nature communications

Article

https://doi.org/10.1038/s41467-024-53806-0

Transfer of mitochondrial DNA into the nuclear genome during induced DNA breaks

Received: 18 January 2024

Accepted: 22 October 2024

Published online: 01 November 2024

Check for updates

Jinchun Wu[1,5], Yang Liu[1,5], Liqiong Ou[1,5], Tingting Gan[1,2,5], Zhengrong Zhangding[1], Shaopeng Yuan[1], Xinyi Liu[1], Mengzhu Liu[1], Jiasheng Li[3,4], Jianhang Yin[1], Changchang Xin[1], Ye Tian[3,4] & Jiazhi Hu[1,2]

Mitochondria serve as the cellular powerhouse, and their distinct DNA makes them a prospective target for gene editing to treat genetic disorders. However, the impact of genome editing on mitochondrial DNA (mtDNA) stability remains a mystery. Our study reveals previously unknown risks of genome editing that both nuclear and mitochondrial editing cause discernible transfer of mitochondrial DNA segments into the nuclear genome in various cell types including human cell lines, primary T cells, and mouse embryos. Furthermore, drug-induced mitochondrial stresses and mtDNA breaks exacerbate this transfer of mtDNA into the nuclear genome. Notably, we observe that mitochondrial editors, including mitoTALEN and recently developed base editor DdCBE, can also enhance crosstalk between mtDNA and the nuclear genome. Moreover, we provide a practical solution by co-expressing TREX1 or TREX2 exonucleases during DdCBE editing. These findings imply genome instability of mitochondria during induced DNA breaks and explain the origins of mitochondrial-nuclear DNA segments.

The human genome consists of two copies of nuclear DNA and a few hundred to a few thousand copies of mitochondrial DNA (mtDNA), which are separated by the membrane system. During the endosymbiotic evolution, mtDNA undergoes natural transfer to the nuclear genome[1]. Physiological and pathological mitochondrial stresses that promote the production of reactive oxygen species damage mtDNA and facilitate the release of fragmented mtDNA[2–4]. Consequently, the human genome contains hundreds of nuclear-mitochondrial DNA segments (NUMTs), and de novo NUMTs were detected in approximately one per 10^4 births and 10^3 cancers[5–7]. Meanwhile, NUMTs in tumor cells tend to be embedded in gene-rich regions of nuclear DNA, suggesting that they may be the causative integration of certain tumors[6,8].

Recently, the focus of gene editing has extended from the nuclear DNA[9–12] to mtDNA[13,14], leading to the development of a series of highly effective mitochondrial editors. These tools aim to eliminate mutated mtDNA by inducing DNA double-stranded breaks (DSBs) with mitochondrial-targeted transcription activator-like effector nucleases (mitoTALEN)[15,16] or correcting mutations with mitochondrial base editors[17,18]. DNA insertions of genomic DNA, vector, or viral DNA origins have been widely observed during gene editing with TALEN or clustered regularly interspaced short palindromic repeats associated system (CRISPR-Cas)[19–21]. Rapid degradation of damaged mtDNA has been found to facilitate the elimination of mutated mtDNA by nucleases, leading to the generation of small mtDNA fragments[15,22,23]. Given the formation of NUMTs, it remains to be investigated whether mitochondrial or nuclear editing can accelerate the transfer of mtDNA segments into the nuclear genome. Moreover, it's worth investigating whether mtDNA fragments can be integrated into various types of DSBs including DSBs induced by gene editors.

[1]State Key Laboratory of Protein and Plant Gene Research, Genome Editing Research Center, School of Life Sciences, PKU-THU Center for Life Sciences, Peking University, Beijing, China. [2]Peking University Chengdu Academy for Advanced Interdisciplinary Biotechnologies, Chengdu, Sichuan, China. [3]State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [4]University of Chinese Academy of Sciences, Beijing, China. [5]These authors contributed equally: Jinchun Wu, Yang Liu, Liqiong Ou, Tingting Gan.
e-mail: hujz@pku.edu.cn

Here we find that genome editing on the nuclear DNA causes mtDNA-nuclear DNA fusion at the CRISPR-Cas9-target sites in vitro and in vivo, validated by both primer-extension-mediated sequencing (PEM-seq) and target sequencing. Moreover, high-fidelity Cas9 variants are unable to eliminate the transfer of mtDNA to nuclear DNA. Mitochondrial stresses increase the level of mtDNA-nuclear DNA fusion, and mitochondrial DSBs introduced by mitoTALEN also cause mtDNA fragility and result in the transfer of mtDNA to nuclear DNA. Moreover, mitochondrial editors including both mitoTALEN and mitochondrial DddA-derived cytosine base editors (DdCBEs) also exhibit a discernible level of mtDNA-nuclear DNA fusion. Finally, we find that co-expression of either TREX1 or TREX2 exonuclease in the mitochondria could be a plausible solution to suppress mtDNA fusion with the nuclear DNA during DdCBE treatment.

Results

Mitochondrial DNA fuses to CRISPR-Cas targeting sites in the nuclear genome

To investigate the integration of mtDNA segments into nuclear CRISPR-Cas target sites, we examined potential mtDNA sequences in the CRISPR-Cas editing products captured by PEM-seq[24]. PEM-seq utilizes a biotinylated primer adjacent to the "bait" target site for a one-round primer extension to generate biotinylated DNA products containing "prey" sequences fused to the target DNA breaks (Fig. 1a). Subsequently, the single-stranded DNA products are isolated by streptavidin beads to remove all the un-biotinylated DNA including nuclear genomic DNA and mtDNA. Then the enriched biotinylated DNA is subjected to two rounds of PCR for library preparation (Fig. 1a and Supplementary Fig. 1a)[24]. Leveraging the information of prey sequence, PEM-seq is capable of identifying insertions, deletions, vector integrations, and chromosomal translocations (Supplementary Fig. 1a)[24]. During data processing, we only counted the chimeric reads where the prey sequences were better aligned to the mtDNA than NUMTs as fusion junctions between mtDNA and the nuclear DNA (termed as mt-nuclear DNA fusions hereafter; Fig. 1a and Supplementary Fig. 1a, b). Moreover, we conducted parallel PEM-seq analysis in both edited and unedited samples to preclude potential errors in primer annealing, library preparation, and sequencing (Fig. 1a). By analyzing previously published PEM-seq libraries[25], we identified multiple mt-nuclear DNA fusions from varied CRISPR target sites, including *CLIC4*, *KLHL29*, *NLRC4*, *COL8A1*, *NUDT16*, *LNX1*, *FGF18*, *VEGFA_1/2*, *HBB*, *IFNγ*, and *P2RX5-TAX1BP3* in HEK293T cells, edited with CRISPR-Cas editors including *Sp*Cas9, *As*Cas12a, *Lb*Cas12a, *Un1*Cas12f, CasMINI, and CasMINI_ge4.1 (Fig. 1b, Supplementary Fig. 1c, Supplementary Table 1)[25]. The frequency of mt-nuclear DNA fusions at CRISPR-Cas target sites varied from one per 10^3 to 10^5 editing events (Fig. 1b), accounting for up to 1% of insertions (>1 bp) in HEK293T cells (Supplementary Fig. 1d), with the Cas12 family exhibiting slightly lower levels of mtDNA integrations in comparison to *Sp*Cas9 (Fig. 1b–e and Supplementary Fig. 1e–g). Of note, Cas12a exhibited similar editing efficiency as *Sp*Cas9 while Cas12f had a slightly lower editing efficiency, suggesting that stagger cutting may affect the formation of DNA fusions (Supplementary Fig. 1c), consistent with previous report[25]. In contrast, unedited samples showed no mt-nuclear DNA fusions for 10 of the 12 tested loci. Only two editing events out of 625,886 for *VEGFA_1* and one editing event out of 900,733 for *HBB* were detected in the unedited cells (Fig. 1b and Supplementary Table 1). However, these three events contained the same sequences as those observed in the corresponding edited cells (Supplementary Table 1), indicating that they might originate from clustering errors during Illumina sequencing.

High-fidelity *Sp*Cas9 variants such as e*Sp*Cas9, HF1, FeCas9, evo-Cas9, HiFiCas9, Hypa, LZ3, and Sniper Cas9, as well as PAM-flexible *Sp*Cas9 variants including Cas9-NG, xCas9, *Sp*G, and *Sp*RY, have been developed to enhance editing specificity or broaden editing scope, respectively[26,27]. Through analyzation of published PEM-seq samples upon *Sp*Cas9 variants editing[27], we found that these *Sp*Cas9 variants exhibited comparable levels of mt-nuclear DNA fusions to wild-type *Sp*Cas9 (Fig. 1f and Supplementary Fig. 2a, b). Considering the inheritability of editing products in embryonic stem cells (ESCs), we performed gene editing with five high-fidelity *Sp*Cas9 variants (Hypa, HF1, FeCas9, e*Sp*Cas9, and SuperFi)[27,28] in parallel with *Sp*Cas9 in mouse ESCs. Consistently, we identified mt-nuclear DNA fusions at the *c-Myc* target site in mouse ESCs after editing of high-fidelity variants (Fig. 1g and Supplementary Fig. 2c, d). Additionally, we identified some mt-nuclear DNA fusions at a dominant off-target site of *Sp*Cas9:*c-MYC_2* (Supplementary Fig. 2e)[29]. In contrast, by largely avoiding DSB generation, base editors decreased mt-nuclear DNA fusion products at target loci (Fig. 1h)[29,30]. In total, we have identified 1063 mt-nuclear DNA fusion events induced by various CRISPR-Cas editing tools in human cells. The mt-nuclear DNA fusion junctions were distributed throughout the mitochondrial genome (Supplementary Fig. 2f), and the D-loop region exhibited a higher junction density than other mtDNA regions (Supplementary Fig. 2g), consistent with previous reports that the D-loop is susceptible to DNA damage[6,31].

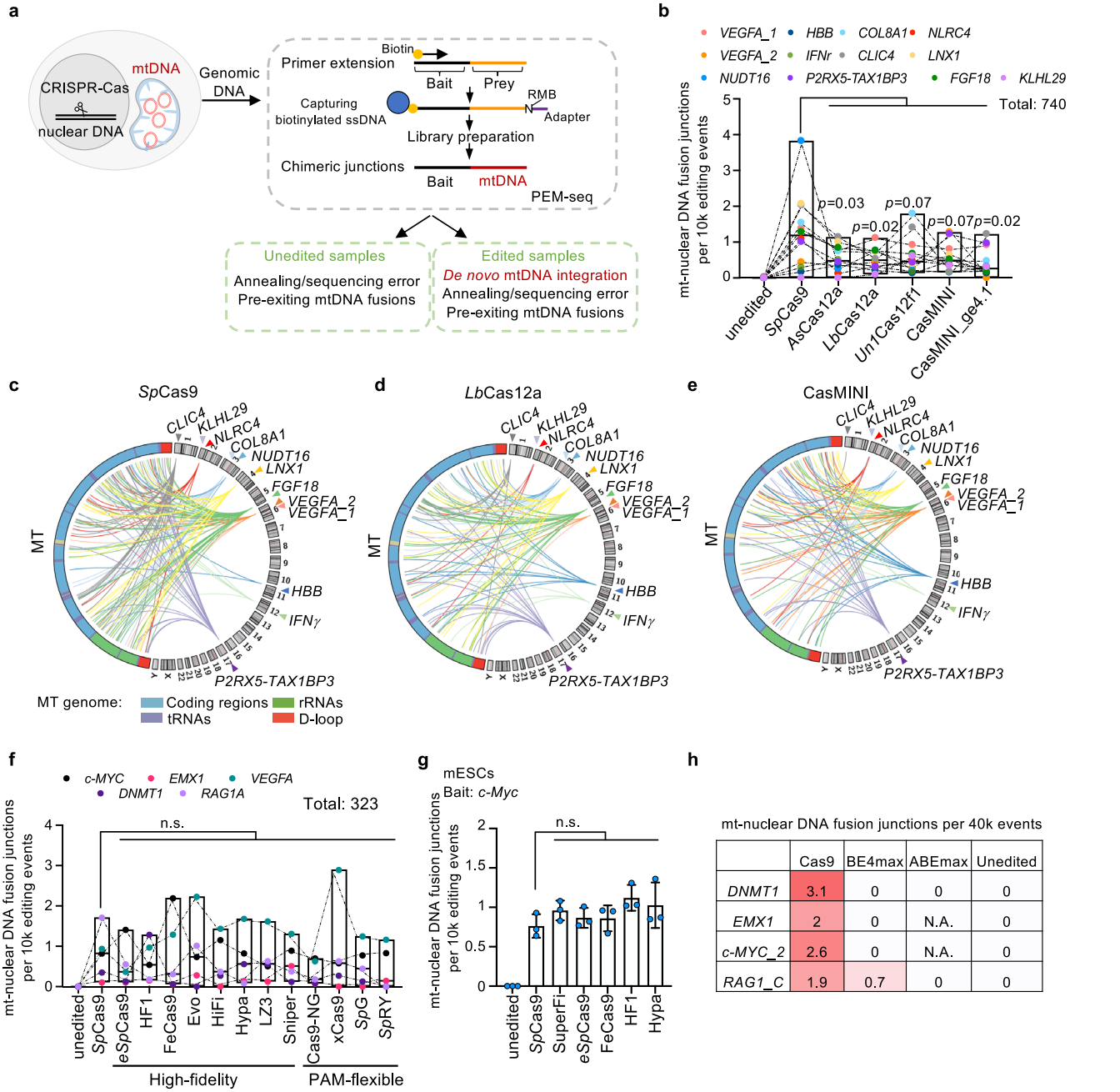
The mt-nuclear DNA fusions exhibit in various systems ex vivo and in vivo

The CRISPR-Cas9 system has been used to generate genetically engineered T cells for cancer immunotherapies (Fig. 2a)[32]. We analyzed mt-nuclear DNA fusion events using published PEM-seq data[29] of engineered human T cells (Supplementary Figs. 1a and 3a). Human T cells from cord blood were isolated, activated, and infected with lentivirus carrying a chimeric antigen receptor (CAR) and a CFP reporter. Following viral infection, 56.4% of T cells were CFP-positive[29]. Subsequently, the T cells were transfected with Cas9 ribonucleoprotein to disrupt the *TRAC*, *TRBC*, and *PDCD1* genes and were cultured for 3, 7, or 14 days (Supplementary Fig. 3a, b)[29]. Although the proportion of successfully edited cells decreased over time (Supplementary Fig. 3b), we still detected mt-nuclear DNA fusions in human CAR T cells after ex vivo culture for 3, 7, or 14 days, with a comparable frequency to those observed in HEK293T cells, regardless of the position of target sites (Fig. 2b and Supplementary Fig. 3c).

In another chronic inflammation murine model, naive T cells from engineered mice expressed an allogenic TCR, HH7-2tg, which specifically targets *Helicobacter hepaticus* (*H. hepaticus*)[33]. Following Cas9-mediated editing at the *c-Myc* locus, the activated TCR-T cells were infused into *H. hepaticus*-colonized *Rag1* deficient mice for 3 weeks (Fig. 2c and Supplementary Fig. 3d)[34]. We analyzed previously published PEM-seq data of both activated T cells and in vivo expanded T cells to detect mt-nuclear DNA fusions (Supplementary Fig. 1a). Editing efficiency of Cas9 exceeded 95% in both activated T cells and in vivo expanded T cells at the *c-Myc* locus[34]. Three different mtDNA-*c-Myc* fusions were identified in 167,611 editing events from activated T cells (Fig. 2d). Remarkably, one fusion junction was captured 38 times in the expanded T cells after 3 weeks of infusion into the recipient mouse (0.0132% of 286,818 editing events; Fig. 2d). The identified 38 reads showed diverse prey lengths and different sequences of random molecular barcodes (RMB; Fig. 2e, f), indicating that these reads were originated from different cells with the same fusion junction but not PCR duplication (Supplementary Fig. 3e), implying the amplification of mt-nuclear DNA fusions during T cell clonal expansion in vivo.

The genome-wide off-target analysis by two-cell embryo injection (GOTI) assay was developed to assess the nuclear base editors, by editing the embryos at the two-cell stage and collecting cells from murine E14.5 embryos for whole-genome sequencing[35]. We employed the NUMTs-detection algorithm to analyze the discordant reads of GOTI libraries with base editor 3 (BE3) targeting the *Tyr-D* and *Tyr-C* sites[6,35]. Collectively, we have identified 37 germline mtDNA

https://doi.org/10.1038/s41467-024-53806-0



h

mt-nuclear DNA fusion junctions per 40k events

	Cas9	BE4max	ABEmax	Unedited
DNMT1	3.1	0	0	0
EMX1	2	0	N.A.	0
c-MYC_2	2.6	0	N.A.	0
RAG1_C	1.9	0.7	0	0

Fig. 1 | Identification of mtDNA fusion to nuclear target sites of CRISPR-Cas systems. a Schematic of mt-nuclear DNA fusions captured by PEM-seq. The biotin-labeled primer located adjacent to the CRISPR-Cas9-target site (scissor) on the nuclear DNA is used to clone editing products (orange line). Then the single-stranded products were ligated with adapters (purple line) containing random molecular barcodes (RMB). and the chimeric reads harboring nuclear DNA around the editing site and mtDNA (red line), were identified as mt-nuclear DNA fusions. For each tested locus, PEM-seq was also conducted in unedited samples. **b** Box plot showing the frequency of mt-nuclear DNA fusions out of editing events at CRISPR-Cas target sites (colorful dots) under editing of CRISPR-Cas enzymes. Boundary of each box indicates the minimum and maximum. The middle line of each box indicates the median. Two-sided paired *t*-tests were conducted between *Sp*Cas9 and other CRISPR nucleases; $N = 12$. Source data are provided as a Source Data file. **c** Circos plot showing the mt-nuclear DNA fusion junctions on mtDNA (MT) and the indicated CRISPR-Cas9-target sites (colorful triangles) on the nuclear DNA of HEK293T cells. The outer circle shows the human genome, labeled with numbers or characters. The colorful lines indicate the fusion between the target site and mtDNA. Annotations of colored regions in mtDNA are shown at the bottom. **d** Circos plot showing the fusion junctions on mtDNA (MT) and the indicated CRISPR-*Lb*Cas12a target sites (colorful triangles) on the nuclear DNA of HEK293T cells. Legends are described in (**c**). **e** Circos plot showing the fusion junctions on mtDNA (MT) and the indicated CRISPR-CasMINI target sites (colorful triangles) on the nuclear DNA of HEK293T cells. Legends are described in (**c**). **f** Box plot showing the frequency of mt-nuclear DNA fusion events out of editing events at CRISPR-Cas target sites (colorful dots) under editing of *Sp*Cas9 variants. Boundary of each box indicates the minimum and maximum. The middle line of each box indicates the median. Two-sided paired *t*-test; n.s., not significant; $N = 5$. Source data are provided as a Source Data file. **g** Frequency of mt-nuclear DNA fusions caused by high fidelity *Sp*Cas9 variants in the mES cells. Mean ± SD; two-sided *t*-test; n.s. not significant; $n = 3$. Source data are provided as a Source Data file. **h** Average frequency of mt-nuclear DNA fusions at *DNMT1*, *EMX1*, *c-MYC_2*, and *RAG1_C* loci after editing by Cas9, BE4max, and ABEmax. *EMX1* and *c-MYC_2* loci were not targetable by ABEmax. N.A. not applicable. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-024-53806-0

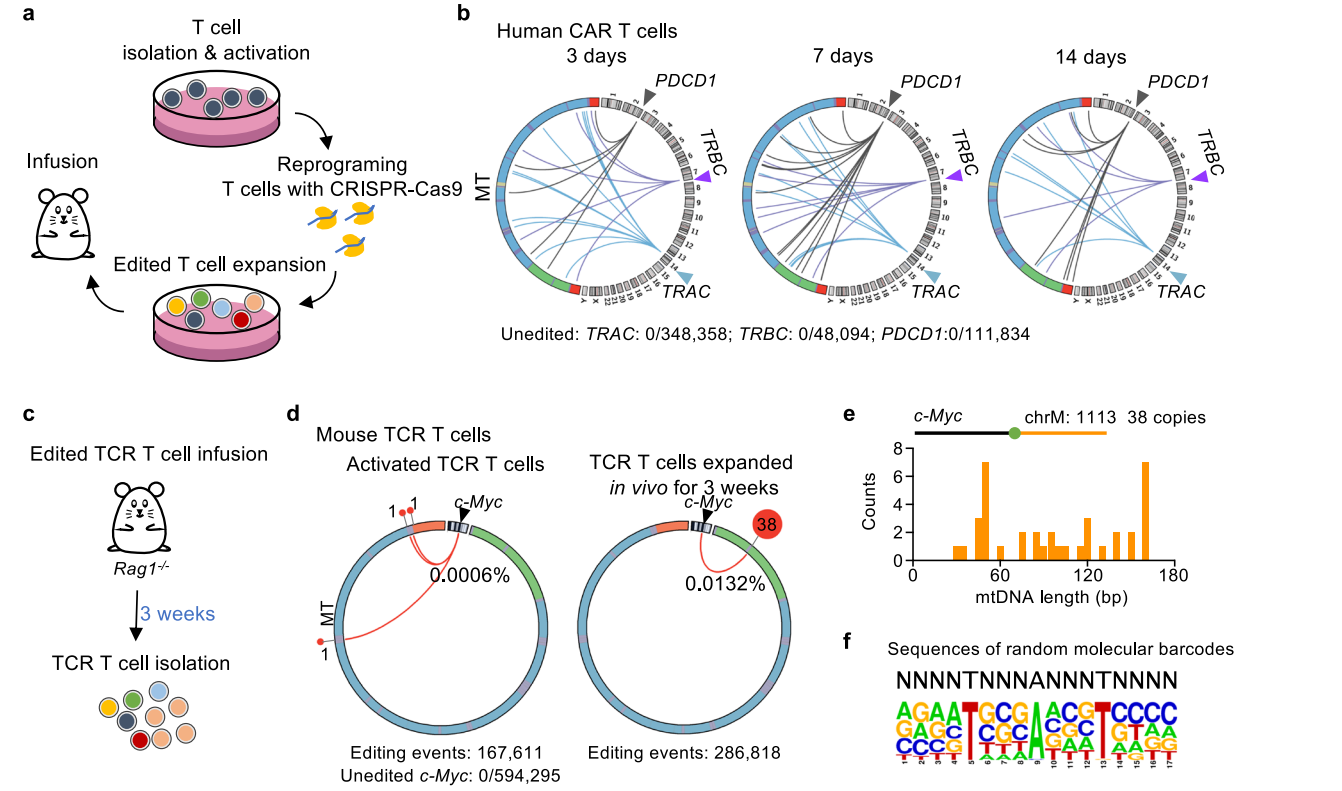


Fig. 2 | Mitochondrial DNA fuses to CRISPR-Cas9-target sites of the nuclear DNA ex vivo and in vivo. a Schematic of universal T cell manufactory. Primary T cells isolated from human or mice were activated and edited by CRISPR-Cas9. After edition, T cells underwent ex vivo culture or were infused into recipient mice. **b** Distribution of mtDNA-nuclear DNA fusions after CRISPR-Cas9 editing for 3, 7, and 14 days in human CAR T cells. Human primary T cells isolated from cord blood were activated for 3 days and subsequently transfected with Cas9/gRNA ribonucleoprotein (RNP) complexes targeting *TRAC*, *TRBC*, and *PDCD1* loci. Cells were collected after 3, 7, or 14 days and subjected to PEM-seq libraries. Legends are described in Fig. 1c. **c** Schematic showing the production of mouse TCR-T cells. TCR-T cells post-editing were infused to *Rag1*$^{-/-}$ recipient mice for 3 weeks, and subsequently isolated for PEM-seq analysis. **d** Distribution of mtDNA-nuclear DNA fusion junctions in CRISPR-Cas9 treated mouse TCR-T cells before and post-infusion for 3 weeks. Legends are described as depicted in (**b**). Percentages show the frequency of each mtDNA-nuclear DNA fusion out of Cas9-induced editing events. **e** Prey lengths of 38 mt-nuclear DNA fusions sharing the same junction from expanded mouse TCR-T cells indicated in (**d**). **f** Sequence logo showing the frequency of nucleotides in random molecular barcodes derived from the 38 reads of mt-nuclear DNA fusion at a single junction in expanded TCR-T cells indicated in (**d**). **g** Distribution of mtDNA integration in the nuclear DNA post base editor (BE3) treatment in mouse embryos. MT, mtDNA. **h** Number of mt-nuclear DNA fusion events identified in BE3-treated and -untreated samples. Two-sided *t*-test.

integrations shared by both edited and unedited embryos in the mouse stream (Fig. 2g). Besides, we detected 0 or two mitochondrial segments in the control embryos, while the BE3-edited embryos had six or eight new mitochondrial segments when targeting the *Tyr-D* or *Tyr-C* loci, respectively (Fig. 2g), indicating an increase in mtDNA integrations after base editing (Fig. 2h). Taken together, mtDNA integrations persisted during the development of mouse embryos.

Mitochondrial perturbance enhances mtDNA transfer into nuclear DNA

Integration of mitochondrial fragments into the nuclear target sites requires the joining of mtDNA and nuclear DNA breaks. To explore the contributions of mtDNA instability to mt-nuclear DNA fusions during gene editing, we induced mitochondrial stress that can generate mtDNA breakage and release fragmented mtDNA into the cytosol (Fig. 3a)[3,36]. We employed carbonyl cyanide m-chlorophenyl hydrazone (CCCP) to disrupt the membrane potential and used paraquat to induce oxidative damage[2,37,38]. Editing efficiency of Cas9 showed no increase after treatment with CCCP or paraquat (Supplementary Fig. 4a). The levels of mt-nuclear DNA fusions were found to be approximately two-fold higher after treatment with CCCP or paraquat, as determined by PEM-seq (Fig. 3b, c and Supplementary Fig. 4b).

To investigate the contributions of mtDNA breaks in mt-nuclear DNA fusions, we next employed mitoTALEN to directly generate DSBs in

https://doi.org/10.1038/s41467-024-53806-0

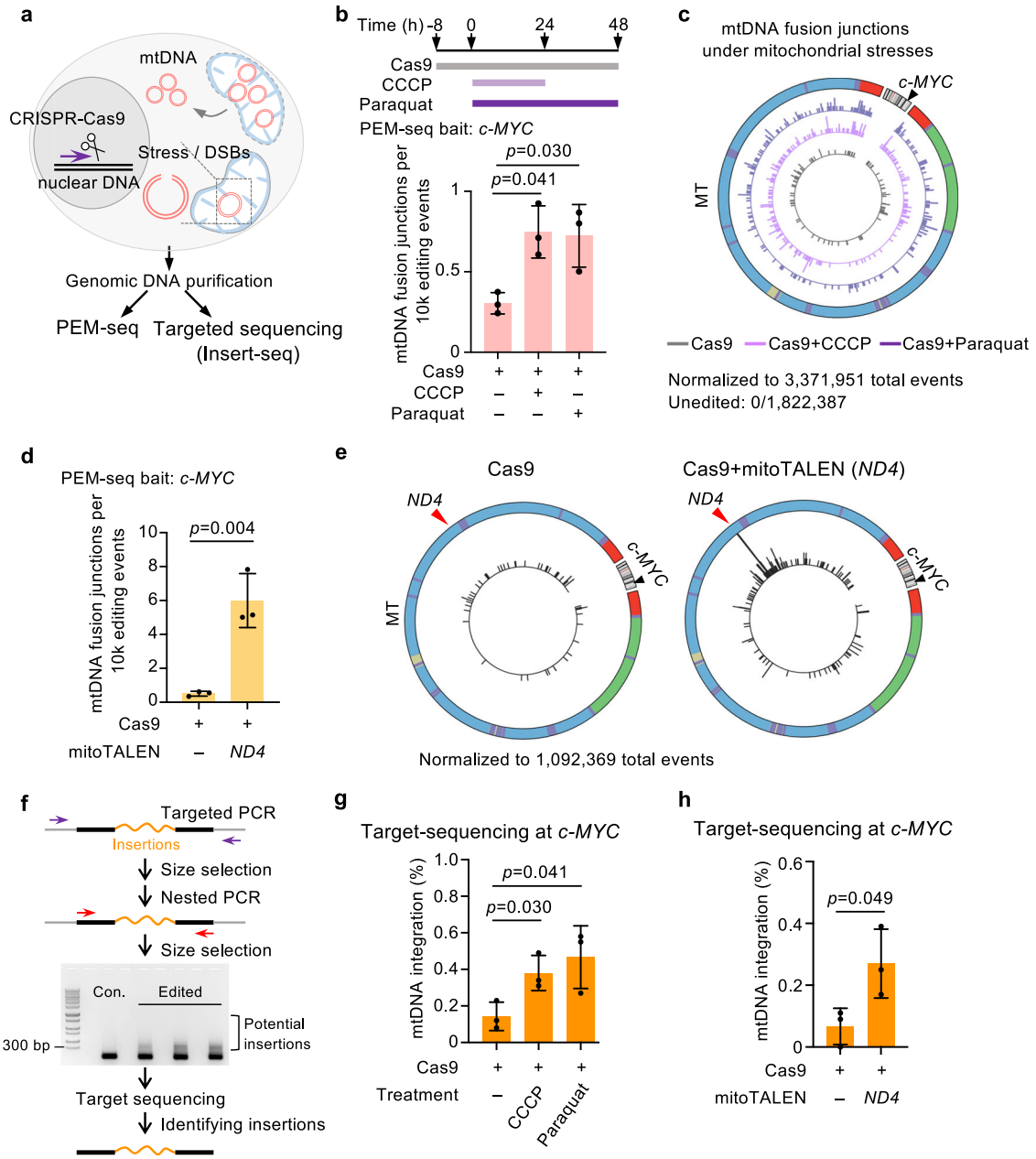


Fig. 3 | Mitochondrial stresses and DSBs exacerbate mtDNA integration into the nuclear genome. a Illustration of mitochondrial stresses and DSBs-induced mt-nuclear DNA fusions captured by PEM-seq and Insert-seq. CRISPR-Cas9 (scissor) targets the *c-MYC* locus in nuclear DNA, and the primer (purple arrow) for PEM-seq is adjacent to the target site. **b** Mitochondrial stresses inducing mt-nuclear DNA fusions captured by PEM-seq. Top: cells were transfected with a plasmid containing Cas9 and gRNA targeting *c-MYC* for 8 hours to allow the assembly of Cas9 and gRNA. Subsequentially, cells are treated with CCCP or paraquat for 24 or 48 hours , respectively, and then harvested for PEM-seq and Insert-seq analysis. Bottom: percentages of mt-nuclear DNA fusions captured by PEM-seq. Each dot represents a biological replicate. Mean ± SD; two-sided *t*-test; $n = 3$. Source data are provided as a Source Data file. **c** Distribution of mt-nuclear DNA fusion junctions captured by the *c-MYC* bait (black triangle) with or without mitochondrial stresses (untreated, gray bars; CCCP, light purple bars; paraquat, dark purple bars) treatment. Legends of mtDNA annotations are described as depicted in Fig.1c. The inner circles show the number of each mtDNA-nuclear DNA fusion point on mtDNA in a log scale. MT, mtDNA. **d** Frequency of PEM-seq-captured mt-nuclear DNA fusion junctions with or without mitoTALEN treatment. Each dot represents a biological replicate. Mean ± SD; two-sided *t*-test; $n = 3$. Source data are provided as a Source Data file. **e** Distribution of mt-nuclear DNA fusion junctions captured by the *c-MYC* bait (black triangle) with or without mitoTALEN (*ND4* site, red triangle) treatment. Legends of mtDNA annotations are described as depicted in Fig. 1b. **f** Workflow of Insert-seq to enrich insertions (orange lines) at the Cas9-editing site (*c-MYC* locus). Briefly, two rounds of targeted PCR (purple and red arrows) are used to clone the editing events around the target site, followed by two rounds of size selection that enriches insertions. Source data are provided as a Source Data file. **g** Percentages of mtDNA integrations within total insertions captured by Insert-seq at the *c-MYC* locus in the presence or absence of mitochondrial stresses. Each dot represents a biological replicate. Mean ± SD; two-sided *t*-test; $n = 3$. **h** Percentages of mtDNA integrations within total insertions captured by Insert-seq at the *c-MYC* site. Each dot represents a biological replicate. Mean ± SD; two-sided *t*-test; $n = 3$.

the mitochondrial *ND4* gene, and used PEM-seq to analyze mt-nuclear DNA fusions from CRISPR-Cas9-induced *c-MYC* bait on the nuclear DNA (Fig. 3a). We found that mitoTALEN treatment resulted in over 11-fold more mt-nuclear DNA fusions than unedited cells (0.059%, 647 of 1,092,369 vs. 0.005%, 60 of 1,203,042 total events; Fig. 3d and Supplementary Fig 4c, d). The mitochondrial fusion junctions were enriched at the mitoTALEN target sites (Fig. 3e), as validated by PCR and Sanger sequencing with two pairs of primers located adjacent to the editing sites of *c-MYC* and *ND4* (Supplementary Fig. 4e, f). We detected many junctions over the mitochondrial genome beyond the *ND4* target site (Fig. 3e), consistent with previous findings that DSBs lead mainly to rapid fragmentation and degradation of the mtDNA[22,23,39].

For further validation, we developed insertion-enriched target sequencing (Insert-seq), which employed two rounds of targeted PCR and two rounds of size selection to enrich fragment insertions at the nuclear editing site (Fig. 3f and Supplementary Fig. 4g). Insert-seq mainly captured DNA fragment insertions less than 500 bp with a median length of 163 bp (Supplementary Fig. 4g, h). We performed Insert-seq analysis on the same genomic DNA used for PEM-seq (Fig. 3a). The mtDNA integrations at the *c-MYC* target site accounted for approximately 0.38% and 0.47% in the presence of CCCP and paraquat, respectively, via Insert-seq. In contrast, the integration level was only 0.14% in the absence of mitochondrial stresses (Fig. 3g). With Insert-seq, we also identified over fourfold more mitochondrial insertions in the presence of mitoTALEN than in mitoTALEN-untreated cells (0.26%, 22 of 8531 vs. 0.06%, 7 of 10,098 total insertions; Fig. 3h and Supplementary Fig. 4i), in line with PEM-seq. Of note, one mtDNA insertion at the *c-MYC* locus captured by Insert-seq harbored the same junctions as one insertion identified by PEM-seq (Supplementary Fig. 4j). Collectively, these results suggest that fragmented mtDNA derived from mitochondrial stresses and DSBs can integrate into nuclear breaks.

MitoTALEN and DdCBE result in mt-nuclear DNA fusions

We next sought to investigate whether mitochondrial editing tools could induce mt-nuclear DNA fusions. We placed the bait primer in the mtDNA to capture the fusion events with nuclear prey sequences (Fig. 4a). To evaluate de novo mt-nuclear DNA fusions, we excluded the junctions with prey sequences falling in NUMTs or *ENCODE* blacklist regions (Fig. 4a)[40]. MitoTALEN generated substantial DSBs within the editing window between two TALE binding sites[13]. Therefore, we only considered PEM-seq chimeric reads harboring mtDNA baits within the editing window to identify bona fide mt-nuclear DNA fusions (Fig. 4b). We observed eightfold more mt-nuclear DNA fusions than in unedited cells when using mitoTALEN to edit the mitochondrial *ND5.1* gene (Fig. 4b, c). The identified nuclear fusion sites were widely distributed on the chromosomes, with about 49.5% (48/97) of these junctions located in gene regions. Similarly, fourfold more mt-nuclear DNA fusions were captured in cells post mitoTALEN editing at the *ND4* gene (Supplementary Fig. 5a, b).

We next investigated potential mt-nuclear DNA fusions during mitochondrial base editing by using DdCBE targeting *ND4* or *ND5.3* site[41] (Fig. 4d). At the *ND4* target site, the editing efficiency exceeded 40% (Fig. 4e). Occurrence of uracil through DdCBE editing might produce DNA breaks during DNA repair[17,30]. We only counted the chimeric junctions containing a G to A mutation or missing the G nucleotide (Fig. 4d, f and Supplementary Fig. 4c). We observed an 11-fold increase in mt-nuclear DNA fusions at the *ND4* locus compared to untreated cells after DdCBE treatment (Fig. 4f, g). Similarly, fivefold more mt-nuclear DNA fusions were observed when DdCBE targeted the *ND5.3* locus (Supplementary Fig. 5d–f). Besides, we also analyzed published GOTI data with DdCBE to induce the C12336T or G12918A mutation in mouse embryos[42]. Apart from germline integrations (Fig. 2g), we detected nine new mtDNA integrations in DdCBE-C12336T-edited embryos, with two mtDNA integrations detected in unedited embryos (Fig. 4h). Similarly, three new mitochondrial integrations were detected in DdCBE-G12918-edited embryos, and no mtDNA integration was detected in unedited embryos (Figs. 2g and 4h, i). These data suggested that DdCBE might accelerate the mtDNA integration to the nuclear genome.

TREX1 and TREX2 suppress the transfer of mtDNA into nuclear DNA

The transfer of mtDNA into the nuclear DNA necessitates mtDNA breakage and the release of fragmented mtDNA into cytoplasm. TREX1 and TREX2 are efficient at degrading free cytosol DNA and have been applied in nuclear gene editing[29]. We found a decrease of mt-nuclear DNA fusions at nuclear target sites by combining TREX2 with CRISPR-Cas9 in comparison to CRISPR-Cas9 (Fig. 5a)[29]. Consistently, Cas9-TREX2-3R (R163A, R165A, R167A mutant, referred to as Cas9TX), combining Cas9 and a TREX2 mutant with abolished DNA binding activity, also decreased mt-nuclear DNA fusions (Supplementary Fig. 6a)[29]. Thereby it is conceivable that ectopic expression of exonucleases might promote the degradation of damaged mtDNA and prevent mtDNA integration into the nuclear DNA. We generated TREX1n by removing the C-terminal cellular trafficking domain from TREX1, which did not affect its exonuclease activity[29]. We fused TREX1n or TREX2 with DdCBE directly or with a mitochondrial targeting sequence (MTS) to co-express with DdCBE (Fig. 5b). The fusion and co-expression of TREX1n/TREX2 had no significant impact on the editing efficiency of DdCBE or the mtDNA copy numbers (Fig. 5c and Supplementary Fig. 6b). We detected 126 mt-nuclear DNA fusions out of 25,732,881 total events after DdCBE editing at the *ND4* locus in HEK293T cells. In comparison, 46 to 90 mt-nuclear DNA fusions were identified when cells were treated with DdCBE either fused or co-expressed with TREX1n/TREX2 with the same sequencing depth (Fig. 5d), implying a potentially lower transfer rate of mtDNA into the nuclear DNA.

To capture mt-nuclear DNA fusions at a higher frequency, we introduced recurrent DSBs at *c-MYC* in the nuclear genome by CRISPR-Cas9 to capture mtDNA post-DdCBE editing via PEM-seq (Supplementary Fig. 6c). The *c-MYC* bait successfully captured 34 to 161 mt-nuclear DNA fusion junctions for each treatment after normalizing each sample to 3,231,948 total events (Fig. 5e and Supplementary Fig. 6d). DdCBE treatment at the *ND4* locus induced twofold more mtDNA integrations into the *c-MYC* Cas9-target sites than cells treated with only CRISPR-Cas9. However, fused expression of DdCBE with TREX1n/TREX2, or co-expression of DdCBE with TREX2 resulted in a significant decrease in mtDNA-nuclear DNA fusions, with levels dropping to that of cells treated with only CRISPR-Cas9 (Fig. 5e). Supportively, a nuclease-dead form of TREX2 (TREX2 H188A) significantly impaired the impact of TREX2 (Fig. 5e), suggesting that TREX1 or TREX2 might help reduce mtDNA integrations during gene editing.

Discussion

The NUMTs with fragmented or full-length mtDNA have been widely observed in somatic and cancer cells[5–7,43–47]. De novo transfer of mtDNA to the nuclear genome can lead to altered gene expression, such as the overactivation of *c-MYC* and *KCNMA1* oncogenes that contribute to tumor progression[7,48], or the disruption of genes that may contribute to the development of Pallister-Hall syndrome and mucolipidosis[49–51]. In addition, mtDNA integrations lead to mitochondrial heteroplasmy, and the level of mitochondrial heteroplasmy in individuals could change rapidly during development[52], similar to our findings that mtDNA integrations can be amplified during the clonal expansion of TCR-T cells (Fig. 2c–f).

Both the nuclear and mitochondrial genomes encounter spontaneous DNA lesions triggered by oxidative stress, ultraviolet (UV) light, chemicals, and replicative stress[53]. Furthermore, the mitochondrial genome undergoes fragmentation during stresses, resulting in the leakage of multiple mtDNA fragments into the cytosol and nucleus[2]. In this study, we utilized the gene editing-induced DSBs to show that any DSB in the nuclear genome might capture these mtDNA fragments, not

https://doi.org/10.1038/s41467-024-53806-0

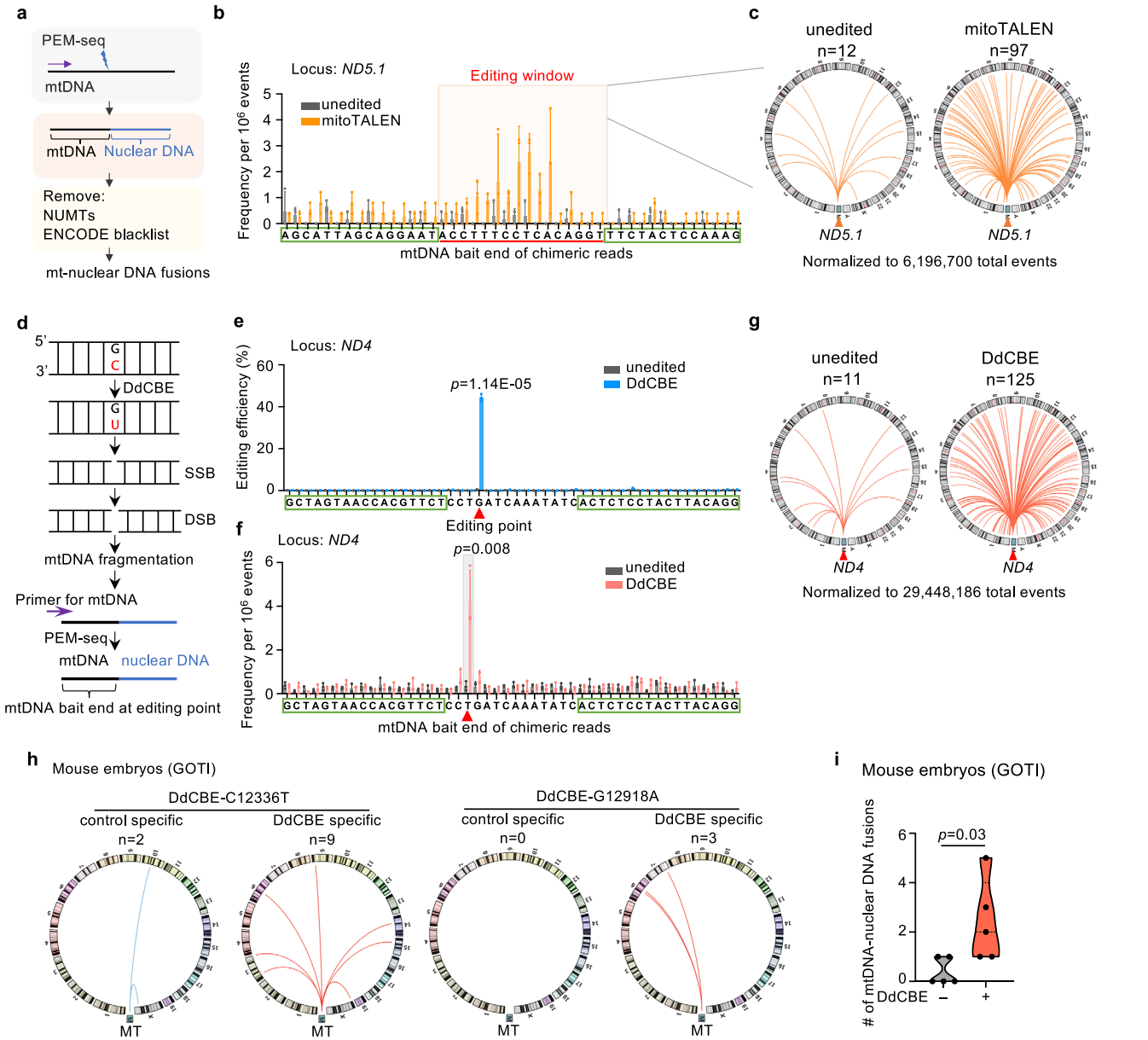


Fig. 4 | MitoTALEN and DdCBE result in mt-nuclear DNA fusions. a Schematic of mt-nuclear DNA fusion captured by PEM-seq cloning from mtDNA. The biotin-labeled primer located adjacent to the target site (spark) on the mtDNA is used to clone editing products, and the chimeric reads harboring mtDNA bait and nuclear DNA prey (blue line) were proceeded to subsequent process. Nuclear DNA prey junctions were filtered by removing those within NUMTs (details in Methods) and ENCODE repeat masker regions. **b** Frequency of mt-nuclear DNA fusions with mtDNA bait junctions at each nucleotide spanning the mitoTALEN target site on *ND5.1* (green boxes). Each dot represents a biological replicate. Chimeric reads containing bona fide mitoTALEN-induced mt-nuclear DNA fusion should harbor mtDNA bait junctions within the editing window (orange shadow). Mean ± SD; $n = 3$. Source data are provided as a Source Data file. **c** Distribution of mtDNA-nuclear DNA fusions with mtDNA bait junctions within the editing window (orange shadow in **b**) of mitoTALEN. **d** Illustration of DdCBE caused mt-nuclear DNA integration. Chimeric reads containing bona fide DdCBE-induced fusion should harbor mtDNA bait junctions at the editing site. **e** Editing efficiency (C to T or G to A) of DdCBE on *ND4*. Green boxes and the red triangle indicate binding sites and editing point of DdCBE. Each dot represents a biological replicate; Mean ± SD; *p*-value was obtained from two-sided *t*-test of editing efficiency at editing point (red triangle); $n = 3$. **f** Frequency of mt-nuclear DNA fusions with mtDNA bait junctions at each nucleotide spanning the DdCBE target sites (green boxes) on *ND4*. The mtDNA bait junctions induced by DdCBE editing was marked by gray shadow. Each dot represents a biological replicate. Mean ± SD; *p*-value was obtained from two-sided *t*-test of fusion frequency at 1-bp before editing point (red triangle); $n = 3$. **g** Distribution of mt-nuclear DNA fusions with mtDNA ending at the editing site (gray shadow in **f**) of DdCBE. **h** Distribution of mt-nuclear DNA fusion post mitochondrial base editor treatment in mouse embryos. Legends are described as depicted in Fig. 2g. **i** Distribution of mt-nuclear DNA fusions posts DdCBE treatment in mouse embryos. Data were re-analyzed from GOTI libraries. Two-sided *t*-test.

limited to DSBs in physiological cellular processes such as V(D)J recombination[54]. In this context, a 41-bp mtDNA fragment was found to be integrated at the junction of a reciprocal constitutional translocation[55]. Thus, NUMTs may arise in different cell types and may become inheritable if occurring in ES cells. Hundreds of NUMTs have been embedded in the human genome. Since most cancer cells are susceptible to DNA damage and possess a high oxidative metabolism, de novo NUMTs are frequent in cancer cells[6,8].

Previous studies have revealed that both nuclear CBEs and ABEs introduce small indels[30], as well as genome-wide translocations at target loci[29,56]. The CRISPR-free DdCBE system and other forms of mitochondrial base editors have demonstrated high efficiency in

https://doi.org/10.1038/s41467-024-53806-0

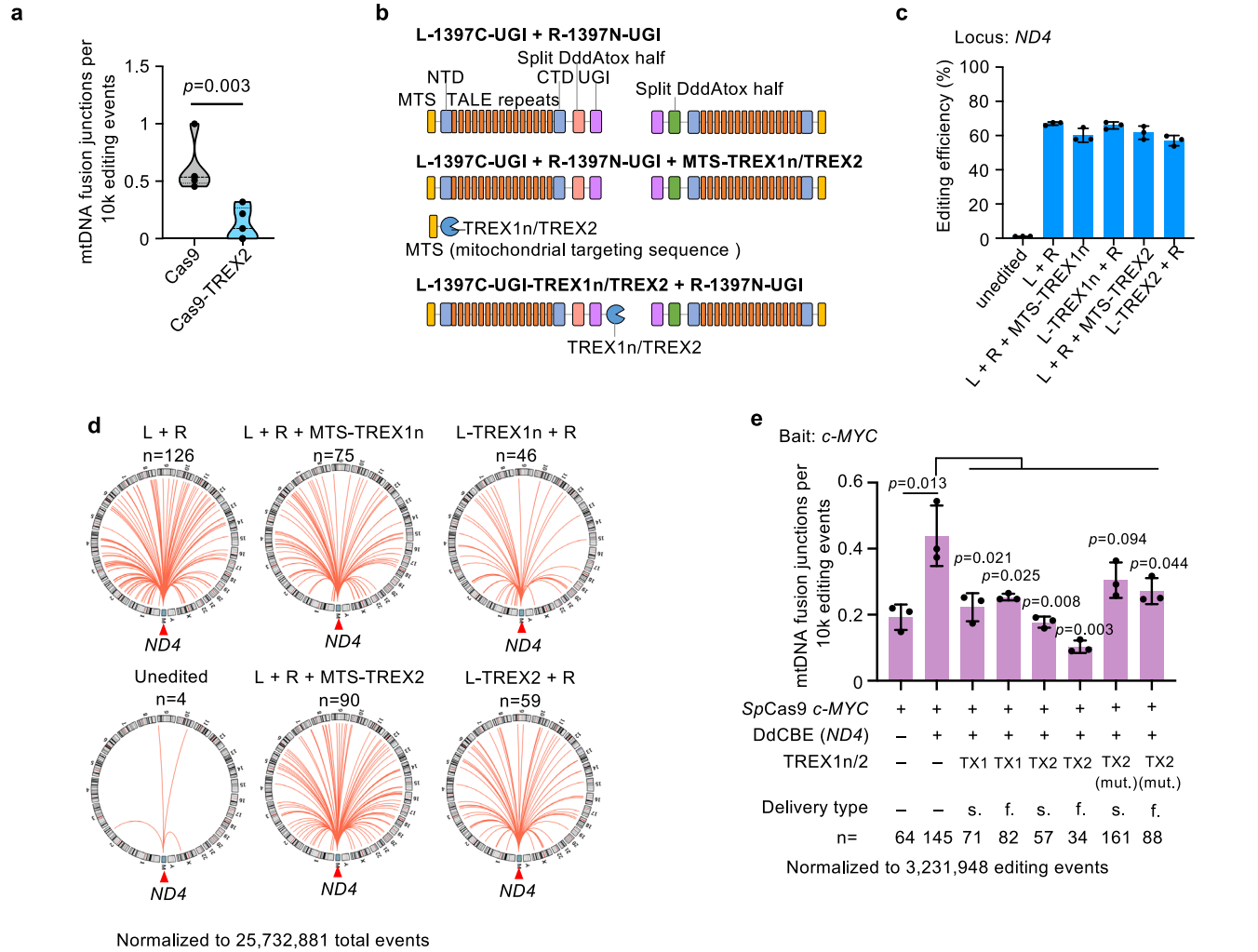


Fig. 5 | TREX1n and TREX2 suppress the transfer of mtDNA into nuclear DNA. **a** Frequency of mt-nuclear DNA fusions at *DNMT1*, *MYC1*, c-*MYC_2*, *MYC3*, *RAG1A loci* after editing with Cas9 or Cas9-TREX2. Two-sided *t*-test. **b** Structures of DdCBE with or without TREX1n/TREX2. For the fusion form, TREX1n or TREX2 was fused to the C-terminal domain of L-1397C-UGI. Regarding separated TREX1n or TREX2, both nucleases were tagged with mitochondrial targeting sequence (MTS) on the N-terminal. L-1397C-UGI, left TALE arrays fused to C-terminal DddAtox half and UGI; R-1397N-UGI, right TALE arrays fused to N-terminal DddAtox half and UGI; NTD N-terminal domain, CTD C-terminal domain. **c** Editing efficiency of DdCBE with or without TREX1n/TREX2 treatment. Mean ± SD. L L-1397C-UGI, R R-1397N-UGI; $n = 3$. **d** Distribution of mt-nuclear DNA fusions (red lines) with mtDNA bait junctions ending at the editing site of DdCBE on *ND4*. The number (n) of fusions in each sample is normalized to the same editing events. L L-1397C-UGI, R R-1397N-UGI. Source data are provided as a Source Data file. **e** Frequency of DdCBE-induced mtDNA fusing with the CRISPR-Cas9-target site with or without TREX1n/TREX2 treatment. Each dot represents a biological replicate. Mean ± SD; two-sided *t*-test; $n = 3$. L L-1397C-UGI, R R-1397N-UGI, TX1 TREX1n, TX2 TREX2, f. fused, s. separated, mut. nuclease-dead mutant. Source data are provided as a Source Data file.

modifying mtDNA and offer great potential for treating mitochondria-related diseases[17,18,57]. Our findings reveal a previously unknown risk that both nuclear and mitochondrial editing systems, such as DdCBE, can cause the transfer of mtDNA into the nuclear genome and nearly half of these mitochondrial integrations occur in gene regions. In the context of 10^8 ~ 10^9 edited CAR or TCR-T cells being infused into patients during therapy[58], one in 10^3 ~ 10^5 infused cells may carry mt-nuclear DNA fusions. Aberrant genome rearrangements such as translocations[34], large deletions[59], and chromosome loss[59,60] have also been documented during T cell manufacturing. Therefore, caution should be exercised when using both nuclear and mitochondrial genome editing tools for therapeutic purposes. In this regard, we have shown that the fusion or co-expression of DdCBE with TREX1/TREX2 can degrade mitochondrial fragments and prevent the fusion of mtDNA with DSBs in the nuclear genome. In support, TREX1 mutations found in autoimmune diseases were accompanied by increased amounts of escaped mtDNA fragments[61,62]. In addition, a recent study reported that chromosome loss can be mitigated by editing non-activated T cells expressing a higher level of p53 protein[59], indicating that genotoxicity such as mtDNA integrations might be reduced by conducting proper protocols during both nuclear and mitochondrial editing in clinic.

Methods

Ethics statement

All research performed in this study conforms to relevant ethical regulations of Peking University.

Cell culture

V6.5 mouse embryonic stem cells (a gift from Dr. Xiong Ji lab, School of Life Sciences, Peking University) were cultured in Knockout™ Dulbecco's Modified Eagle's Medium (DMEM, Gibco) containing 15% fetal bovine serum (Gibco), MEM non-essential amino acids solution (Sigma), nucleotides (Millipore), penicillin–streptomycin (Gibco), L-glutamine solution (Sigma), 2-mercaptoethanol (Sigma), LIF (Millipore), CHIR99021 (Selleck), and PD0325901 (Selleck) at 37 °C under 5% CO_2. HEK293T cells (a gift from Dr. Frederick Alt lab, Harvard Medical School) were cultured in DMEM (Gibco) supplemented with 10% fetal

bovine serum (ExCell Bio), penicillin–streptomycin (Gibco), and L-glutamine solution (Sigma) at 37 °C with 5% CO_2.

Cas9 variants in mouse embryonic cells

The plasmids expressing Cas9 variants were purchased replacing SpCas9 with different variants in the pX330 backbone, meanwhile, a puromycin resistance gene was inserted in the backbone for selection. One million cells were transfected with 2 μg plasmid containing gRNA targeting the *c-MYC* locus by nucleofector (Lonza, 4D-Nucleofector X). After one day post-transfection, the transfected cells were selected with 1 μg/mL puromycin for 36 hours, and then cultured with fresh medium for another 24 hours. Genomic DNA was extracted for PEM-seq analysis.

Mitochondrial stresses treatments

To evaluate the effects of mitochondrial stresses in Cas9 editing, 15 μg plasmid expressing *Sp*Cas9 targeting *c-MYC* locus was transfected into HEK293T cells by calcium phosphate. The transfected cells were then treated with 10 μM CCCP for 1 day and then cultured with fresh medium for another day, or treated with 500 μM paraquat for 2 days. After that, cells were harvested and the genomic DNA was extracted for PEM-seq analysis.

Mitochondrial editing

We obtained original DdCBE plasmids from Yi lab[41]. To obtain mitoTALEN system, we introduced ForkI to replace DddAtoX C-terminal domain and UGI in ND4-L1397C-LEFT plasmid (ND4-L1397C-LEFT-TALEN), and DddAtoX N-terminal domain and UGI in ND4-L1397C-RIGHT plasmid (ND4-L1397C-RIGHT-TALEN). HEK293T cells cultured in 10-cm dishes were transfected with pX330 plasmid targeting *c-MYC* locus (15 μg) by calcium phosphate, with or without additional mitoTALEN plasmid targeting the *ND4* locus (12 μg *ND4*-L1397C-LEFT-TALEN + 12 μg *ND4*-L1397C-RIGHT-TALEN).

To determine the occurrence of mtDNA-nuclear DNA fusion caused by mtDNA editors. HEK293T cells were transfected with mitoTALEN targeting *ND5.1* (20 μg *ND5.1*-L1397N-LEFT-TALEN + 20 μg *ND5.1*-L1397N-RIGHT-TALEN) in 10-cm dishes for 3 days. Regarding DdCBEs, HEK293T cells were also transfected with DdCBE plasmids targeting *ND4* locus (20 μg *ND4*-L1397C-LEFT + 20 μg *ND4*-L1397C-RIGHT) or *ND5.3* locus (20 μg *ND5.3*-L1397C-LEFT + 20 μg *ND5.3*-L1397C-RIGHT) for 3 days. After transfection, cells were harvested and permeabilized with NP-40 buffer (10 mM HEPES-NaOH, pH 7.6; 10 mM KCl; 0.1 mM EDTA; 0.3% NP-40) for 10 min, followed by centrifugation for nuclei collection and genomic DNA extraction.

For the TREX1/2 co-expression analysis, HEK293T cells were transfected with *Sp*Cas9 plasmid targeting the *c-MYC* locus (15 μg) and DdCBE-*ND4*-L1397C plasmids (20 μg + 20 μg) containing TREX1n/TREX2 at the C-terminus of the UGI of the left arm of DdCBE. Besides, DdCBE-*ND4*-L1397C plasmids and MTS-tagged TREX1n/TREX2 driven by a separate promoter were also introduced into HEK293T cells. All sample cells were collected at 3 days post-transfection.

PEM-seq library preparation and analysis

Reagents of PEM-seq method can be found in Supplementary Table 2. 50 μg of genomic or nuclear DNA from HEK293T cells, or 20 μg of genomic DNA from mES cells were used for each library. Sonicated DNA within 200–500-bp was applied to a single-round primer-extension step using a biotinylated primer placed within 200-bp from the cleavage site. Sequences of biotinylated primers for different target sites are listed in Supplementary Table 3. The products were conducted to size selection using 1.2× AxyPrep Mag PCR Clean-Up beads (Axygen, US) to remove excessive biotinylated primers. Biotinylated products were enriched by Dynabeads™ MyOne™ Streptavidin C1 (Thermo Fisher), and ligated with a bridge adapter (forward and reverse strand sequences listed in the Supplementary Table 3) carrying RMB. Ligated products were amplified through on-beads nested PCR with primers containing I5 and I7 sequences for 13–16 cycles. Then the PCR products were recovered by size selection Mag PCR Clean-Up beads (Axygen, US) and amplified with Illumina P5 and P7 sequences. All the primers used in this study were listed in Supplementary Table 3. A step-by-step PEM-seq protocol can be found in the previous report[24].

PEM-seq libraries were sequenced on Illumina Hiseq platforms, 2 × 150 bp. Primers used in this study were listed in Supplenmentary Table 3.

The PEM-seq data was analyzed using the PEM-Q pipeline [https://github.com/JiazhiHuLab] and aligned to the hg38 genome assembly using BWA[24]. To identify mtDNA fusions, we only kept chimeric reads mapped to "chrM" with a higher BWA mapping score than any nuclear genome regions. When analyzing PEM-seq data cloned from the mtDNA, we only kept chimeric reads mapped to the nuclear DNA and then removed those reads mapped to mtDNA-resembled nuclear contigs and ENCODE blacklist from [https://github.com/caleblareau/mitoblacklist/tree/master] and repetitive genomic regions annotated in the UCSC Database.

PEM-seq data from human CAR T and murine TCR-T cells

We analyzed and identified the mt-nuclear DNA fusion events of human CAR T cells after Cas9 editing from previously published data[29]. The human primary T cells (CD3+) isolated from human cord blood by negative selection kits (STEMCELL Technologies) were cultured with recombinant human interleukin-2 (IL-2) and activated with anti-CD3/28 beads for 3 days. pMD2.G, pxPAX2, and anti-CD19 scFV 4-1BB plasmids were co-transfected into HEK293T cells for CAR lentivirus preparation. After lentiviral infection, *Sp*Cas9 protein and sgRNA (synthesized with 2′-O-methyl and phosphorothioate modifications) were delivered into T cells through electroporation. Subsequently, the T cells were then collected after 3, 7, or 14 days of culture. Genomic DNA of these T cells was subjected to PEM-seq analysis.

We analyzed and identified the persistence of mt-nuclear DNA fusion in mouse TCR-T cells from previously published PEM-seq data[34]. Naïve TCR-T cells were isolated from HH7-2tg Cas9 mice[33] and activated with α-mouse CD3ε and α-mouse CD28 for 2 days. Activated T cells were infected with pST-sgRNA retrovirus expressing CD90.1 and rested with hIL-2, mouse IL-7, α-mouse IFNγ and α-mouse IL-4 for 3 days. CD90.1-positive T cells were sorted and injected into *H. hepaticus*-colonized *Rag1*$^{-/-}$ mice through the tail vein (300,000 T cells per mouse). After 3 weeks, the recipient mice were sacrificed and their intestinal tissues were collected by dissection. Inflammatory TCR-T cells were sorted with cell markers $CD3^+CD4^+CD45.1^+CD90.1^+$ and their genomic DNA was obtained for PEM-seq analysis.

Insertion-enriched target sequencing and analysis

Genomic DNA (gDNA) from cells with mitochondrial stresses or mitoTALEN treatment were subjected to both PEM-seq and insertion-enriched target sequencing analysis. Regarding insertion-enriched target sequencing, 2 μg gDNA was amplified by two primers (sequences were listed in Supplementary Table 3) flanking the CRISPR-Cas9-target site and Taq DNA polymerase in 100 μL PCR mixture and then performed as follows: 95 °C, 5 min; 95 °C, 30 s, 59 °C, 30 s, 72 °C, 4 min, 20 cycles; 72 °C, 5 min. DNA products were size selected by AMpure beads to keep insertions. After that, another two primers (sequences were listed in Supplementary Table 3) were used to further enrich DNA products on the Cas9-editing site by nested PCR, and the program was set as follows: 95 °C, 3 min; 95 °C, 30 s, 59 °C, 30 s, 72 °C, 3 min, 15 cycles; 72 °C, 5 min. A second size selection was performed to keep insertions by AMpure beads and the recovered DNA was tagged with Illumina adapter sequences. Library DNA was sequenced on Hi-seq platforms, 2 × 150 bp.

Raw reads were aligned to the human reference genome, assembly hg38, by BWA-MEM with the default parameters. The chimera reads, with MAPQ > 30 for each segment, were kept for insertion identification. Reads beginning at the primer start sites, covering at least 40 bp of both upstream and downstream of the Cas9-editing sites, and containing insertion fragments of less than 2000 bp were potential insertions after genome editing. Insertions with ≥4 counts were kept for further analysis.

Genome-wide off-target analysis by two-cell embryo injection (GOTI) data analysis

The GOTI datasets were aligned to the mm10 genome assembly by BWA (0.7.12) and samtools (0.1.19). The bam files were subjected to identification of discordant reads and breakpoints containing mitochondrial fragments using the NUMT detection algorithm from [https://github.com/WeiWei060512/NUMTs-detection][6]. The germline mitochondrial integration loci in the nuclear genome were obtained from shared breakpoints that showed up in both untreated and BE3-/DdCBE-treated embryos from SRP119022 and PRJNA786071 datasets in Sequence Read Archive (SRA). The breakpoints that only appeared in BE3-/DdCBE-treated were defined as BE3- or DdCBE-induced mt-nuclear DNA fusion sites.

Statistics and reproducibility

Statistics were performed using Student's *t*-test or paired *t*-test. Data were presented in mean ± SD for three or more biological repeats. The Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

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

Data availability

All sequencing data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO) database under the accession code GSE226884. PEM-seq datasets generated by previous studies are available in the following directories: Cas12 libraries[25] are from GEO under the accession number GSE213149; high-fidelity and PAM-less Cas9 variants libraries[27] from National Omics Data Encyclopedia (NODE) database with accession code OEP001824; comparison between Cas9 and base editors, *c-MYC_2* off-target data, and human T cell data[29] from NODE database with accession code OEP000911 [https://www.biosino.org/node/project/detail/OEP00000911]; mouse TCR-T cell data[34] from GEO under the accession number GSE202887; GOTI data of BE3[35] from SRA under project accession SRP119022; GOTI data of DdCBE[42] from SRA under project accession PRJNA786071. Source data are provided with this paper.

Code availability

The PEM-Q pipeline[24] is available at [https://github.com/JiazhiHuLab]. NUMT detection algorithm[6] for GOTI data is available at [https://github.com/WeiWei060512/NUMTs-detection].

References

1. Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
2. O'Malley, J., Kumar, R., Inigo, J., Yadava, N. & Chandra, D. Mitochondrial stress response and cancer. *Trends Cancer* **6**, 688–701 (2020).
3. Xian, H. et al. Oxidized DNA fragments exit mitochondria via mPTP- and VDAC-dependent channels to activate NLRP3 inflammasome and interferon signaling. *Immunity* **55**, 1370–1385.e8 (2022).
4. McArthur, K. et al. BAK/BAX macropores facilitate mitochondrial herniation and mtDNA efflux during apoptosis. *Science* **359**, 6378 (2018).
5. Wei, W. et al. Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans. *Nat. Commun.* **11**, 1740 (2020).
6. Wei, W. et al. Nuclear-embedded mitochondrial DNA sequences in 66,083 human genomes. *Nature* **611**, 105–114 (2022).
7. Ju, Y. S. et al. Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells. *Genome Res.* **25**, 814–824 (2015).
8. Kopinski, P. K., Singh, L. N., Zhang, S., Lott, M. T. & Wallace, D. C. Mitochondrial DNA variation and cancer. *Nat. Rev. Cancer* **21**, 431–445 (2021).
9. Jiang, F. & Doudna, J. A. CRISPR-Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
10. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
11. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
12. Liu, G., Lin, Q., Jin, S. & Gao, C. The CRISPR-Cas toolbox and gene editing technologies. *Mol. Cell* **82**, 333–347 (2021).
13. Silva-Pinheiro, P. & Minczuk, M. The potential of mitochondrial genome engineering. *Nat. Rev. Genet.* **23**, 199–214 (2022).
14. Barrera-Paez, J. D. & Moraes, C. T. Mitochondrial genome engineering coming-of-age. *Trends Genet.* **38**, 869–880 (2022).
15. Bacman, S. R., Williams, S. L., Pinto, M., Peralta, S. & Moraes, C. T. Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. *Nat. Med.* **19**, 1111–1113 (2013).
16. Bacman, S. R. et al. MitoTALEN reduces mutant mtDNA load and restores tRNAAla levels in a mouse model of heteroplasmic mtDNA mutation. *Nat. Med.* **24**, 1696–1700 (2018).
17. Mok, B. Y. et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631–637 (2020).
18. Cho, S.-I. et al. Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. *Cell* **185**, 1764–1776.e1712 (2022).
19. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
20. Min, J. et al. Mechanisms of insertions at a DNA double-strand break. *Mol. Cell* **83**, 2434–2448.e2437 (2023).
21. Yin, J. & Hu, J. The origin of unwanted editing byproducts in gene editing. *Acta Biochim. Biophys. Sin.* **54**, 767–781 (2022).
22. Peeva, V. et al. Linear mitochondrial DNA is rapidly degraded by components of the replication machinery. *Nat. Commun.* **9**, 1727 (2018).
23. Nissanka, N., Bacman, S. R., Plastini, M. J. & Moraes, C. T. The mitochondrial DNA polymerase gamma degrades linear DNA fragments precluding the formation of deletions. *Nat. Commun.* **9**, 2491 (2018).
24. Liu, Y. et al. PEM-seq comprehensively quantifies DNA repair outcomes during gene-editing and DSB repair. *STAR Protoc.* **3**, 101088 (2022).
25. Xin, C. et al. Comprehensive assessment of miniature CRISPR-Cas12f nucleases for gene disruption. *Nat. Commun.* **13**, 5623 (2022).
26. Schmid-Burgk, J. L. et al. Highly parallel profiling of Cas9 variant specificity. *Mol. Cell* **78**, 794–800.e798 (2020).
27. Zhang, W. et al. In-depth assessment of the PAM compatibility and editing activities of Cas9 variants. *Nucleic Acids Res* **49**, 8785–8795 (2021).
28. Bravo, J. P. K. et al. Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* **603**, 343–347 (2022).
29. Yin, J. et al. Cas9 exo-endonuclease eliminates chromosomal translocations during genome editing. *Nat. Commun.* **13**, 1204 (2022).

30. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
31. Mousavizadeh, K., Rajabi, P., Alaee, M., Dadgar, S. & Houshmand, M. Usage of mitochondrial D-loop variation to predict risk for Huntington disease. *Mitochondrial DNA* **26**, 579–582 (2015).
32. Yin, H., Xue, W. & Anderson, D. G. CRISPR–Cas: a tool for cancer research and therapeutics. *Nat. Rev. Clin. Oncol.* **16**, 281–295 (2019).
33. Xu, M. et al. c-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature* **554**, 373–377 (2018).
34. Wu, J. et al. CRISPR/Cas9-induced structural variations expand in T lymphocytes in vivo. *Nucleic Acids Res.* **50**, 11128–11137 (2022).
35. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
36. Kim, J. et al. VDAC oligomers form mitochondrial pores to release mtDNA fragments and promote lupus-like disease. *Science* **366**, 1531–1536 (2019).
37. Fessler, E. et al. A pathway coordinated by DELE1 relays mitochondrial stress to the cytosol. *Nature* **579**, 433–437 (2020).
38. Riley, J. S. et al. Mitochondrial inner membrane permeabilisation enables mtDNA release during apoptosis. *EMBO J.* **37**, e99238 (2018).
39. Moretton, A. et al. Selective mitochondrial DNA degradation following double-strand breaks. *PLoS ONE* **12**, e0176795 (2017).
40. Lareau, C. A. et al. Mitochondrial single-cell ATAC-seq for high-throughput multi-omic detection of mitochondrial genotypes and chromatin accessibility. *Nat. Protoc.* **18**, 1416–1440 (2023).
41. Lei, Z. et al. Mitochondrial base editor induces substantial nuclear off-target mutations. *Nature* **606**, 804–811 (2022).
42. Wei, Y. et al. Mitochondrial base editor DdCBE causes substantial DNA off-target editing in nuclear genome of embryos. *Cell Discov.* **8**, 27 (2022).
43. Dayama, G., Emery, S. B., Kidd, J. M. & Mills, R. E. The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res.* **42**, 12640–12649 (2014).
44. Mishmar, D., Ruiz-Pesini, E., Brandon, M. & Wallace, D. C. Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. *Hum. Mutat.* **23**, 125–133 (2004).
45. Woischnik, M. & Moraes, C. T. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* **12**, 885–893 (2002).
46. Hazkani-Covo, E., Zeller, R. M. & Martin, W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* **6**, e1000834 (2010).
47. Yuan, Y. et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet.* **52**, 342–352 (2020).
48. Shay, J. W., Baba, T., Zhan, Q. M., Kamimura, N. & Cuthbert, J. A. HeLaTG cells have mitochondrial DNA inserted into the c-myc oncogene. *Oncogene* **6**, 1869–1874 (1991).
49. Turner, C. et al. Human genetic disease caused by de novo mitochondrial-nuclear DNA transfer. *Hum. Genet.* **112**, 303–309 (2003).
50. Goldin, E. et al. Transfer of a mitochondrial DNA fragment to MCOLN1 causes an inherited case of mucolipidosis IV. *Hum. Mutat.* **24**, 460–465 (2004).
51. Borensztajn, K. et al. Characterization of two novel splice site mutations in human factor VII gene causing severe plasma factor VII deficiency and bleeding diathesis. *Br. J. Haematol.* **117**, 168–171 (2002).
52. Schwartz, M. & Vissing, J. Paternal inheritance of mitochondrial DNA. *N. Engl. J. Med.* **347**, 576–580 (2002).
53. Tubbs, A. & Nussenzweig, A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell* **168**, 644–656 (2017).
54. Lebedin, M. et al. Different classes of genomic inserts contribute to human antibody diversity. *Proc. Natl. Acad. Sci. USA* **119**, e2205470119 (2022).
55. Willett-Brozick, J. E., Savul, S. A., Richey, L. E. & Baysal, B. E. Germ line insertion of mtDNA at the breakpoint junction of a reciprocal constitutional translocation. *Hum. Genet.* **109**, 216–223 (2001).
56. Huang, M. E. et al. C-to-G editing generates double-strand breaks causing deletion, transversion and translocation. *Nat. Cell Biol.* **26**, 294–304 (2024).
57. Mok, B. Y. et al. CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat. Biotechnol.* **40**, 1378–1387 (2022).
58. Stadtmauer, E. A. et al. CRISPR-engineered T cells in patients with refractory cancer. *Science* **367**, eaba7365 (2020).
59. Tsuchida, C. A. et al. Mitigation of chromosome loss in clinical CRISPR-Cas9-engineered T cells. *Cell* **186**, 4567–4582.e20 (2023).
60. Nahmad, A. D. et al. Frequent aneuploidy in primary human T cells after CRISPR–Cas9 cleavage. *Nat. Biotechnol.* **40**, 1807–1813 (2022).
61. Hemphill, W. O. & Perrino, F. W. Measuring TREX1 and TREX2 exonuclease activities. *Methods Enzymol.* **625**, 109–133 (2019).
62. Skopelja-Gardner, S., An, J. & Elkon, K. B. Role of the cGAS–STING pathway in systemic and organ-specific diseases. *Nat. Rev. Nephrol.* **18**, 558–572 (2022).

Acknowledgements

We acknowledge all members of the Hu laboratory for their critical comments and helpful discussions. DdCBE plasmids were kind gifts from Dr. Chengqi Yi at Peking University. We also thank the Flow Cytometry Core at the National Center for Protein Sciences, Peking University, for the technical help. This work was supported by the National Key R&D Program of China (2022YFC3400201 to J.H. and 2022YFA1303000 to Y.T.), the NSFC grant (32122018 to J.H., 32225025 and 92254305 to Y.T.), and the Ministry of Agriculture and Rural Affairs of China. J.H. is an investigator at the PKU-TSU Center for Life Sciences.

Author contributions

J.H. conceived and supervised the project. J.W., Y.L., Y.T. and J.H. designed the experiments; J.W., Y.L., L.O., T.G., J.L., S.Y., C.X. and J.Y. performed the experiments; J.W., Y.L., Z.Z., X.L., M.L., Y.T. and J.H. analyzed the data; and J.W., Y.L., Y.T. and J.H. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53806-0.

Correspondence and requests for materials should be addressed to Jiazhi Hu.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

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

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024