan Jiang and Meisong Shi contributed equally to the work.

*Correspondence:
zuoerwei@caas.cn;
zhouchangyang@ion.ac.cn;
ydsun@ion.ac.cn

[1] Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[2] Epigenic Therapeutics, Inc, Shanghai 201315, China
[3] Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangdong, China
[4] Key Laboratory of Genetic Evolution & Animal Models, Chinese Academy of Sciences, Shanghai, China

Abstract

Background: The safety of CRISPR-based gene editing methods is of the utmost priority in clinical applications. Previous studies have reported that Cas9 cleavage induced frequent aneuploidy in primary human T cells, but whether cleavage-mediated editing of base editors would generate off-target structure variations remains unknown. Here, we investigate the potential off-target structural variations associated with CRISPR/Cas9, ABE, and CBE editing in mouse embryos and primary human T cells by whole-genome sequencing and single-cell RNA-seq analyses.

Results: The results show that both Cas9 and ABE generate off-target structural variations (SVs) in mouse embryos, while CBE induces rare SVs. In addition, off-target large deletions are detected in 32.74% of primary human T cells transfected with Cas9 and 9.17% of cells transfected with ABE. Moreover, Cas9-induced aneuploid cells activate the P53 and apoptosis pathways, whereas ABE-associated aneuploid cells significantly upregulate cell cycle-related genes and are arrested in the G0 phase. A percentage of 16.59% and 4.29% aneuploid cells are still observable at 3 weeks post transfection of Cas9 or ABE. These off-target phenomena in ABE are universal as observed in other cell types such as B cells and Huh7. Furthermore, the off-target SVs are significantly reduced in cells treated with high-fidelity ABE (ABE-V106W).

Conclusions: This study shows both CRISPR/Cas9 and ABE induce off-target SVs in mouse embryos and primary human T cells, raising an urgent need for the development of high-fidelity gene editing tools.

Keywords: ABE, Cas9, Off-target structure variation, Large deletion, T cell

Background

The paramount consideration in genome editing is the assurance of clinical safety and the prevention of undesirable genotoxicities. Multiple methods have been applied to evaluate the off-target effects of various editing tools [1–4], and high-fidelity versions of CRISPR-based editing systems have been subsequently developed [4–7]. In addition

Wu *et al. Genome Biology*   (2024) 25:291

Page 2 of 22

to single nucleotide variations, structural variations (e.g., large deletion, translocation) induced by gene editing tools have been of significant concern for clinical application [4, 8, 9]. Previous studies have reported that CRISPR/Cas9 induced large deletions at both on-target and off-target loci in mouse [10] and human embryos [11], embryonic stem cells [12], and induced pluripotent stem cells [13]. Chromosome losses have been directly detected using single-cell RNA sequencing (scRNA-seq) analysis in CRISPR/Cas9-edited primary human T cells [4, 7]. In addition, Tsuchida et al. demonstrated that the CRISPR/Cas9 editing-induced chromosomal alteration was closely linked to p53 activation and cell death, raising safety concerns of CRISPR/Cas9 cleavage in T cell therapy [7].

Base editors (BEs) have been widely used in gene therapies to effectively convert base pairs (A/T to G/C; C/G to T/A) without generating double-strand breaks (DSBs). Studies reported before have shown that the deaminase domains in cytosine base editor (CBE) [14] and adenine base editors (ABE) [15] both lead to off-target single nucleotide variations (SNVs) at the DNA [16] and RNA level [5]. Subsequent studies have adopted various methods to minimize these off-target effects [17–21]. Apart from SNVs, cytidine deaminases have been shown to induce genomic DNA breaks and chromosomal instability by triggering DSBs [22, 23]. Consistently, CBE has been reported to be associated with DSB-related DNA damage risks, primarily due to the deaminase activity [24–26]. Recently, CGBE is reported to generate double-strand breaks and causes structure variations [21]. However, the evaluation of such side effects after editing with ABE has been insufficient. In this study, we have conducted a systematic evaluation of potential off-target structural variations for widely used gene editing tools (CRISPR/Cas9, ABE, and CBE) using whole genome sequencing analysis in mouse embryos. Assisted by single-cell RNA-seq analysis, we further evaluated the genome-wide aneuploidy and truncations of chromosomes in primary human T cells transfected with CRISPR/Cas9 or base editors. We detected frequent structural variations associated with CRISPR/Cas9 or ABE editing in both mouse embryos and human cells.

Results

Genome-wide detection of structure variations induced by CRISPR/Cas9 and base editors

To explore whether various gene editing approaches would induce off-target structure variations (SVs), we applied Gridss [27] and Manta [28] to call potential SVs in paired whole genome sequencing (WGS) samples using the genome-wide off-target analysis by two-cell embryo injection (GOTI) method, where the edited and non-edited samples derive from the same zygote and thus avoid the influence of genetic background [16]. We called structure variations in tdTomato + (T +) samples with tdTomato − (T −) samples from the same embryo as the reference. We defined high-confident (HC) off-target SVs that were called by both algorithms, and the union set of structure variations was shown in Additional file 3: Table S2. In comparison to the control group, HC off-target SVs were detected in Cas9-treated, ABE-treated (ABE7.10: TadA-TadA*-nCas9) and CBE-treated (BE3: rAPOBEC1-nCas9-UGI) embryos in 50% samples while not in controls (Fig. 1a and Additional file 1: Fig. S1a). Interestingly, Cas9-treated and ABE-treated embryos showed higher number of off-target structure variations, while CBE-treated embryos had similar number of SVs

Wu *et al. Genome Biology*   (2024) 25:291

Page 3 of 22

compared with that of the control group (Fig. 1a and Additional file 1: Fig. S1b), likely spontaneous structure variations during embryo development.

Among the detected SVs, HC large deletion (DEL) accounted for 56%, 75%, and 100% of all off-target SVs in Cas9-treated, ABE-treated, and CBE-treated embryos, respectively (Fig. 1b). For Cas9, the off-target deletion rate (85.71%) showed much higher frequency than that of the on-target large deletion (14.29%) (Fig. 1c and Additional file 1: Fig. S1c), indicating that off-target SVs were more common events than on-target large deletions. ABE-induced large deletions had prominently larger sizes than those in the Cas9 group (median 2414 bp vs. 1334 bp), and with more extreme outliers than Cas9 and CBE (Fig. 1d). The maximum deletion detected in ABE reached to 4.1Mb, which was only observed in T + other than T- sample (Fig. 1e). The large deletions in Cas9-treated, ABE-treated and CBE-treated mice were visualized at the corresponding genomic loci and validated by PCR analysis both for HC and non-HC large deletions (Fig. 1f, and Additional file 1: Fig. S1d, S1e, S1f; Additional file 3: Table S2). In addition to large deletions, we also observed chromosomal translocation in ABE-treated and Cas9-treated groups (Fig. 1b). The off-target translocations in Cas9-treated were visualized at the breakpoints and further confirmed by PCR analysis (Fig. 1g). For these off-target large deletions in ABE-treated mice, we examined the breakpoints (flanking 500 bp) of SVs in different groups and found that the percentage of A was significantly higher in ABE-treated mice than that observed in the Cas9 group ($P = 0.0046$) or random background ($P = 0.003$; Additional file 1: Fig. S1g). In addition, motif analysis of the SVs (flanking 10bp) revealed that TAT was the most recurrent motif in ABE-treated mice (Fig. 1h), consistent with the binding preference of TadA previously reported [5]. We next used Cas-OFFinder [1] to predict potential off-target sites showing sequence similarity with the on-target sites, and the predicted off-target sites with mismatches less than 8 were rare for all designed sgRNAs (Fig. 1i) in the whole genome, demonstrating high specificity of designed sgRNAs. In addition, local sequence alignment revealed that the off-target SV sequences showed poor similarity with the on-target sequences (Fig. 1j, and Additional file 1: Fig. S1h, S1i), indicating that these off-target SVs were more likely to be sgRNA-independent.

(See figure on next page.)

Fig. 1 Off-target structural variations in the mouse embryo genome. **a** The number of high-confidential structural variations detected in non-editing control, ABE, Cas9, and CBE-edited mouse embryos. **b** Proportion of structural variation types in non-editing control, ABE, Cas9, and CBE-edited mouse embryos. **c** Percentage of samples with on-target large deletions and off-target large deletions. **d** Size distribution of large deletions in Cas9, ABE, and CBE. **e** Sequencing depth distribution around two large deletions of T + (tdTomato +) and T − (tdTomato −) in ABE. **f** PCR validation of large deletions detected in Cas9- and ABE-treated embryos. **g** Supporting sequencing reads of Chr7-Chr14 translocation shown in IGV and PCR-validated translocations detected in Cas9. **h** Enriched motifs of sequences within 10 bp flanking the breakpoints of SVs in ABE-treated mice. **i** Proportion of off-target sites with different mismatch numbers predicted by Cas-Offinder for designed sgRNAs. **j** Similarity of the on-target sequence and sequences surrounding breakpoints of SVs. The ratio of matching bases between the on-target sequence and all off-target SVs at each position was displayed at the top heatmap. The top 5 off-target sequences showing the highest similarity with the on-target sequence were shown. Dots represent matched bases. **k** Annotation of high-confidential structural variations in non-editing control, ABE, Cas9, and CBE-edited mouse embryos. **l** Inferred copy number distribution in ABE-edited HEK293T cells
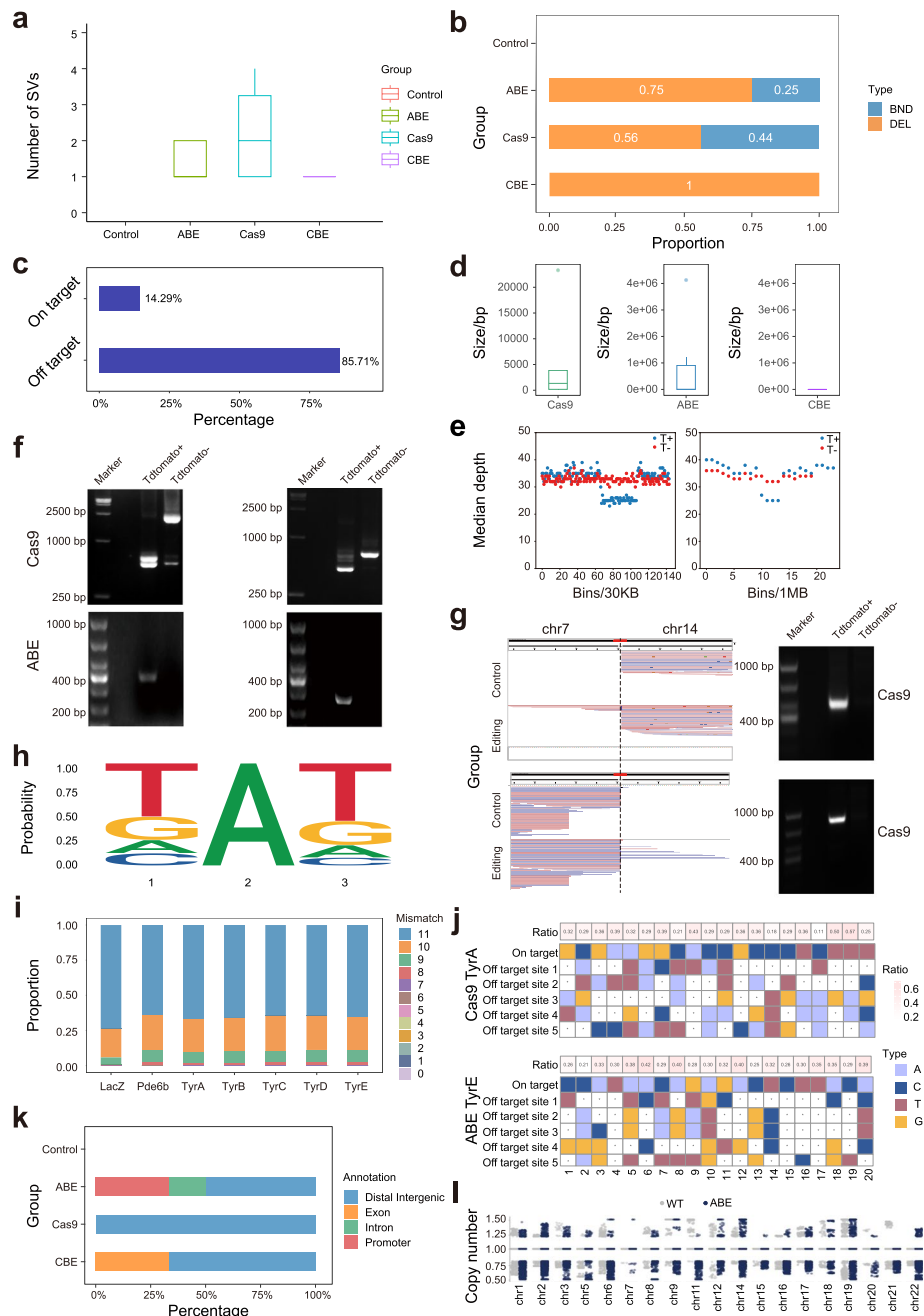
Fig. 1 (See legend on previous page.)

To explore whether off-target SVs could affect gene expression, we annotated and mapped the genomic positions of these SVs onto annotated genes using ChIPseeker [29]. Distal intergenic regions were the most common annotated regions and ABE-detected off-targets located in promoter region for HC off-target SVs (Fig. 1k), and the same trend was found for all off-target SVs (Additional file 1: Fig. S1j). Then we used bedtools [30] to intersect off-target SVs with gene locations and normalized the deletion size for ABE and Cas9. The result showed that Cas9 affected the expression of a higher number of genes than that of ABE (Additional file 1: Fig. S1k). The existence of ABE-induced large

Wu *et al. Genome Biology*   (2024) 25:291

Page 5 of 22

deletions was further validated in ABE-treated human cells using previously published single-cell RNA-seq data [5] (Fig. 1l).

Single-cell RNA-seq analysis reveals chromosomal aneuploidy after CRISPR/Cas9 or ABE cleavage in primary human T cells

To further explore whether off-target SVs induced by SpCas9 or ABE8.8 would occur in a clinically relevant setting, we performed CRISPR/Cas9 or ABE electroporation together with sgRNAs targeting *PDCD1* or *B2M* into human primary T cells with high efficiency (Fig. 2a). The sgRNAs used for Cas9 were the same as those used clinically in T cells reported before [4, 31]**.** We collected single cells at days 0, 3, 7, and 21 after transfection for single-cell RNA sequencing. Both Cas9 and ABE achieved high editing efficiency on targeted genes (Fig. 2b), which was further validated by the downregulation in the gene expression level of *PDCD1* and *B2M* (Additional file 1: Fig. S2a). For single-cell analysis, we obtained 217,033 single cells (Additional file 4: Table S3). After quality control, 209,469 single cells were kept which were classified into 8 cell types by unsupervised clustering (Fig. 2c). Based on the expression of canonical marker genes, we manually annotated the cell clusters as naive CD4 T cell (*CD4*, *CD40LG*, and *IL7R*), naive CD8 T cell (*CD8B*, *CD8A*, and *IL7R*), cytotoxic CD8 T cell (*CD8B*, *CD8A*, *NKG7*, *GZMB*, and *GZMA*), CD4 effector memory T cell (*CD4*, *ILR7*, *SELL*, *MKI67*, and *HMGN2*), CD8 effector memory T cell (*CD8A*, *CD8B*, *SELL*, *NKG7*, *MKI67*, and *HMGN2*), proliferating T cell (*MKI67* and *HMGN2*), gamma delta T cell (*TRDC*) and B cell (*MS4A1*, *CD79A*, and *BANK1*; Fig. 2c and d; Additional file 5: Table S4).

We next explored chromosomal abnormalities in each annotated cell type using a newly developed Bayesian model to remove false positives based on inferred copy number changes from InferCNV (see Methods). Significant changed genes from the model were shown in Additional file 6: Table S5. Finally, we detected 2 and 3 inferred CNV regions in ABE-treated T cells, and 13 and 47 in Cas9-treated T cells with significant aneuploid ratio change targeting *B2M* and *PDCD1*, respectively. To further reveal the characteristics of cells showing chromosomal abnormalities, we classified cells with chromosome deletions as aneuploid cells. To further validate the reliability of detected aneuploid cell proportion in the editing group, we did a reverse calling process to obtain aneuploid cell ratios (proportion/reverse proportion; see the "Methods" section). Compared to control group (blank vs. none-targeting control), the overall aneuploid cell ratios significantly increased by 1.70 and 4.28 folds for ABE ($P = 0.012$) and CRISPR/Cas9 ($P = 9.1e - 09$), respectively (Fig. 2e). Strikingly, we detected 32.74% of aneuploid cells transfected with Cas9 and 9.17% of cells transfected with ABE (Fig. 2f).

Specifically, we observed significant chromosomal loss in chr15 in Cas9-transfected naive CD4 T cells, and chr13 and chr2 in ABE-transfected CD8 effector memory T cells and proliferating T cells (Additional file 1: Fig. S2b). For CRISPR/Cas9 or ABE-transfected cells targeting *PDCD1* or *B2M*, we found frequent deletions across the genome in addition to the on-target deletion at chr2 or chr15 (Additional file 1: Fig. S2b). Chromosomal deletions with significant aneuploid ratio change were observed in all T cell types in Cas9-transfected group targeting *PDCD1* (Fig. 2g), and in two subtypes of ABE-transfected cells (Fig. 2h). Similarly, deletions occurred more often in Cas9-transfected group targeting *B2M* than ABE-treated cells (Additional file 1: Fig. S2c).
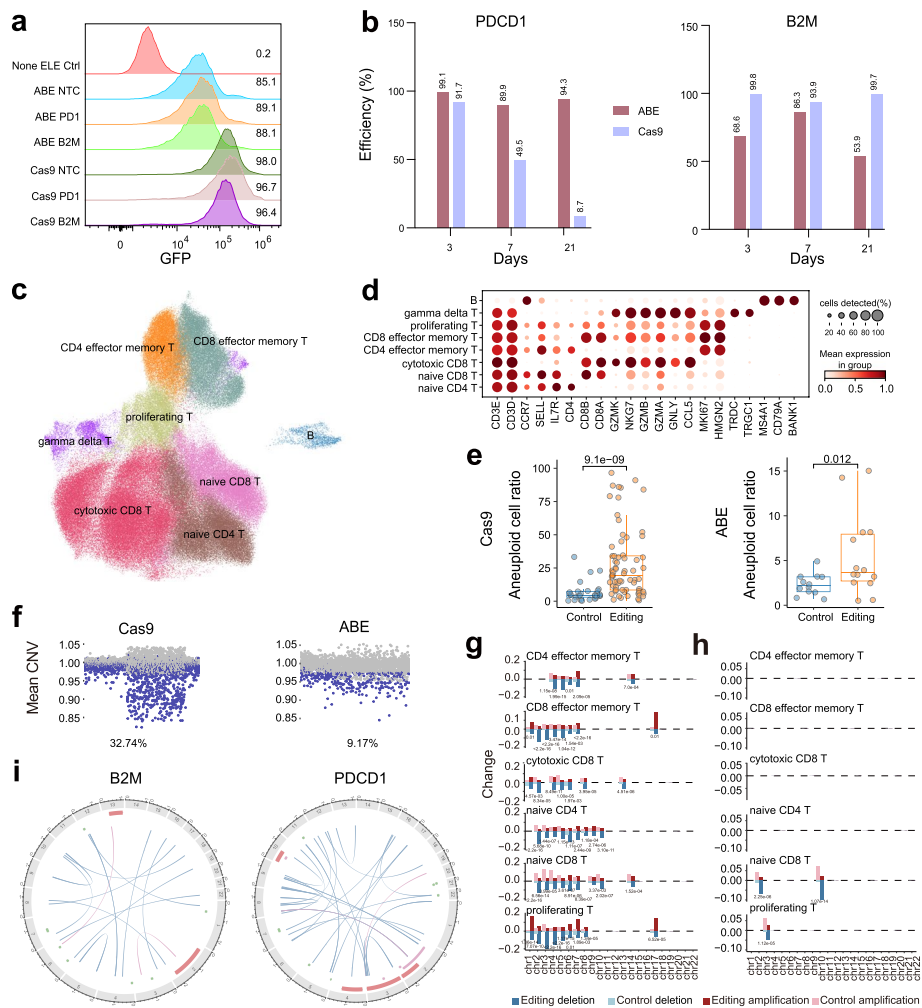
Wu *et al. Genome Biology*   (2024) 25:291

Page 6 of 22

Fig. 2 Off-target structural variations detected in primary human T cells using single-cell RNA-seq. **a** Quantification of the efficiency of mRNA electroporation for editing tool through assessment of GFP fluorescence levels. **b** Editing efficiency of *PDCD1* and *B2M* for CRISPR/Cas9 and ABE at days 3, 7, and 21. **c** Uniform manifold approximation and projection (UMAP) plot depicting 209,469 cells representing 8 cell lineages. **d** Dot plot showing averaged expression levels of marker genes across 8 cell types. **e** Distribution of aneuploid cell ratio across the whole genome for Cas9 and ABE. Dot represents the aneuploid cell ratio (proportion/reverse proportion) for a particular cell type detected at chromosomes. **f** Proportions of cells with deletions among all cells corresponding to Cas9 and ABE groups. The blue dot represents cells with inferred copy number values lower than 2 standard deviations and the gray dot represents normal cells. **g** Proportions of cells harboring deletions marked blue in the Cas9 editing group for *PDCD1* compared to the control group marked light blue (non-targeting control vs. blank). **h** Proportions of cells harboring deletions marked blue in the ABE editing group for *PDCD1* compared to the control group marked light blue (non-targeting control vs. blank). **i** Circos plot showing all HC SVs in ABE-treated cells targeting *B2M* or *PDCD1*. Pink bars represent regions detected by scRNA-seq data. Pink lines represent SVs located in regions detected by scRNA-seq data

To validate whether these off-target SVs from scRNA-seq could be detected at the genome level, we also performed WGS for ABE-treated T cells. The same SV calling pipeline was applied as that for GOTI samples. Compared to scRNA-seq that only reveals SVs located in coding regions, WGS analyses revealed more SVs at the whole-genome level in ABE-treated T cells (Additional file 1: Fig. S2d; Additional file 6: Table S5). Motif analysis revealed a TAT motif for detected SVs in T cells targeting

Wu *et al. Genome Biology*   (2024) 25:291

Page 7 of 22

either *B2M* or *PDCD1* (Additional file 1: Fig. S2e), in line with that found in GOTI samples. In *B2M*-targeting cells, we found two translocations located in the off-target SV regions detected from scRNA-seq (Fig. 2i, marked red). In *PDCD1*-targeting cells, WGS revealed SVs showed high overlap with those detected in scRNA-seq data (Fig. 2h, i and Additional file 1: Fig. S2c). The in silico visualization by sequencing depths and IGV further supported the high confidence of these overlapped SVs identified from both WGS and scRNA-seq data (Additional file 1: Fig. S2f and S2g).

The majority of aneuploid cells only had deletions in single chromosome for both Cas9 and ABE, while an increased proportion of cells with multiple chromosomal deletions was found in Cas9-treated cells targeting *PDCD1* (Fig. 3a). Differential expression analysis showed that compared to normal ploid cells, aneuploid cells significantly upregulated 31 genes (Fig. 3b). Among them, *CD70* and *MDM2* occurred frequently in multiple cell types in Cas9-transfected cells targeting *B2M* except naive CD8 T cells (Additional file 1: Fig. S3a). *MDM2* has been identified as a p53 interacting protein that represses p53 transcriptional activity and was associated with many human malignancies [32]. Overexpression of *CD70* in cancer cells has been shown to promote cell proliferation and survival [33, 34]. In addition, a previous study has reported that *CD70* and *MDM2* as the most frequently upregulated genes in cells regardless of which chromosome was lost [7]. For Cas9-transfected cells targeting *PDCD1*, we only detected three upregulated genes (*DNAJB11*, *TNFSF10*, and *JUN*) in aneuploid cells with multi chromosomes lost (Additional file 1: Fig. S3a). In ABE-treated naive CD8 T cells targeting *PDCD1*, we also detected two genes with
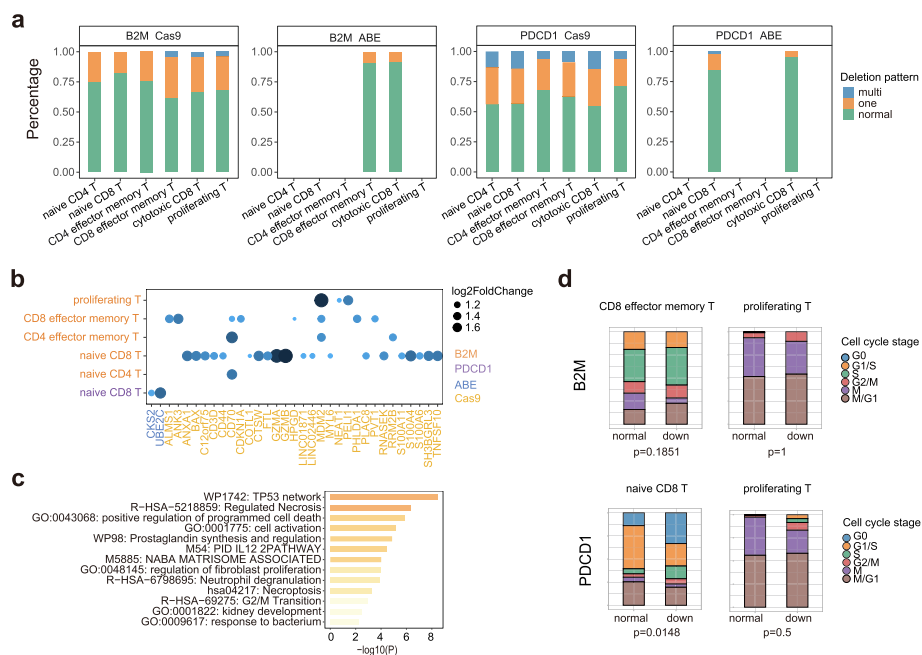
Fig. 3 Deletion patterns, significant changed genes, and biological functional changes in aneuploid cells. **a** Proportions of cells with multi-deletion, one deletion, and no deletions among the six cell types. **b** Log2 fold change of differentially expressed genes from aneuploid cells and normal cells in six cell types for Cas9 and ABE. **c** Enrichment results of differentially expressed genes from aneuploid cells and normal cells for Cas9 edited groups. **d** Cell cycle distribution of normal cells and cells with deletions in ABE cells

Wu *et al. Genome Biology*   (2024) 25:291

Page 8 of 22

significant upregulation, *UBE2C* and *CKS2* (Fig. 3b), which were both reported to be functional in cell cycle and cancer progression [35, 36].

The upregulated genes in aneuploid cells of the Cas9-transfected group targeting *B2M* were significantly enriched in the *TP53* network, programmed cell death, and G2/M transition (Fig. 3c), consistent with previous studies showing that a single Cas9-induced DSB can lead to p53 upregulation [37]. The activity of the p53 signaling pathway and apoptosis signaling were also significantly increased in aneuploid cells of 3 out of 6 cell types in the Cas9-transfected group targeting *B2M* but not in the Cas9-transfected group targeting *PDCD1* (Fig. 3c and Additional file 1: Fig. S3b). By contrast, the activity of cell cycle-related genes was significantly expressed in ABE-transfected cells (Fig. 3b). Cell cycle is an important signature and it is closely related to the p53 pathway [38] and *TP53* gene expression showed a positive correlation with *UBE2C* in several cancers [35]. Based on significantly differentially expressed genes and pathways enriched in aneuploid cells, we further calculated cell proportion in different cycle stages for each cell type, and found that the cell proportion in early cell cycle phases significantly increased in the aneuploid cells for the majority of cell types compared to normal cells in ABE-transfected groups (Fig. 3d). In conclusion, we detected aneuploidy and chromosomal deletions associated with functionally important genes when targeting *PDCD1* and *B2M* by either Cas9 or ABE.

Using longitudinal single-cell data, aneuploid cells were monitored to explore their fitness. Despite the gradual decrease in chromosome loss cells, Cas9 or ABE-induced chromosome loss in T cells is still observable at 3 weeks after transfection (Fig. 4a). Overall, the percentage of aneuploid cells of Cas9 targeting *B2M* and *PDCD1* decreased from 11.27%, 8.17% at day 3 to 6.67%, and 3.17% at day 21, respectively. For ABE, the percentage of aneuploid cells targeting *B2M* and *PDCD1* decreased from 8.51%, 6.79% at day 3 to 4.83%, and 2.33% at day 21 (Fig. 4b). For example, aneuploid cells with chromosome 15 loss decreased from 22.50% at day 3 to 8.99% at day 21 in Cas9-transfected naive CD4 T cells, and the aneuploid frequency of chr10 loss in ABE-transfected naive CD8 T cells decreased from 10.2% at day 3 to 3.67% at day 21 (Fig. 4c). In addition, all other aneuploid cells changed over time were shown in Additional file 1: Fig. S4a. These long-lived aneuploid cells highly expressed genes enriched in cell cycle, DNA repair, and DNA damage pathways (Additional file 7: Table S6), suggesting the safety concern of Cas9- or ABE-editing in clinical therapy.

Chromosomal aneuploidy in different cell types

To determine whether these SVs in ABE were specific to T cells, we also collected B cells from PBMC in addition to T cells. A total of 378 cells and 129 cells were identified in control and ABE-treated cells targeting *PDCD1*, respectively (Fig. 2c). In the Cas9-treated group, 226 and 25 cells were found in control and edited groups targeting *B2M*, and 226 and 276 cells were in control and edited group targeting *PDCD1*, respectively (Fig. 2c). Targeted genes *B2M* and *PDCD1* were both significantly repressed in ABE- or Cas9-treated B cells (Additional file 1: Fig. S5a). Notably, a significantly higher aneuploidy cell ratio was detected in ABE- or Cas9-treated cells than that in the control group (Additional file 1: Fig. S5b, marked red). We detected 3 and 3 off-target SVs in ABE and Cas9 (Fig. 5a and Additional file 1: Fig.
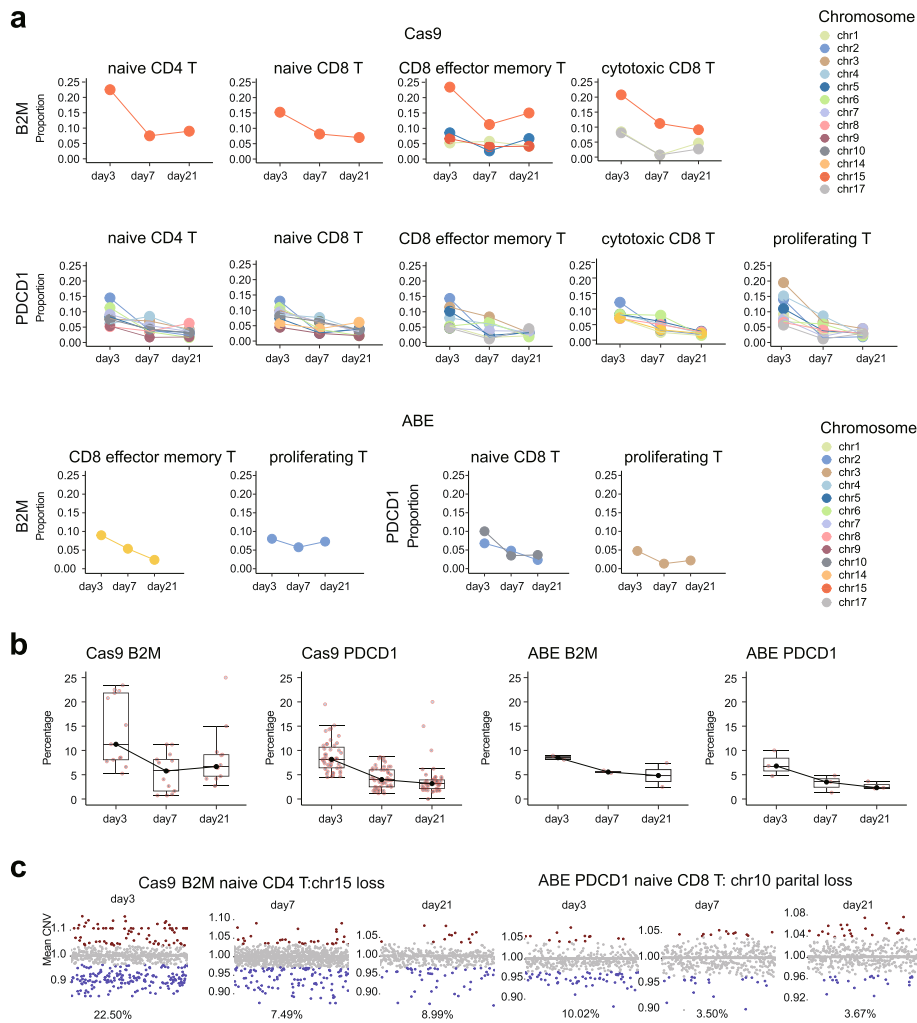
Wu *et al. Genome Biology*   (2024) 25:291

Page 9 of 22

Fig. 4 Longitudinal single-cell data monitoring the fitness of aneuploid cells. **a** Changes in the proportion of aneuploid cells in different cell types from day 3 to day 21 at every detected chromosome for Cas9 and ABE edited *B2M* and *PDCD1*. **b** Overall proportion changes of aneuploid cells from day 3 to day 21, categorized by editing tools and targets. **c** Dot plot representing specific examples of aneuploid cell changes from day 3 to day 21 in naive CD4 T cells on chr15 of Cas9 editing *B2M* and in naive CD8 T cells on chr10 of ABE editing *PDCD1*. Each dot represents the mean inferred copy number value of cells

S5c) targeting *PDCD1*. To further verify the universality of off-target phenomena, we also performed ABE and Cas9 editing targeting *B2M* in Huh7 cells. High editing efficiency was achieved in both Cas9- and ABE-treated cells (Fig. 5b). We detected 5 regions with chromosomal abnormality in ABE and 2 in the Cas9 group (Fig. 5c and Additional file 1: Fig. S5d). Similar to that found in ABE- or Cas9-treated T cells targeting *B2M*, off-target SVs were also detected in chr15 and chr2 in Cas9- and ABE-treated Huh7 cells, respectively (Fig. 5c). Taking together, our results indicated that the off-target SVs are not specific to T cells but are generally detectable in other cell types including B cells and Huh7 cells.
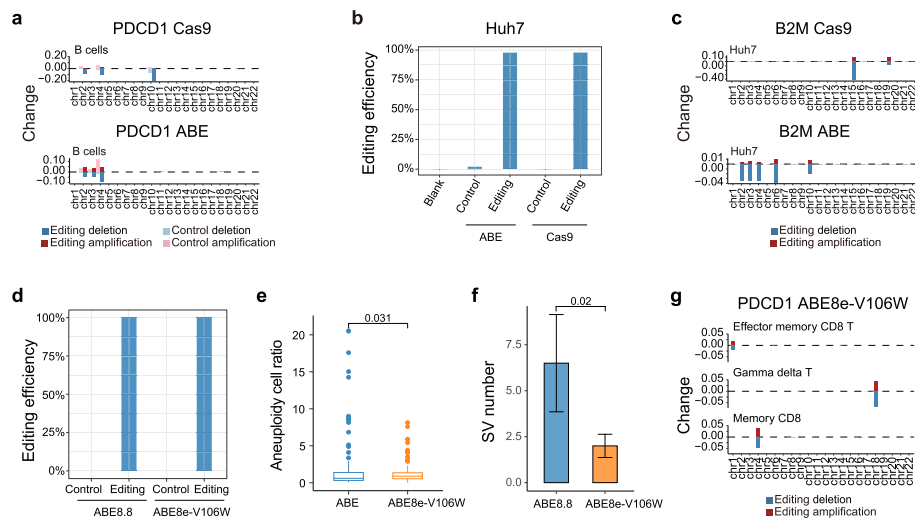
Fig. 5 Off-target SVs detected in other cell types and high-fidelity ABE. **a** Proportions of cells harboring deletions marked blue in Cas9 and ABE editing group for *PDCD1* compared to control group marked light blue (Non-targeting control vs. Blank) in B cells. **b** Editing efficiency of ABE and Cas9 in Huh7 cells. **c** Proportions of cells harboring deletions marked blue in Cas9 and ABE editing group for *B2M* compared to the control group marked light blue (Non-targeting control vs. Blank) in Huh7. **d** Editing efficiency of ABE8.8 and ABE8e-V106W targeting *PDCD1*. **e** Aneuploidy cell ratio across the whole transcriptome in ABE and ABE8e-V106W. **f** Number of off-target SVs among different T cell types with P value 0.02 between ABE and ABE8e-V106W. **g** Proportions of cells harboring deletions marked blue in ABE8e-V106W editing group for *PDCD1* in T cells

Off-target SVs in primary T cells treated with high-fidelity ABE

To explore whether the off-target SVs could be reduced, we next performed single-cell data analysis in primary T cells using high-fidelity ABE (ABE8e-V106W) [18] to target *PDCD1*. Similar editing efficiency was achieved by ABE8e-V106W in comparison to that achieved by ABE (Fig. 5d). These cells were clustered and annotated into 8 cell types (Additional file 1: Fig. S5e and S5f). Off-target analysis revealed that off-target aneuploid cell ratios in ABE8e-V106W-treated cells were significantly reduced than those in ABE-treated cells (Fig. 5e). In addition, we found off-target SV numbers were also significantly reduced (Fig. 5f, g and Additional file 1: Fig. S5g). These results supported that the off-target SVs were induced by deaminase in a sgRNA-independent manner and could be reduced by a high-fidelity version of ABE.

Discussion

In the current study, we comprehensively assessed the off-target structure variations induced by CRISPR/Cas9, ABE, and CBE. In mouse embryos, we detected a significantly increased number of off-target structure variations in ABE and Cas9, whereas no detectable off-target SVs were observed for CBE. Further single-cell RNA-seq analyses of human primary T cells showed that 32.74% of cells in Cas9 and 9.17% in ABE were aneuploid. Even after 3 weeks of culture, 16.59% and 4.29% of aneuploid cells were still detected. Furthermore, we found these off-target structural variants were also prevalent in B cells and Huh7 cell types. In Cas9-induced aneuploid cells, the p53 pathway and apoptosis pathway were significantly upregulated. On the other hand, aneuploid cells in

Wu *et al. Genome Biology*   (2024) 25:291

Page 11 of 22

ABE highly expressed two cell cycle-related genes and tended to be quiescent as the proportion of cells in the early stage of the cell cycle increased. These results highlight the importance of minimizing off-target structure variations in CRISPR/Cas9 and ABE.

CRISPR/Cas9 generates double-strand breaks (DSBs) in DNA, which can be repaired through homology-directed repair (HDR) and nonhomologous end joining (NHEJ) mechanisms [39–41]. Repair of a single DSB can result in perfect re-joinings, small indels, microhomology-mediated deletions, and large deletions [42]. On-target large deletions caused by Cas9 have been reported in several studies [8, 9, 43, 44]. Besides, off-target SVs caused by Cas9 were also observed in fertilized zebrafish eggs [8], in line with our findings in mouse embryos. In contrast to CRISPR/Cas9, ABE is an attractive gene-editing modality as it can introduce precisely targeted alterations without DSBs [15]. Off-target effects of ABE mainly focus on the deaminase activity on RNA [5, 20] and sgRNA-dependent off-targets on DNA [45]. In genomic DNA, in addition to nickase activity, deaminase activity could lead to genomic instability [23, 25]. Thus, the off-target SVs might be generated through two mechanisms in BEs: (1) sgRNA-dependent off-target effects induced by the non-specific binding of sgRNA subsequent to DNA cleavage of nCas9 [46] and (2) sgRNA-independent off-target effects induced by the deaminases in base editors [5, 16, 47]. A previous study detected nearly no sgRNA-dependent off-target [48]. In contrast, base editors have been reported to generate significant amounts of off-target single-nucleotide mutations, and these off-target mutations are mostly induced by deaminases in a sgRNA-independent manner [5, 16, 47]. Through Cas-OFFinder prediction, the breakpoints of off-target SVs showed poor sequence similarity with the protospacer sequence, and no difference was found among ABE, CBE, and Cas9 groups. This evidence indicated that these off-target SVs in base editors were unlikely sgRNA-dependent. Motif analysis further showed that ABE-induced SVs had a strong preference for the TAT motif and significantly higher A percentage around breakpoints, consistent with previous findings [5]. For CBE, we did not find any motif preference, which in part explained why ABE induced more SVs than CBE in GOTI samples. Further WGS data in T cells validated the motif pattern of ABE-induced off-target SVs. These results suggested that the higher number of off-target SVs induced by ABE was primarily attributed to TadA, leading to the activation of endogenous DNA damage repair pathways such as base excision repair (BER) pathway-mediated deoxyinosine repair [49]. The high-fidelity version of ABE (ABE8e-V106W) has been reported to substantially reduce off-target A to I in RNA and A•T-to-G•C in DNA [18]. Our results supported that ABE8e-V106W indeed reduced the off-target SVs compared with ABE8.8. However, off-target SVs were still detectable in ABE8e-V106W, raising a direction for further evolution of high-fidelity versions of ABE. Cellular repair-related genes have been reported to play important roles in generating large deletions in Cas9 [50] and CGBE [21]. Thus, the repair mechanism responsible for the formation of SVs in the genome by ABE needs further genetic screening.

Aneuploidy is a distinct feature of cancer cells and is associated with most types of cancers [51]. Loss of chr15 has been reported to be associated with several tumors, such as head and neck carcinomas [52]. In adult T-cell leukemia, chr15 loss infrequently occurs [53]. Deletions of 15q have been reported in childhood acute lymphoblastic leukemia [54]. Both ABE and Cas9 systems have been widely applied in the treatment of

Wu *et al. Genome Biology*   (2024) 25:291

Page 12 of 22

various diseases [45, 55–57]. In immunotherapies, Cas9 has been extensively used to engineer T cells for the development of potent immunotherapies [58–60]. Our results, along with previous findings [4, 7], suggest the possibility that Cas9-induced aneuploidy and chromosomal truncation might be associated with an increased risk of tumorigenesis. For the first time, we found severe off-target structure variations in the genome of ABE-edited cells and in different cell types. We observed differences between Cas9- and ABE-induced aneuploid cells. Though similar numbers of genomic structural variations were detected in both Cas9- and ABE-treated cells, different pathways were activated at the transcriptomic level. For ABE, cell cycle-related genes significantly expressed, and cells are arrested in the early stage. Cell cycle can be arrested by p53 under stress signals such as DNA damage, and further apoptosis is induced [38]. Previous study shows that Cas9 selects for p53-inactivating mutations [61]. Although we did not detect significantly unregulated apoptosis scores in ABE, these cell cycle arrest shows DNA-damaged cells were found by a guard of cells [38]. This indicates the necessity of detecting off-target genomic structure variations in clinical applications for ABE. Nevertheless, with future development of high-fidelity versions of base editors, the impact of off-target variations could be significantly minimized.

Conclusions

The GOTI off-target detection system and our developed Bayesian method based on InferCNV provide an unbiased, high-throughput approach to identify nuclease-induced chromosomal aberrations. Our discovery of chromosomal truncation in mouse embryos and human T cells as well as other cell types highlights the potential oncogenic risks associated with both ABE and Cas9. The comprehensive characterization of these off-target SVs and an examination of high-fidelity ABE (ABE8e-V106W) support that these off-target SVs were induced by TadA in a sgRNA-independent manner and could be reduced by deaminase engineering. Together, our study emphasizes the need for mitigation strategies to minimize the off-target effects caused by ABE and CRISPR/Cas9 in further clinical applications.

Methods

Whole genome sequencing

A total amount of 0.2 μg DNA of T cells per sample was used as input material for the DNA library preparations. Human PBMCs were provided by Milestone Biological Science & Technology (Shanghai, China). PBMCs were resuspended at 1×10^6 cells/mL in X-VIVOTM 15 Serum-free Hematopoietic Cell Medium (Lonza), supplemented with 10mM neutralized NAC (Sigma-Aldrich), and 10 ng/mL IL-7/IL-15. Briefly, the genomic DNA sample was fragmented by Covaris LE220R-plus (Covaris, USA) to a size of 350 bp. Then DNA fragments were endpolished, A-tailed, and ligated with the full-length adapter for Illumina sequencing, followed by further PCR amplification. PCR products were purified by the AMPure XP system (Beckman Coulter, Beverly, USA). Subsequently, library quality was assessed on the Agilent 5400 system (AATI) and quantified by real-time PCR (1.5 nM). The qualified libraries were pooled and sequenced on Illumina platforms with PE150 strategy in Novogene Bioinformatics Technology Co., Ltd (Beijing, China), according to effective library concentration and the data amount required. The

original fluorescence image files are transformed into short reads (Raw data) by base calling and these short reads are recorded in FASTQ format, which contains sequence information and corresponding sequencing quality information. Sequence artifacts, including reads containing adapter contamination, low-quality nucleotides, and unrecognizable nucleotide (N), undoubtedly set the barrier for the subsequent reliable bioinformatics analysis. Hence, quality control is an essential step applied to guarantee meaningful downstream analysis. We used Fastp to perform basic statistics on the quality of the raw reads. The steps of data processing were as follows:

(1) Discard a pair of reads if either one of them contains adapter contamination (>10 nucleotides aligned to the adapter, allowing $\leq 10\%$ mismatches).
(2) Discard a pair of reads if more than 10% of bases are uncertain in either one of the reads.
(3) Discard a pair of reads if the proportion of low-quality (Phredquality <5) bases is over 50% in either one of the reads.

Structure variant calling

Based on a previous comprehensive benchmark of structure variation callers [62], Gridss and Manta were selected in this study to perform structure variation (SV) calling. Whole genome sequencing data using the genome-wide off-target analysis by two-cell embryo injection (GOTI) method were obtained from previously published data [16]. Briefly, a mixture of Cre, gene editing messenger RNA (Cas9/BE3:rAPOBEC1-nCas9-UGI/ABE7.10:TadA-TadA*-nCas9), and single guide RNA was injected into one blastomere of two-cell mouse embryo deriving from Ai9 male mice with wildtype female mice [63]. Cre was only injected into one of two cells of the embryo and generated a chimeric embryo with half of the cells labeled with tdTomato (colored red) with the Cre-loxP system [63, 64]. Thus, the cells that received editing reagents cells can be distinguished. After embryonic day 14.5, through digestion into a single-cell suspension, the tdTomato $+$ cells and tdTomato $-$ cells were then separated by fluorescence-activated cell sorting. The two populations were independently processed for WGS [65]. The sequence data was mapped to the reference genome (GRCm38.p6) and duplicates were removed using SpeedSeq [66]. Somatic SVs were called in the tdtomato $+$ (T $+$) sample compared to the uninjected tdTomato $-$ (T $-$) sample, with Gridss and Manta being used for the calling process. For Gridss [27], each sample was called separately, and a normal panel (PON) was created for subsequent somatic filtering. For Manta [28], the workflow was configured using the "configManta.py" command, with T $+$ and T $-$ samples as inputs. The configured workflow was then executed to call the somatic SVs. An in-house filtering process was developed specifically for large deletions. For each candidate site identified by the callers, an in-house script was used to retrieve all supported reads and calculate the total read depth. Deletions with a size less than 15 kb and an allele frequency larger than 0.15 were considered as qualified SVs, while all deletions larger than 15 kb were retained. For translocations, the presence of direct split reads was considered a necessary criterion for qualifying translocations. Additionally, for deletions of different sizes, the mean depth was calculated using a reasonable bin size. The depth values were

then plotted in bins for the $T+$ and $T-$ samples to validate the deletions in silico. The identified SVs were further visualized using the integrative Genomics Viewer (IGV) [67]. After filtering and manual inspection, the union of SVs detected by Gridss and Manta constituted all the detected SVs, while the high-confidence (HC) SVs were those identified by both callers. ChIPseeker [29] was used for annotating all off-target structure variations, and bedtools [30] was utilized to intersect large deletions with genes. The effect of large deletions on genes was calculated by normalizing the size of the deletions.

Validation of translocation and large deletion

For validation of the detected translocations or large deletions, target sites of interest were amplified by 2 rounds of nested PCR using site-specific primers. Each round of PCR was performed at 95 °C for 3 min, 30 cycles at 95 °C for 30 s, 59 °C for 30 s, 72 °C for 60–120 s, and a final extension at 72 °C for 5 min. The PCR products were analyzed on an agarose gel electrophoresis and then subjected to Sanger sequencing.

Single guide-RNA dependent and independent off-target assessment

The potential off-target sites of all previously designed single guide-RNA (sgRNA) were predicted by Cas-OFFinder [1]. The mismatch numbers were calculated from the outputs from Cas-OFFinder. In addition, sgRNAs were blasted against the flanking site (1KB) of breakpoint for every off-target deletion to assess the sequence similarity between sgRNA and break sites.

T cell, Huh7 cell editing, and single-cell sequencing

In this study, two target genes, *PDCD1* and *B2M*, were edited. The same single guide RNA (sgRNA) used by Gaudelli et al. [31] to target *B2M* was also used in our study, successfully targeting both ABE and Cas9 through splice site disruption. For *PDCD1*, we used the same sgRNA as Nahmad et al. [4] did for Cas9 and Gaudelli et al. [31] did for ABE. For high-fidelity ABE, the same sgRNA targeting *PDCD1* was used. For Huh7, the same sgRNA targeting *B2M* was used. The Huh-7 cell line was purchased from the National Collection of Authenticated Cell Cultures (China) (Serial No.: TChu182; Identifier No.: CSTR:19,375.09.3101HUMTCHu182) and was confirmed to be free of mycoplasma contamination. Prior to the experiment, T Cell Medium (Lonza X vivo5 + 5% human AB serum + NAC + 10 ng/ml IL7 and 10 ng/ml IL15) was prepared and allowed to equilibrate at room temperature for 20 min within a biological safety cabinet. Cryopreserved PBMCs were then retrieved from a liquid nitrogen tank, thawed in a 37 °C water bath, and transferred into the cabinet after removing any remaining ice clumps. The outer surface of the container was sterilized with 5% ethanol, dried, and transferred into the cabinet. The cells were then transferred into 9 mL of T cell medium and centrifuged at 400 g for 5 min at room temperature. After discarding the supernatant, the cells were resuspended in 5 mL of medium and counted using a ThermoFisher Attune NxT flow cytometer. Subsequently, the cells were seeded at a density of 1e6 cells/mL in plates or culture flasks. After 6 h of PBMC revival, CD3/28 dynabeads were added for activation. Following 3 days of cell activation, the activation beads were removed using a magnet and the culture was continued for 24 h. For electroporation, 24-well plates were prepared by adding 900 μL of PBMC complete medium per well and incubating

at 37 ℃ for 10 min. The Lonza 4D Nucleofector System with the P3 Primary Cell 4D kit was used, following the kit instructions for electroporation. Each electroporation reaction consisted of 1E6 cells, 1.2 pmol of EPIREG mRNA, and 40 pmol of sgRNA for each target gene. Immediately after electroporation, 100 μL of cell-free medium was added to each well, followed by a 1-h incubation at 37 ℃. The cells were then transferred to 24-well plates for further culture. To assess changes in the protein levels of *PDCD1* and *B2M* on T cell surfaces using flow cytometry, approximately 1e5 cells were transferred to 1.5 mL Eppendorf tubes and centrifuged at 400 g for 5 min at room temperature. The supernatant was discarded and 50 μL of the corresponding dilution ratio of antibodies was added. The tubes were incubated at 4 ℃ in the dark for 30 min, followed by the addition of 500 μL of MACS buffer to each tube. After centrifuging at 400 g for 5 min at room temperature, the supernatant was discarded and 300 μL of MACS buffer was added to each tube, gently mixed by pipetting. The samples were then analyzed using a ThermoFisher Attune NxT flow cytometer. The $10 \times$ Genomics® Cell Preparation Guide was followed for washing, counting, and concentrating cells from both abundant and limited cell suspensions (greater than or less than 100,000 total cells, respectively) in preparation for use in $10 \times$ Genomics Single Cell Protocols. The cell suspension was loaded into Chromium microfluidic chips with 3′ (v2 or v3, depending on the project) chemistry and barcoded with a $10 \times$ Chromium Controller (10X Genomics). RNA from the barcoded cells was subsequently reverse-transcribed and sequencing libraries were constructed using reagents from a Chromium Single Cell 3′ v2 (v2 or v3, depending on the project) reagent kit (10X Genomics) according to the manufacturer's instructions. Sequencing was performed with Illumina (HiSeq 2000 or NovaSeq, depending on the project) according to the manufacturer's instructions (Illumina). The same library construction process was used in Huh7 cells.

Single-cell RNA-seq data analysis for CRISPR/Cas9 or ABE editing-treated T cells, B cells, and Huh7 cells

The scRNA-seq reads were aligned to the GRCh38 reference genome and quantified using cellranger count ($10 \times$ Genomics, v7.1.0). A filtered and qualified feature matrix was generated, which only included barcodes with unique molecular identifier (UMI) counts surpassing the threshold for cell detection. The raw expression matrix was processed using the scanpy packages in Python. A total of 217,033 cells were detected initially. After applying quality control metrics (total counts < 50,000, mitochondrial ratio < 10, and detected gene numbers between 500 and 8000), 209,469 cells remained for further analysis. The qualified raw count data was log-normalized per cell with a size factor of 10,000. The top 3000 genes with the highest variations across all cells were selected as "highly variable genes." A subset of the expression matrix was extracted and scaled using these highly variable genes. Principal component analysis (PCA) was then applied to derive meaningful features from the scaled data. The elbow point in the variance ratio distributions of PCA was used to select fifty principal components for computing the neighboring cluster graph. To further reduce dimensionality, Uniform Manifold Approximation Projection (UMAP) with the Leiden algorithm was applied (resolution = 1.2). For ABE8.8, a total of 19 clusters were obtained and visualized using UMAP. We removed cluster 13 based on the marker genes specifically expressed in

Wu *et al. Genome Biology*   (2024) 25:291

Page 16 of 22

MT-related genes. For high-fidelity ABE, a total of 19 clusters were obtained and visualized using UMAP. Next, we performed differential expression analysis in each cluster using the Wilcoxon rank sum-test method. The top differentially expressed genes in each cluster, as determined by ranking, were defined as marker genes and used for subsequent cell type annotations.

InferCNV and identification of aberrant genes and cells using the Bayesian model

Before conducting InferCNV analysis, for *B2M*, we used control expression as a reference to establish a normal distribution and extracted cells that showed significant repression. For *PDCD1*, we randomly sampled 600 cells for efficient InferCNV analysis. Raw gene expression data were extracted from the AnnData object following the recommendations in the "Using 10 × data" section of the InferCNV of the Trinity CTAT Project (https://github.com/broadinstitute/inferCNV). InferCNV is a single-cell method developed to infer copy number changes from gene expression profiles [68, 69]. In our analysis, the non-targeted T cell population served as the reference, while the edited-targeted population (CRISPR/Cas9, ABE) was tested. The following parameters were used: "denoise" with default hidden Markov model (HMM) settings and a cutoff value of 0.1. Additionally, we developed a Bayesian model to eliminate background noise.

From the InferCNV reference, we obtained two matrices for the reference and observation data. We then used the fitter package to fit the data using the generalized Gaussian distribution (GGD) and obtained the estimated parameter values. The probability density function of the generalized Gaussian distribution was

$$f(x, \beta, \alpha, \gamma) = \frac{1}{2 * \alpha * \Gamma\left(1 + \frac{1}{\beta}\right)} e^{-\left(\frac{|x-\gamma|}{\alpha}\right)^{\beta}}$$

Next, use a Bayesian model to fit the data. We assumed that the parameters in the model had a wide uniform distribution as the prior distribution (as shown in Additional file 2: Table S1).

We then utilized the Markov chain Monte Carlo (MCMC) method to fit the generalized normal distribution model to the data. The MCMC method was implemented using the Gibbs sampling approach through the R package nimble. By using the sample mean as the parameter estimates for the model, we obtained α_{oi}, β_{oi} and γ_{oi}. Finally, we employed the stats.gennorm.cdf() function from the scipy Python package to calculate the cumulative probability density value for each data point in the posterior distribution. If a data point had a value less than 0.05, the corresponding gene was considered to be affected by the editing tool. This process was applied to the data of each gene, resulting in the identification of an affected gene list.

Next, for each cell, we computed the mean InferCNV scores on chromosomes 1–22 using the obtained gene list. Cells with a mean inferred copy number variation (CNV) value higher or lower than 2 standard deviations from the mean of the reference cells were classified as aneuploid cells with gains or losses, respectively. To confirm the reliability of the signals, we performed a reverse process using observation cells as the reference. The aneuploid cell ratio was calculated as the percentage of forward cells divided by the percentage of reverse cells in both the editing and non-targeting

control (NTC) groups. To measure the background noise level, we applied the same process to the NTC blank groups using genes from the entire chromosome, calculating the aneuploid cell ratio.

We employed this method in 8 cell types to detect aberrant genes and cells. Subsequently, we identified cells that harbored deletions in single and multiple chromosomes. For long-term cell fitness evaluation, we utilized the same gene list detected at day 3 to monitor changes in the percentage of aneuploid cells at day 7 and day 21. We treated ABE-treated Huh7 cells as a cluster and applied the same analysis for detecting aberrant genes and cells.

Differential gene expression analysis and enrichment analysis

Differentially expressed genes (DEGs) were identified using DESeq2 [70] in six cell types targeting *PDCD1* and *B2M* in the CRISPR/Cas9 and ABE groups. Normal cells were used as the reference group, while single and multi-chromosome loss cells were used as the case group.

To further analyze the enrichment of DEGs, we utilized the web-server Metascape [71] (https://metascape.org/gp/index.html). Metascape provides comprehensive functional annotation and enrichment analysis for gene sets.

Cell cycle analysis

Cell cycle states were determined using the data and methods described in previous studies [7, 72]. In our dataset, a total of 577 marker genes were identified and used to define the different cell cycle states, including M, M/G1, G1/S, S, and G2/M. To ensure the accuracy of the marker genes, only those genes that showed a correlation with the average normalized expression across the other marker genes in the dataset were retained. Subsequently, each cell was assigned a score for each of the five cell cycle states, which was calculated as the average normalized expression of the marker genes associated with that state. These scores were then normalized across cells to obtain z-values.

If all five z-values were negative, indicating low expression in all cell cycle states, the cell was labeled as non-cycling (G0). On the other hand, if at least one z-value was positive, the cell was assigned a label corresponding to the cell cycle state with the highest z-value.

TP53 and apoptosis pathway scoring

TP53 and apoptosis pathway-related genes were obtained from the GSEA Molecular Signatures Database (MSigDB) portal (https://www.gseamsigdb.org/gsea/msigdb/). To assess the activity of these genes in each cell, we employed GSVA (Gene Set Variation Analysis) [73] with the "ssgsea" method.

To determine significant differences between normal and aneuploidy cells, a one-sided t-test was performed. This test allowed us to examine whether the expression levels of *TP53* and apoptosis pathway-related genes significantly differed between the two groups.

Wu *et al. Genome Biology*   (2024) 25:291

Page 18 of 22

ABE editing efficiency analysis

High-throughput sequencing data were processed with CRISPResso2 to assess the editing efficiency of ABE and ABE-V106W targeting the *B2M* or *PDCD1* gene in PBMC or Huh7 cells. Paired-end FASTQ files, along with the amplicon sequence or WGS data and guide RNA sequence, were used as the inputs for the analysis. Quality filtering of input reads was performed using a Phred quality score threshold of 33. Subsequently, the reads were aligned to the reference amplicon sequence using the alignment algorithm of CRISPResso2 [74] with default parameters. Upon determining the conversion of A•T-to-G•C based on the guide RNA sequence provided, editing events were assigned and quantified to calculate the proportion of reads exhibiting modifications. The modification percentage, which was calculated by comparing the number of reads showing the expected editing event against the total number of reads processed, was used to inform editing efficiency at the target site.

Statistics

To assess the knock-out efficiency in flow cytometry, statistical analyses were conducted using Prism (GraphPad). A *t*-test was performed on the averaged values to determine significant differences between experimental groups. Additionally, a Fisher exact test was utilized to compare the proportions of cell cycle distributions among different conditions. This allowed for the evaluation of any significant variations in the distribution patterns between experimental groups.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-024-03434-0.

Additional file 1: Fig. S1 Sequencing and PCR evidence for detected off-target large deletions. a The percentage of samples with HC structural variations in control, ABE, Cas9, and CBE-edited mouse embryos. b The number of all structural variations detected in non-editing control, ABE, Cas9, and CBE-edited mouse embryos. c IGV plot showing supported reads of on-target large deletions in Cas9-edited mouse embryos. d-f Depth plot showing depth changes around large deletions in ABE, Cas9, and CBE. For Cas9 and ABE, two off-target large deletions were validated by PCR. g Distribution of the A base percentage within 500bp flanking the breakpoints of SVs in ABE-, Cas9-treated mice and randomly sampled background controls. h-i Similarity of the on-target sequence and sequences surrounding break points of SVs. The ratio of matching bases between the on-target sequence and all off-target SVs at each position was displayed at the top heatmap. The top 5 off-target sequences showing the highest similarity with the on-target sequence was shown. Dots represent matched bases. j Distribution of different annotated types for off-target structural variations. k Affected genes number normalized by deletion size for ABE and Cas9 after adjusting overlapping fractions from 0.05 to 1.0. Fig. S2 Evidence of off-target structural variations in primary human T cells using single-cell RNA-seq. a The expression levels of the B2M and PDCD1 genes in different cell groups, where the y-axis represents the normalized and log-transformed values. b InferCNV results for CD4 effector memory T cells, CD8 effector memory T cells, cytotoxic CD8 T cells, naive CD4 T cells, naive CD8 T cells, and proliferating T cells on Cas9 and ABE for the B2M and PDCD1 targets. c Proportions of cells harboring deletions in Cas9 and ABE editing group for B2M. d The number of high-confidential (HC) SVs in ABE-treated cells targeting B2M or PDCD1. e Enriched motifs of sequences within 10bp flanking the breakpoints of HC SVs in ABE-treated cells targeting B2M or PDCD1. f Depth plot showing depth changes around big deletions and supporting sequencing reads of translocation shown by IGV in ABE-treated cells targeting B2M. The translocation located in the inferred region in B2M. g Depth plot showing depth changes around big deletions and supporting sequencing reads of translocation shown by IGV in ABE-treated cells targeting PDCD1. Fig. S3 Frequency of significant genes and pathway scores among different cell types in Cas9 and ABE. a The frequency of differentially expressed genes among different cell types. b The scores of p53 and apoptosis pathways for all cell types in the Cas9-editing B2M group. Fig. S4 Longitudinal single-cell data monitoring the fitness of aneuploid cells at different cell types over time. a The proportions of cells with deletions (represented in blue) for CD4 effector memory T cells, CD8 effector memory T cells, cytotoxic CD8 T cells, naive CD4 T cells, naive CD8 T cells, and proliferating T cells on Cas9 and ABE-edited B2M and PDCD1 at day 3, day 7, and day 21. Fig. S5 Off-target SVs in B cells and Huh7 cells and in ABE-V106W. a The expression levels of the B2M and PDCD1 genes in B cells, where the y-axis represents the normalized and log-transformed values. b Distribution of aneuploid cell ratio across the whole genome for Cas9 and ABE. Red Dot represents aneuploid cell ratio (proportion/reverse proportion) for B cell type detected at chromosomes. c InferCNV results for B cells in Cas9 and ABE targeting PDCD1. d InferCNV results for

Huh7 cells in Cas9 and ABE targeting B2M. e Uniform manifold approximation and projection (UMAP) plot depicting 19650 cells representing 8 cell lineages. f Dot plot showing averaged expression levels of marker genes across 8 cell types. g InferCNV results for different T cell types in ABE-V106W targeting PDCD1

Additional file 2: Table S1. Prior distribution of parameters.

Additional file 3: Table S2. Off-target structure variations detected by Gridss or Manta.

Additional file 4: Table S3. Raw cell numbers detected in each sample.

Additional file 5: Table S4. Top 100 markers for each leiden.

Additional file 6: Table S5. List of SVs from WGS and gene list from MCMC in scRNA-seq.

Additional file 7: Table S6. Aberrant gene number, annotaion and enriched pathway in every chromose for every cell types.

Additional file 8: Review history.

Authors' contributions
LL-W, MS-S, PZ-H, and YingQi-Li analyzed the data and interpreted results. LL-W wrote the manuscript. ST-J and YaQin-Li performed the T cell editing of Cas9 and ABE. TL-Y performed GOTI experiments and structure variation validation. EW-Z, CY-Z, and YD-S designed and supervised the study. YD-S revised the manuscript. All authors approved the manuscript.

Peer review information
Kevin Pang and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history
The review history is available as Additional file 8.

Data availability
The source codes for the detection of off-target SVs from single-cell transcriptome data are available under the MIT license on GitHub at https://github.com/zhaodalv/ABE_off_target_SVs [75] and on Zenodo at https://zenodo.org/records/13991107 [76]. Whole genome sequencing data of mouse embryos were downloaded from https://www.biosino.org/node/project/detail/OEP000195 [16, 65]. The single-cell transcriptome data of PBMC and Huh-7 cells have been deposited in the National Center for Biotechnology Information (NCBI) GEO dataset with accession GSE280767 [77].

Declarations

Ethics approval and consent to participate
Ethical approval was not needed for this study.

Competing interests
The authors declare that they have no competing interests.

References
1. Bae S, Park J, Kim J-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics. 2014;30:1473–5.
2. Yin J, Liu M, Liu Y, Wu J, Gan T, Zhang W, Li Y, Zhou Y, Hu J. Optimizing genome editing strategy by primer-extension-mediated sequencing. Cell Discovery. 2019;5:18.
3. Zhang C, Yang Y, Qi T, Zhang Y, Hou L, Wei J, Yang J, Shi L, Ong S-G, Wang H, et al. Prediction of base editor off-targets by deep learning. Nat Commun. 2023;14:5358.
4. Nahmad AD, Reuveni E, Goldschmidt E, Tenne T, Liberman M, Horovitz-Fried M, Khosravi R, Kobo H, Reinstein E, Madi A, et al. Frequent aneuploidy in primary human T cells after CRISPR–Cas9 cleavage. Nat Biotechnol. 2022;40:1807–13.
5. Zhou C, Sun Y, Yan R, Liu Y, Zuo E, Gu C, Han L, Wei Y, Hu X, Zeng R, et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. Nature. 2019;571:275–8.
6. Doman JL, Raguram A, Newby GA, Liu DR. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. Nat Biotechnol. 2020;38:620–8.

Wu *et al. Genome Biology*   (2024) 25:291

Page 20 of 22

7.   Tsuchida CA, Brandes N, Bueno R, Trinidad M, Mazumder T, Yu B, Hwang B, Chang C, Liu J, Sun Y, et al. Mitigation of chromosome loss in clinical CRISPR-Cas9-engineered T cells. Cell. 2023;186:4567-4582.e4520.

8.   Höijer I, Emmanouilidou A, Östlund R, van Schendel R, Bozorgpana S, Tijsterman M, Feuk L, Gyllensten U, den Hoed M, Ameur A. CRISPR-Cas9 induces large structural variants at on-target and off-target sites in vivo that segregate across generations. Nat Commun. 2022;13:627.

9.   Adikusuma F, Piltz S, Corbett MA, Turvey M, McColl SR, Helbig KJ, Beard MR, Hughes J, Pomerantz RT, Thomas PQ. Large deletions induced by Cas9 cleavage. Nature. 2018;560:E8–9.

10. Papathanasiou S, Markoulaki S, Blaine LJ, Leibowitz ML, Zhang C-Z, Jaenisch R, Pellman D. Whole chromosome loss and genomic instability in mouse embryos after CRISPR-Cas9 genome editing. Nat Commun. 2021;12:5855.

11. Alanis-Lobato G, Zohren J, McCarthy A, Fogarty NME, Kubikova N, Hardman E, Greco M, Wells D, Turner JMA, Niakan KK. Frequent loss of heterozygosity in CRISPR-Cas9–edited early human embryos. Proc Natl Acad Sci. 2021;118: e2004832117.

12. Weisheit I, Kroeger JA, Malik R, Klimmt J, Crusius D, Dannert A, Dichgans M, Paquet D. Detection of Deleterious On-Target Effects after HDR-Mediated CRISPR Editing. Cell Rep. 2020;31: 107689.

13. Boutin J, Rosier J, Cappellen D, Prat F, Toutain J, Pennamen P, Bouron J, Rooryck C, Merlio JP, Lamrissi-Garcia I, et al. CRISPR-Cas9 globin editing can induce megabase-scale copy-neutral losses of heterozygosity in hematopoietic cells. Nat Commun. 2021;12:4922.

14. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533:420–4.

15. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature. 2017;551:464–71.

16. Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, Yuan L, Steinmetz LM, Li Y, Yang H. Cytosine base editor generates sub-stantial off-target single-nucleotide variants in mouse embryos. Science. 2019;364:289–92.

17. Liu Y, Zhou C, Huang S, Dang L, Wei Y, He J, Zhou Y, Mao S, Tao W, Zhang Y, et al. A Cas-embedding strategy for mini-mizing off-target effects of DNA base editors. Nat Commun. 2020;11:6073.

18. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat Biotechnol. 2020;38:883–91.

19. Zuo E, Sun Y, Yuan T, He B, Zhou C, Ying W, Liu J, Wei W, Zeng R, Li Y, Yang H. A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. Nat Methods. 2020;17:600–4.

20. Li J, Yu W, Huang S, Wu S, Li L, Zhou J, Cao Y, Huang X, Qiao Y. Structure-guided engineering of adenine base editor with minimized RNA off-targeting activity. Nat Commun. 2021;12:2287.

21. Huang ME, Qin Y, Shang Y, Hao Q, Zhan C, Lian C, Luo S, Liu LD, Zhang S, Zhang Y, et al. C-to-G editing generates double-strand breaks causing deletion, transversion and translocation. Nat Cell Biol. 2024;26:294–304.

22. Hasham MG, Donghia NM, Coffey E, Maynard J, Snow KJ, Ames J, Wilpan RY, He Y, King BL, Mills KD. Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombina-tion. Nat Immunol. 2010;11:820–6.

23. Wörmann SM, Zhang A, Thege FI, Cowan RW, Rupani DN, Wang R, Manning SL, Gates C, Wu W, Levin-Klein R, et al. APOBEC3A drives deaminase domain-independent chromosomal instability to promote pancreatic cancer metasta-sis. Nat Cancer. 2021;2:1338–56.

24. Berríos KN, Evitt NH, DeWeerd RA, Ren D, Luo M, Barka A, Wang T, Bartman CR, Lan Y, Green AM, et al. Controllable genome editing with split-engineered base editors. Nat Chem Biol. 2021;17:1262–70.

25. Wang X, Ding C, Yu W, Wang Y, He S, Yang B, Xiong YC, Wei J, Li J, Liang J, et al. Cas12a Base Editors Induce Efficient and Specific Editing with Low DNA Damage Response. Cell Rep. 2020;31: 107723.

26. Yuan B, Zhang S, Song L, Chen J, Cao J, Qiu J, Qiu Z, Chen J, Zhao XM, Cheng TL. Engineering of cytosine base edi-tors with DNA damage minimization and editing scope diversification. Nucleic Acids Res. 2023;51(20):e105.

27. Cameron DL, Schröder J, Penington JS, Do H, Molania R, Dobrovic A, Speed TP, Papenfuss AT. GRIDSS: sensi-tive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res. 2017;27:2050–60.

28. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32:1220–2.

29. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visuali-zation. Bioinformatics. 2015;31:2382–3.

30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

31. Gaudelli NM, Lam DK, Rees HA, Solá-Esteves NM, Barrera LA, Born DA, Edwards A, Gehrke JM, Lee S-J, Liquori AJ, et al. Directed evolution of adenine base editors with increased activity and therapeutic application. Nat Biotechnol. 2020;38:892–900.

32. Hou H, Sun D, Zhang X. The role of MDM2 amplification and overexpression in therapeutic resistance of malignant tumors. Cancer Cell Int. 2019;19:216.

33. Junker K, Hindermann W, von Eggeling F, Diegmann J, Haessler K, Schubert J. CD70: a new tumor specific biomarker for renal cell carcinoma. J Urol. 2005;173:2150–3.

34. Held-Feindt J, Mentlein R. CD70/CD27 ligand, a member of the TNF family, is expressed in human brain tumors. Int J Cancer. 2002;98:352–6.

35. Dastsooz H, Cereda M, Donna D, Oliviero S. A Comprehensive Bioinformatics Analysis of UBE2C in Cancers. Int J Mol Sci. 2019;20:2228.

36. Yu K, Ji Y, Liu M, Shen F, Xiong X, Gu L, Lu T, Ye Y, Feng S, He J. High Expression of CKS2 Predicts Adverse Outcomes: A Potential Therapeutic Target for Glioma. Frontiers in Immunology. 2022;13:881453.

Wu *et al. Genome Biology*   (2024) 25:291

Page 21 of 22

37. Schiroli G, Conti A, Ferrari S, della Volpe L, Jacob A, Albano L, Beretta S, Calabria A, Vavassori V, Gasparini P, et al. Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. Cell Stem Cell. 2019;24:551-565.e558.

38. Hernández Borrero LJ, El-Deiry WS. Tumor suppressor p53: Biology, signaling pathways, and therapeutic targeting. Biochim Biophys Acta Rev Cancer. 2021;1876:188556.

39. Wu Y, Zhou H, Fan X, Zhang Y, Zhang M, Wang Y, Xie Z, Bai M, Yin Q, Liang D. Correction of a genetic disease by CRISPR-Cas9-mediated gene editing in mouse spermatogonial stem cells. Cell Res. 2015;25:67–79.

40. Fu Y-W, Dai X-Y, Wang W-T, Yang Z-X, Zhao J-J, Zhang J-P, Wen W, Zhang F, Oberg KC, Zhang L. Dynamics and competition of CRISPR–Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. Nucleic Acids Res. 2021;49:969–85.

41. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8:2281–308.

42. Liu Y, Yin J, Gan T, Liu M, Xin C, Zhang W, Hu J. PEM-seq comprehensively quantifies DNA repair outcomes during gene-editing and DSB repair. STAR Protoc. 2022;3: 101088.

43. Egli D, Zuccaro MV, Kosicki M, Church GM, Bradley A, Jasin M. Inter-homologue repair in fertilized human eggs? Nature. 2018;560:E5–7.

44. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol. 2018;36:765–71.

45. Musunuru K, Chadwick AC, Mizoguchi T, Garcia SP, DeNizio JE, Reiss CW, Wang K, Iyer S, Dutta C, Clendaniel V, et al. In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. Nature. 2021;593:429–34.

46. Pacesa M, Lin CH, Cléry A, Saha A, Arantes PR, Bargsten K, Irby MJ, Allain FHT, Palermo G, Cameron P, et al. Structural basis for Cas9 off-target activity. Cell. 2022;185:4067-4081.e4021.

47. Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, Liang C, Wang D, Qiu J-L, Zhang F, Gao C. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science. 2019;364:292–5.

48. Kim D, Lim K, Kim ST, Yoon SH, Kim K, Ryu SM, Kim JS. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nat Biotechnol. 2017;35:475–80.

49. Gu S, Bodai Z, Cowan QT, Komor AC. Base Editors: Expanding the Types of DNA Damage Products Harnessed for Genome Editing. Gene Genome Ed. 2021;1:100005.

50. Kosicki M, Allen F, Steward F, Tomberg K, Pan Y, Bradley A. Cas9-induced large deletions and small indels are controlled in a convergent fashion. Nat Commun. 2022;13:3422.

51. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. Nat Rev Genet. 2020;21:44–62.

52. Poetsch M, Kleist B. Loss of heterozygosity at 15q21.3 correlates with occurrence of metastases in head and neck cancer. Modern Pathology. 2006;19:1462–9.

53. Kamada N, Sakurai M, Miyamoto K, Sanada I, Sadamori N, Fukuhara S, Abe S, Shiraishi Y, Abe T, Kaneko Y, et al. Chromosome abnormalities in adult T-cell leukemia/lymphoma: a karyotype review committee report. Cancer Res. 1992;52:1481–93.

54. Heerema NA, Sather HN, Sensel MG, La MKL, Hutchinson RJ, Nachman JB, Reaman GH, Lange BJ, Steinherz PG, Bostrom BC, et al. Abnormalities of chromosome bands 15q13-15 in childhood acute lymphoblastic leukemia. Cancer. 2002;94:1102–10.

55. Rothgangl T, Dennis MK, Lin PJC, Oka R, Witzigmann D, Villiger L, Qi W, Hruzova M, Kissling L, Lenggenhager D, et al. In vivo adenine base editing of PCSK9 in macaques reduces LDL cholesterol levels. Nat Biotechnol. 2021;39:949–57.

56. Frangoul H, Altshuler D, Cappellini MD, Chen Y-S, Domm J, Eustace BK, Foell J, de la Fuente J, Grupp S, Handgretinger R, et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. N Engl J Med. 2020;384:252–60.

57. Xu L, Wang J, Liu Y, Xie L, Su B, Mou D, Wang L, Liu T, Wang X, Zhang B, et al. CRISPR-Edited Stem Cells in a Patient with HIV and Acute Lymphocytic Leukemia. N Engl J Med. 2019;381:1240–7.

58. Zhang J, Hu Y, Yang J, Li W, Zhang M, Wang Q, Zhang L, Wei G, Tian Y, Zhao K, et al. Non-viral, specifically targeted CAR-T cells achieve high safety and efficacy in B-NHL. Nature. 2022;609:369–74.

59. Lu Y, Xue J, Deng T, Zhou X, Yu K, Deng L, Huang M, Yi X, Liang M, Wang Y, et al. Safety and feasibility of CRISPR-edited T cells in patients with refractory non-small-cell lung cancer. Nat Med. 2020;26:732–40.

60. Stadtmauer EA, Fraietta JA, Davis MM, Cohen AD, Weber KL, Lancaster E, Mangan PA, Kulikovskaya I, Gupta M, Chen F, et al. CRISPR-engineered T cells in patients with refractory cancer. Science. 2020;367:eaba7365.

61. Enache OM, Rendo V, Abdusamad M, Lam D, Davison D, Pal S, Currimjee N, Hess J, Pantel S, Nag A, et al. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. Nat Genet. 2020;52:662–8.

62. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol. 2019;20:117.

63. Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR, et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci. 2010;13:133–40.

64. Wang L, Li M-Y, Qu C, Miao W-Y, Yin Q, Liao J, Cao H-T, Huang M, Wang K, Zuo E, et al. CRISPR-Cas9-mediated genome editing in one blastomere of two-cell embryos reveals a novel Tet3 function in regulating neocortical development. Cell Res. 2017;27:815–29.

65. Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, Yuan L, Steinmetz LM, Li Y, Yang H. GOTI, a method to identify genome-wide off-target effects of genome editing in mouse embryos. Nat Protoc. 2020;15:3009–29.

66. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, Marth GT, Quinlan AR, Hall IM. SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat Methods. 2015;12:966–8.

67. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–92.

68. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;344:1396–401.

Wu *et al. Genome Biology*   (2024) 25:291

Page 22 of 22

69. Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352:189–96.

70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

71. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10:1523.

72. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015;161:1202–14.

73. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 2013;14: 7.

74. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, Pinello L. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol. 2019;37:224–6.

75. Wu L, Shi M, Sun Y. ABE_off_target_SVs. Github; 2024. https://github.com/zhaodalv/ABE_off_target_SVs.

76. Wu L, Shi M, Sun Y. ABE_off_target_SVs: Code for adenine base editors induce off-target structure variations. Zenodo. 2024. https://doi.org/10.5281/zenodo.13991107.

77. Wu L, Sun Y. Adenine base editors induce off-target structure variations in mouse embryos and primary human T cells. In: DataSets. Gene Expression Omnibus. 2024. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE280767.

Publisher's Note