nature protocols

https://doi.org/10.1038/s41596-023-00877-w

Protocol

Check for updates

Design and application of the transformer base editor in mammalian cells and mice

Wenyan Han[1,6], Bao-Qing Gao[2,3,6], Junjie Zhu[1,6], Zongxing He[1,6], Jianfeng Li[1] ✉, Li Yang[3] ✉ & Jia Chen[1,4,5] ✉

Abstract

Fusing apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like cytidine deaminase with catalytically impaired Cas proteins (e.g., nCas9 or dCas9) provides a novel gene-editing technology, base editing, that grants targeted base substitutions with high efficiency. However, genome-wide and transcriptome-wide off-target mutations are observed in base editing, which raises safety concerns regarding therapeutic applications. Previously, we developed a new base editing system, the transformer base editor (tBE), to induce efficient editing with no observable genome-wide or transcriptome-wide off-target mutations both in mammalian cells and in mice. Here we describe a detailed protocol for the design and application of the tBE. Steps for designing single-guide RNA (sgRNA) and helper sgRNA pairs, making constructs, determining the genome-wide and transcriptome-wide off-target mutations, producing the tBE-containing adeno-associated viruses, delivering adeno-associated viruses into mice and examining the in vivo editing effects are included in this protocol. High-precision base editing by the tBE can be completed within 2–3 weeks (in mammalian cells) or within 6–8 weeks (in mice), with sgRNA–helper sgRNA pairs. The whole process can be collaboratively accomplished by researchers using standard techniques from molecular biology, bioinformatics and mouse husbandry.

Key points

- This protocol describes the transformer base editor system to induce efficient editing with no observable genome-wide or transcriptome-wide off-target mutations, both in mammalian cells and in mice.
- The transformer base editor system overcomes the problems of guide RNA-independent and guide RNA-dependent off-target editing as it remains inactive at off-target sites but can be transformed to be active for base editing after binding at the on-target site.

Key references

Wang, L. et al. *Nat. Cell Biol.* **23**, 552-563 (2021): https://doi.org/10.1038/s41556-021-00671-4

[1]Gene Editing Center, School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [2]Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [3]Center for Molecular Medicine, Children's Hospital, Fudan University and Shanghai Key Laboratory of Medical Epigenetics, International Laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, Institutes of Biomedical Sciences, Fudan University, Shanghai, China. [4]Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China. [5]Shanghai Clinical Research and Trial Center, Shanghai, China. [6]These authors contributed equally: Wenyan Han, Bao-Qing Gao, Junjie Zhu, Zongxing He. ✉e-mail: lijf1@shanghaitech.edu.cn; liyang_fudan@fudan.edu.cn; chenjia@shanghaitech.edu.cn

Introduction

The fusion of cytidine deaminase and catalytically impaired Cas protein led to the development of base editors (BEs)[1,2], which induced efficient C-to-T base editing at target sites without DNA double-strand breaks[3–8]. Although BEs have great potential in biological research and disease treatment, recent studies have found that original BEs (with a canonical fusion protein configuration containing a Cas protein and an active cytidine deaminase) can induce genome-wide and transcriptome-wide off-target (OT) mutations[9–12]. Notably, the majority of these OT mutation sites are guide RNA (gRNA) independent as the cytidine deaminase family members in canonical BEs can directly bind single-stranded (ss) nucleic acids to catalyze the deamination of cytosine[13–15], without the requirement of the clustered regularly interspaced short palindromic repeats (CRISPR)–Cas–gRNA complex[16]. Meanwhile, since BEs consist of the CRISPR–Cas system, they exhibit gRNA-dependent OT mutations caused by the binding of gRNA at OT sites that have sequence similarity to on-target sites[17,18]. These gRNA-independent and gRNA-dependent OT mutations thus hinder the application of BEs in both biological and biomedical studies, especially for treating human diseases, e.g., knocking out disease-related genes[19–22], modulating gene expression[23,24] and correcting pathogenic mutations[25,26].

To comprehensively address the problems of gRNA-independent and gRNA-dependent OT editing, we recently developed a transformer base editor (tBE) system, which remains inactive at OT sites but can be transformed to be active for base editing after binding at the on-target site[27]. The tBE can induce efficient C-to-T base editing at a broad range of target sites in mammalian cells (including at GpC dinucleotides and in methylated DNA regions) with no observable gRNA-dependent DNA OT mutations, gRNA-independent DNA OT mutations or gRNA-independent RNA OT mutations. Furthermore, the tBE is composed of a few small components, which can be packaged into a dual-adeno-associated virus (AAV) system without intein splicing. As a result, the tBE can be efficiently delivered into target organs or tissues via AAVs and induce potent base editing in vivo. Here, we describe how to design gRNAs for the tBE, how to determine OT mutations and how to produce tBE-containing AAVs for in vivo base editing.

Development of the tBE

In contrast to the large fusion protein used in canonical BEs, the tBE system is composed of four protein components that together achieve efficient C-to-T base editing with no observable OT mutations[27]. These components are as follows: (1) a *Streptococcus pyogenes* Cas9 (SpCas9) nickase D10A (nSpCas9), (2) the N-terminal fragment of tobacco etch virus (TEV) protease (TEVn), (3) a fusion protein of an RNA aptamer binding protein (N22p) and the C-terminal fragment of TEV protease (TEVc) and (4) an effector fusion protein called tBE–V5–mA3 that contains a cytidine deaminase (mouse apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like (APOBEC)3 cytidine deaminase domain 1, mA3CDA1), a deoxycytidine deaminase inhibitor (dCDI) from mouse APOBEC3 (mA3dCDI), a uracil DNA glycosylase inhibitor (UGI) and another RNA aptamer binding protein (MCP) (Fig. 1a). In addition, a boxB aptamer-containing single-guide RNA (sgRNA–boxB) and an MS2 aptamer-containing helper sgRNA (hsgRNA–MS2) are both used in the tBE system, in contrast to the canonical base editing system, which uses only one regular sgRNA (Fig. 1a).

When the effector protein tBE–V5–mA3 is expressed in the cytoplasm of target cells, its cytidine deaminase is inactive because of the linkage of a dCDI with a TEV protease cleaving site (Fig. 1b, cytoplasm), which ensures that no gRNA-dependent or gRNA-independent OT mutations are triggered. After the nSpCas9–hsgRNA and nSpCas9–sgRNA complexes bind to the appropriate genomic positions, the MS2-containing hsgRNA recruits the effector protein tBE–V5–mA3 and the boxB-containing sgRNA recruits the TEVc (Fig. 1b, colocalization). A free TEVn and the sgRNA-recruited TEVc combine to form an intact TEV protease, which cleaves off the dCDI that is fused to cytidine deaminase via the TEV protease cleaving site (Fig. 1b, activation). As a result, cytidine deaminase is activated at the on-target site and able to induce efficient base editing (Fig. 1b, base editing). As the split version of the TEV protease is utilized in the tBE system, the tri-molecular reaction that results in the cleavage of dCDI and the activation

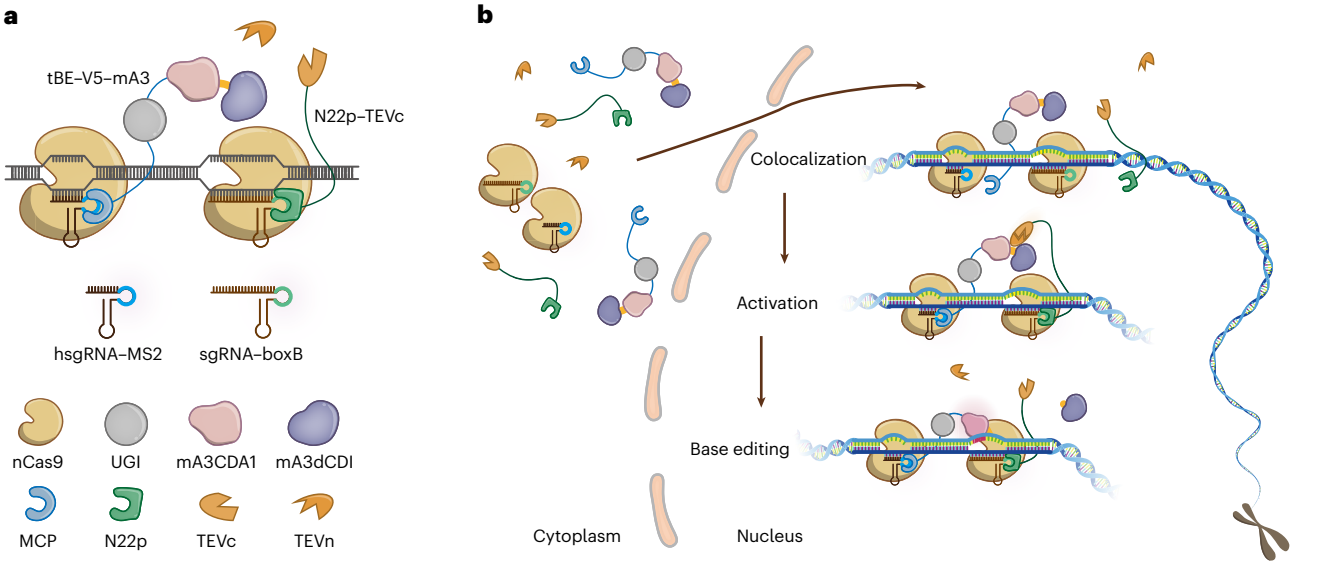


Fig. 1 | Schematic of the tBE mechanism. a, The components of the tBE system. The orange linker between mouse APOBEC3 (mA3) cytidine deaminase domain 1 (mA3CDA1) and deoxycytidine deaminase inhibitor (dCDI) from mA3 (mA3dCDI) is the TEV protease cleaving site. tBE–V5–mA3, the effector fusion protein contains an mA3dCDI, an mA3CDA1, a UGI and a bacteriophage MS2 coat protein (MCP). N22p–TEVc, a fusion protein of bacteriophage lambda N peptide 1–22 (N22p) and the C-terminal fragment of TEV protease (TEVc). V5, version 5. TEVn, the N-terminal fragment of TEV protease. nCas9, Cas9 nickase D10A. **b**, The mechanism of the tBE. After the colocalization of paired hsgRNA–sgRNA on genomic DNA, the assembled TEV protease cleaves the mA3dCDI off and mA3CDA1 is activated, which induces efficient base editing at the on-target site.

of cytidine deaminase occurs only at on-target sites (Fig. 1b, activation). At any OT site (gRNA dependent or gRNA independent), the random collision of any two molecules of TEVn, TEVc and the TEV protease cleaving site does not trigger cleavage, which prevents the generation of OT mutations. Moreover, only two active cytidine deaminases are theoretically released from the two alleles of a target site following the successful on-target editing by tBE as SpCas9 dissociates slowly from the on-target site[28,29].

The simplicity of packaging the tBE into a dual-AAV system for in vivo editing is another advantage in employing it to treat disease. Canonical BEs are too large to be packaged into AAVs, which can carry cargos up to ~4.5 kb in length[30], much smaller than the size of a canonical BE with the fusion protein configuration[1,2,18,31–37]. Although the protein splicing system can be used to split BEs into two halves[38], the religation of two protein fragments mediated by intein can lead to lower editing efficiency. The tBE, in contrast, can be naturally packed into a dual-AAV system (Fig. 2b) due to its small components, allowing for efficient delivery and base editing in vivo without the need for protein splicing. However, the tBE system requires the use of six components (four protein components and two gRNA components), compared with conventional BEs that comprise only two components (the BE protein and the gRNA). Thus, the design of tBE is more complicated than conventional BEs and the total size of tBE constructs is larger too.

Applications of the tBE

As the tBE induces efficient C-to-T base editing without observable gRNA-dependent or gRNA-independent OT mutation, it can be used to disrupt regulatory *cis*-elements to modulate gene expression and provide therapeutic benefits, e.g., triggering the expression of fetal hemoglobin to treat beta-hemoglobinopathies[23,24]. Additionally, the editing efficiency of the tBE is not compromised for hard-to-edit dinucleotides (e.g., GpC) or in methylated DNA regions; therefore, it can also be conveniently used to generate premature stop codons to knockout genes for therapeutic purposes, e.g., knocking out *PCSK9* to treat familial hypercholesterolemia[27,39].

Furthermore, as the tBE is composed of naturally small components and can be efficiently packaged into a dual-AAV system, it is suitable for in vivo editing in the organs or tissues that

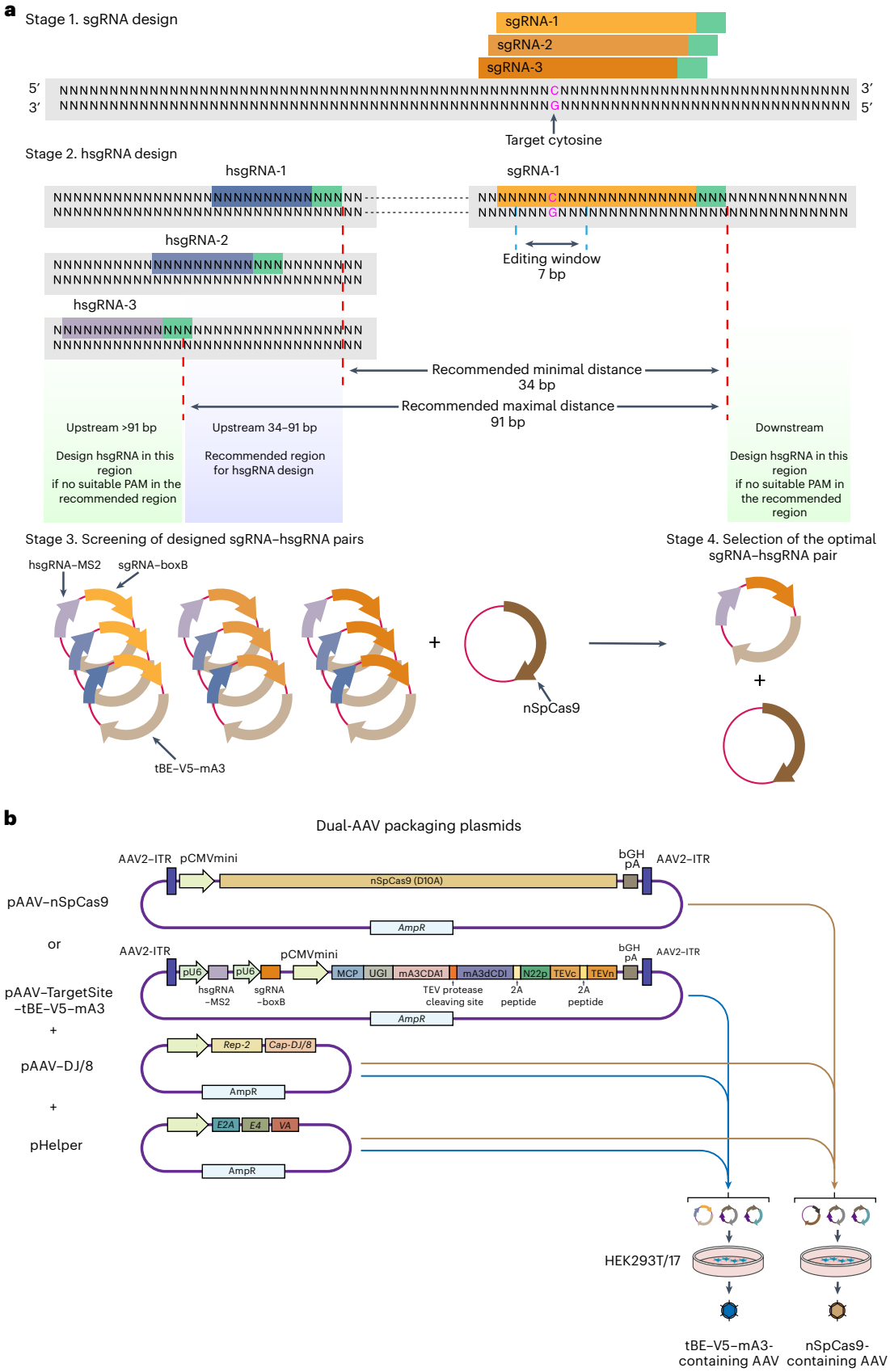

a
Stage 1. sgRNA design
sgRNA-1
sgRNA-2
sgRNA-3
5′
3′
Target cytosine
Stage 2. hsgRNA design
hsgRNA-1
sgRNA-1
hsgRNA-2
hsgRNA-3
Editing window
7 bp
Recommended minimal distance
34 bp
Recommended maximal distance
91 bp
Upstream >91 bp
Design hsgRNA in this region
if no suitable PAM in the recommended region
Upstream 34–91 bp
Recommended region for hsgRNA design
Downstream
Design hsgRNA in this region
if no suitable PAM in the recommended region
Stage 3. Screening of designed sgRNA–hsgRNA pairs
hsgRNA–MS2
sgRNA–boxB
tBE–V5–mA3
nSpCas9
Stage 4. Selection of the optimal sgRNA–hsgRNA pair
b
Dual-AAV packaging plasmids
pAAV–nSpCas9
AAV2–ITR
pCMVmini
nSpCas9 (D10A)
bGH pA
AmpR
or
pAAV–TargetSite –tBE–V5–mA3
pU6
hsgRNA –MS2
sgRNA –boxB
MCP
UGI
mA3CDA1
mA3dCDI
N22p
TEVc
TEVn
TEV protease cleaving site
2A peptide
pAAV–DJ/8
Rep-2
Cap-DJ/8
pHelper
E2A
E4
VA
HEK293T/17
tBE–V5–mA3-containing AAV
nSpCas9-containing AAV

Fig. 2 | Design of sgRNA–hsgRNA pairs for in vivo editing. a, The design strategy for sgRNA–hsgRNA pairs. When multiple sgRNA–hsgRNA pairs can be designed for a target cytosine, we suggest testing all the pairs and selecting the pair with the highest editing efficiency for further in vivo editing. Stage 1, according to the location of the target cytosine, search for proper PAMs around to ensure the target cytosine located in the editing window of tBE. Stage 2, design all suitable sgRNA–boxB following the rule shown in stage 1. Stage 3, screen all pairs of hsgRNA–MS2 and sgRNA–boxB designed. Stage 4, select the pair of hsgRNA–MS2–sgRNA–boxB of tBE with the highest editing efficiency at the target cytosine. NNN highlighted in green, PAM for sgRNA and hsgRNA. **b**, Dual-AAV packaging plasmids for the tBE. nSpCas9 is inserted between two AAV2-ITRs to construct pAAV–nSpCas9. The selected sgRNA–hsgRNA pair, the effector protein tBE–V5–mA3, the N22p–TEVc protein and TEVc protein are inserted between another two AAV2-ITRs to construct pAAV–TargetSite–tBE–V5–mA3. pAAV–nSpCas9 and pAAV–TargetSite–tBE–V5–mA3 are each separately cotransfected with pAAV–DJ/8 and pHelper into packaging cells to produce the two types of AAV for the tBE. pAAV–DJ/8 contains a *Rep-2* gene, which is involved in AAV genome replication, and a *Cap-DJ/8* gene, which constructs AAV protein shell and determines AAV serotype. pHelper contains an *E2A* gene, which is involved AAV genome replication, an *E4* gene, which is involved in the transcription of the AAV genome, and a *VA* gene, which prevents the activation of the antiviral system in AAV-packaging cells. *AmpR*, ampicillin resistance gene; ITR, the inverted terminal repeat; pCMV, cytomegalovirus promoter; pU6, *U6* promoter; bGH pA, bovine growth hormone polyadenylation signal; 2A peptide, self-cleaving peptide.

AAVs of the appropriate serotype can target, e.g., liver, brain, retina or muscle[30]. Additionally, ex vivo editing in hematopoietic stem cells[24,40] or T cells[22] could be induced by the tBE because electroporation can efficiently transfer multiple RNA molecules into a target cell, e.g., an sgRNA, an hsgRNA, an mRNA encoding nSpCas9 and an mRNA encoding the other protein components linked by 2A self-cleaving peptides.

We have used this protocol to induce efficient base editing in the livers of C57BL/6 mice (female, 6–8 weeks old)[27]. As a mismatch repair pathway, which is triggered by nCas9 to stall C-to-T editing, is constitutively active in most of the cells and tissues[41] and other BEs have been used in various animals successfully[21,42–45], the tBE procedure described in this protocol could be able to induce efficient editing in other types, ages and sexes of mouse. We have also used this protocol to induce efficient base editing in 293FT, U2OS and N2A cells[27], and the same procedure can be used in other easy-to-transfect cell lines.

Here, we present a detailed protocol including the design and screening of sgRNA–hsgRNA pairs, determination of OT mutations, packaging of tBE-containing AAVs and examination of in vivo base editing effects (Fig. 3).

Comparison with other methods

Several BEs have been developed by introducing various amino acid substitutions into the substrate-binding domain or the catalytic domain of cytidine deaminases to reduce their binding or catalytic activities on ssDNA or RNA, which reduced the gRNA-independent OT mutations[46–48]. However, these amino acid substitutions inevitably result in reduced on-target editing efficiencies and/or shrunken editing windows, which are not desirable when BEs are used to generate stop codons[19,20] or disrupt regulatory *cis*-elements[23,24]. Similarly, while gRNA-dependent OT mutations can be reduced by using the Cas variants with high DNA targeting fidelity in BEs[17], the on-target editing efficiency is usually compromised[49–51], as most Cas variants with high DNA targeting fidelity have lower DNA-binding activities than the wild-type Cas proteins at on-target sites.

Limitations

As an hsgRNA located in an appropriate region (i.e., flanking the on-target site) is required in the tBE system, the protospacer adjacent motif (PAM) restriction of hsgRNA is a limitation. Nevertheless, we have shown that PAM-flexible Cas9 variants (e.g., Cas9–NG[52]) can be used in the tBE to induce efficient editing[27]; thus, further engineering of Cas9 (ref. 53) to realize PAM-less targeting will greatly facilitate the selection of appropriate sgRNA–hsgRNA pairs for most genomic loci. By contrast, when multiple PAMs in an appropriate region can serve as hsgRNA, we suggest testing all possible hsgRNAs (usually two to six) in a common cell line with high transfection efficiency (such as 293FT) to select the best hsgRNA, which can have relatively high time and labor demands.

In addition, as the deaminase used in the tBE (i.e., mA3CDA1) catalyzes robust cytidine deamination, the editing window of the tBE is relatively large (~7 bp), which may not be suitable for precision editing to correct specific genetic mutations. Thus, when a single C-to-T edit is required for disease treatment, BEs with high editing efficiency and narrow editing

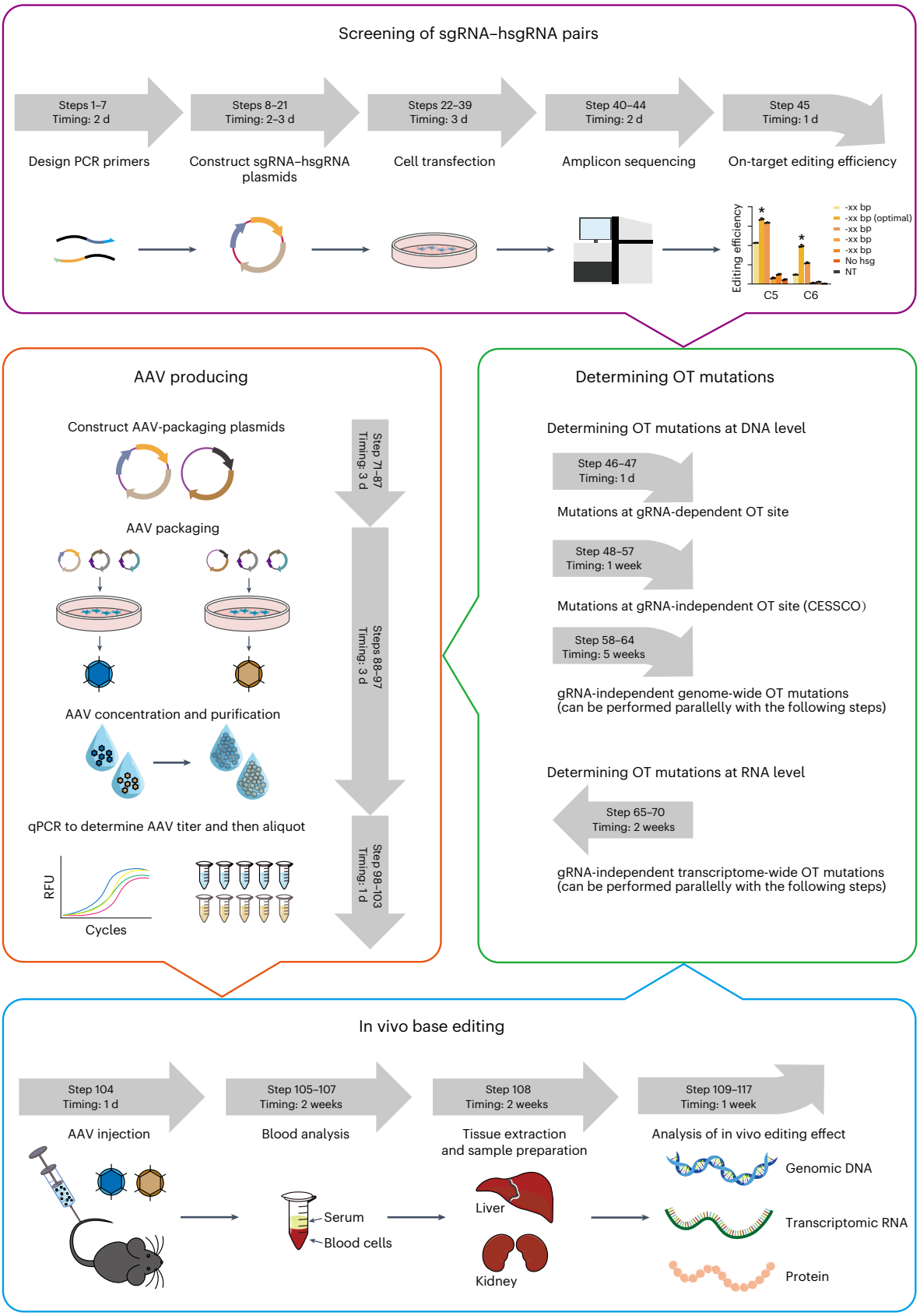

Screening of sgRNA–hsgRNA pairs
Steps 1–7
Timing: 2 d
Design PCR primers
Steps 8–21
Timing: 2–3 d
Construct sgRNA–hsgRNA plasmids
Steps 22–39
Timing: 3 d
Cell transfection
Step 40–44
Timing: 2 d
Amplicon sequencing
Step 45
Timing: 1 d
On-target editing efficiency
Editing efficiency
-xx bp
-xx bp (optimal)
-xx bp
-xx bp
-xx bp
No hsg
NT
C5
C6
AAV producing
Construct AAV-packaging plasmids
AAV packaging
AAV concentration and purification
qPCR to determine AAV titer and then aliquot
RFU
Cycles
Step 71–87
Timing: 3 d
Steps 88–97
Timing: 3 d
Step 98–103
Timing: 1 d
Determining OT mutations
Determining OT mutations at DNA level
Step 46–47
Timing: 1 d
Mutations at gRNA-dependent OT site
Step 48–57
Timing: 1 week
Mutations at gRNA-independent OT site (CESSCO)
Step 58–64
Timing: 5 weeks
gRNA-independent genome-wide OT mutations
(can be performed parallelly with the following steps)
Determining OT mutations at RNA level
Step 65–70
Timing: 2 weeks
gRNA-independent transcriptome-wide OT mutations
(can be performed parallelly with the following steps)
In vivo base editing
Step 104
Timing: 1 d
AAV injection
Step 105–107
Timing: 2 weeks
Blood analysis
Serum
Blood cells
Step 108
Timing: 2 weeks
Tissue extraction
and sample preparation
Liver
Kidney
Step 109–117
Timing: 1 week
Analysis of in vivo editing effect
Genomic DNA
Transcriptomic RNA
Protein

Fig. 3 | Overview of tBE experimental procedures. tBE-mediated editing starts with the design and screening of sgRNA–hsgRNA pairs. After the sgRNA–hsgRNA pair with the highest editing efficiency is selected, OT mutations at the DNA and RNA levels are determined to confirm that no observable OT mutations are induced. If in vivo editing is needed, this sgRNA–hsgRNA pair and the other tBE components are subcloned into the dual AAV packaging plasmids and the corresponding two types of AAV are produced and titers are determined. Then, the two types of AAV are injected into mice and after a certain time period, blood samples and/or tissue samples are collected. Genomic DNA, RNA, proteins and other target components are extracted from blood and/or tissue samples to determine the in vivo editing effects of the tBE, including on-target editing efficiency, potential therapeutic efficacy and OT mutations. d, days.

windows[18,36,54] are recommended. In the future, engineering mA3CDA1 to restrict its ssDNA binding activity may help to narrow the editing window of the tBE.

As the design of tBE system is more complicated than the conventional BEs (six components versus two components) and the total size of tBE vectors is also larger, the construction of the tBE system has higher time and labor demands. Moreover, both the hsgRNA and sgRNA used in the tBE contain RNA aptamers (i.e., MS2 and boxB)[27] and thus are longer than the canonical sgRNA[55–58]. Therefore, in applications that require chemically synthesized gRNAs, the yield of full-length aptamer-containing hsgRNA and sgRNA could be lower than that of canonical sgRNA, which would increase the cost of research.

Experimental design

Design of sgRNA–hsgRNA pairs (Steps 1–7)

The design of the target-site complementary sgRNA for the tBE is similar to that for canonical BEs. First, an appropriate NGG PAM (when nSpCas9 is used in tBE) is selected to ensure that the target cytosine is located in the 7 bp editing window of the tBE (position 3–9 in the spacer region when counting the PAM distal position in the spacer region as the first position; Fig. 2a, sgRNA design). If more than one appropriate PAM can serve as an sgRNA, we suggest designing and testing all possible sgRNAs according to all the appropriate PAMs; typically, the sgRNA with an editing window that accommodates the target cytosine at positions 5–7 will provide optimal editing efficiency.

After a specific sgRNA is selected, the hsgRNAs corresponding to the PAMs in the recommended region (34–91 bp upstream of sgRNA, counting the PAM-to-PAM distance) should be designed (Fig. 2a, hsgRNA design). The hsgRNAs should be designed to target the strand that the sgRNA targets, otherwise the nCas9 in tBE system would generate off-set nicks in opposite DNA strands, which often trigger nonhomologous end joining to induce unintended insertions or deletions (indels). Similarly, if more than one hsgRNA can be designed in the recommended region, we suggest testing all possible hsgRNAs. In contrast, if no hsgRNA can be designed in the recommended region, the sgRNA-downstream region and the further upstream region (>91 bp) may be used to design hsgRNAs. As C-to-T base editing can occasionally occur in the hsgRNA-binding site, we highly recommend using hsgRNAs with a 10 bp spacer region; the use of such hsgRNAs eliminates unintended base editing in the hsgRNA-binding site while having no effect on editing efficiency in the sgRNA-binding site[27].

When more than one sgRNA and multiple corresponding hsgRNAs can be designed for a target site, we suggest screening all the combinations of sgRNAs and hsgRNAs (Fig. 2a, screening of sgRNA–hsgRNA pairs) to select the best sgRNA–hsgRNA pair (Fig. 2a, selection of the optimal sgRNA–hsgRNA pair).

Controls

A potent BE with a high cytidine deamination activity (e.g., BE3 or hA3A–BE3)[1,18] is suggested as the positive control for determining on-target editing, gRNA-dependent DNA OT mutations, gRNA-independent DNA OT mutations, gRNA-independent genome-wide OT mutations and gRNA-independent transcriptome-wide OT mutations. For determining on-target editing, gRNA-dependent DNA OT mutations, gRNA-independent DNA OT mutations and gRNA-independent transcriptome-wide OT mutations, nontreated or irrelevant protein (e.g., GFP)-treated samples can be used as the negative controls. For determining gRNA-independent genome-wide OT mutations, nontreated or Cas9-treated samples can be used as the negative controls. The control groups should be treated same as the test groups and sequenced with a similar depth.

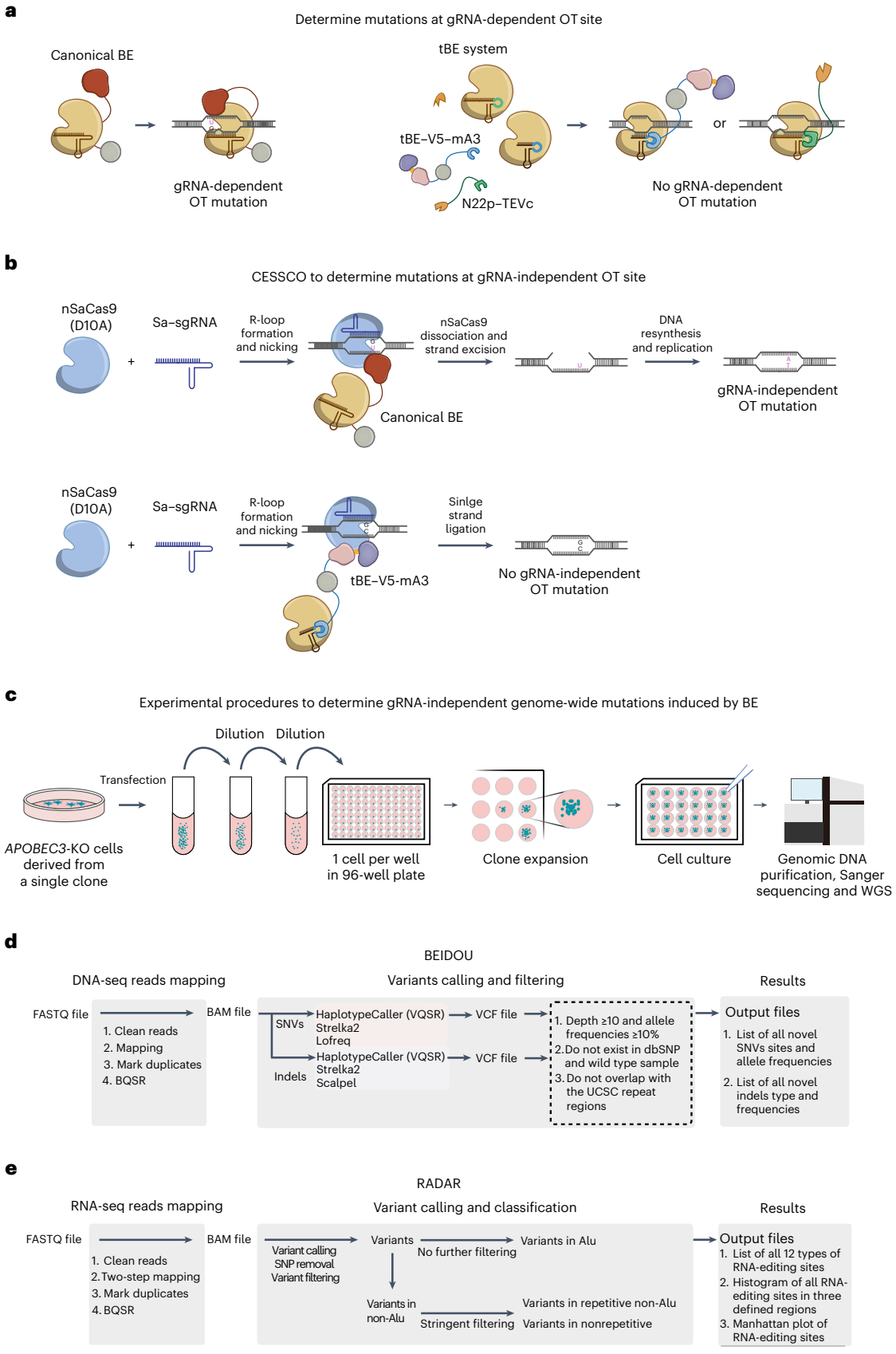

a
Determine mutations at gRNA-dependent OT site
Canonical BE
gRNA-dependent OT mutation
tBE system
tBE–V5–mA3
N22p–TEVc
or
No gRNA-dependent OT mutation
b
CESSCO to determine mutations at gRNA-independent OT site
nSaCas9 (D10A)
Sa–sgRNA
R-loop formation and nicking
nSaCas9 dissociation and strand excision
DNA resynthesis and replication
gRNA-independent OT mutation
Canonical BE
nSaCas9 (D10A)
Sa–sgRNA
R-loop formation and nicking
Sinlge strand ligation
tBE–V5-mA3
No gRNA-independent OT mutation
c
Experimental procedures to determine gRNA-independent genome-wide mutations induced by BE
Dilution
Dilution
Transfection
APOBEC3-KO cells derived from a single clone
1 cell per well in 96-well plate
Clone expansion
Cell culture
Genomic DNA purification, Sanger sequencing and WGS
d
BEIDOU
DNA-seq reads mapping
Variants calling and filtering
Results
FASTQ file
1. Clean reads
2. Mapping
3. Mark duplicates
4. BQSR
BAM file
SNVs
HaplotypeCaller (VQSR)
Strelka2
Lofreq
VCF file
Indels
HaplotypeCaller (VQSR)
Strelka2
Scalpel
VCF file
1. Depth ≥10 and allele frequencies ≥10%
2. Do not exist in dbSNP and wild type sample
3. Do not overlap with the UCSC repeat regions
Output files
1. List of all novel SNVs sites and allele frequencies
2. List of all novel indels type and frequencies
e
RADAR
RNA-seq reads mapping
Variant calling and classification
Results
FASTQ file
1. Clean reads
2. Two-step mapping
3. Mark duplicates
4. BQSR
BAM file
Variant calling
SNP removal
Variant filtering
Variants
No further filtering
Variants in Alu
Variants in non-Alu
Stringent filtering
Variants in repetitive non-Alu
Variants in nonrepetitive
Output files
1. List of all 12 types of RNA-editing sites
2. Histogram of all RNA-editing sites in three defined regions
3. Manhattan plot of RNA-editing sites

Fig. 4 | Determination of OT mutations. a, At gRNA-dependent OT sites, a canonical BE is recruited by the gRNA to trigger OT mutations. In contrast, the inactive tBE–mA3–V5 effector protein and the N22p–TEVc protein of the tBE system do not trigger gRNA-dependent OT mutations. **b**, At the R-loop region generated by Sa–sgRNA and nSaCas9, a canonical BE triggers gRNA-independent OT mutations because its cytidine deaminase is constitutively active. For **a** and **b**, the gRNA-dependent OT mutations and the gRNA-independent OT mutations at the R-loop regions, respectively, can be determined with amplicon sequencing. **c**, Experimental procedures to determine gRNA-independent genome-wide OT mutations by BEs. In brief, *APOBEC3*-KO cells are edited by a canonical BE (e.g., hA3A–BE3) and tBE–V5–mA3 and then single-cell clones are isolated and cultured. The whole genomes of the edited single-cell clones are sequenced for mutation analysis. **d**, The gRNA-independent genome-wide mutations can be identified by using BEIDOU. BQSR, base quality score recalibration; VQSR, variant quality score recalibration. **e**, To determine gRNA-independent transcriptome-wide mutations, the RNA-seq data can be analyzed with RADAR. Image **c** adapted from ref. 27, Springer Nature Limited.

Determination of gRNA-dependent DNA OT mutations (Steps 46–47)

A cell line with a high-transfection efficiency (such as 293FT) is transfected with a canonical BE/sgRNA (e.g., hA3A–BE3, the positive control for determining OT mutations), tBE/sgRNA/hsgRNA or nontransfected (the negative control for determining OT mutations). The in silico-predicted OT sites (e.g., the top 20 OT sites predicted by Cas-OFFinder[59]) can be used to determine the gRNA-dependent DNA OT mutations after the genomic DNA from transfected and nontransfected cells has been extracted. At these predicted gRNA-dependent genomic OT sites, the canonical BE can induce OT mutations but tBE does not trigger OT mutations (Fig. 4a).

Determination of gRNA-independent DNA OT mutations (Steps 48–57)

The gRNA-independent DNA OT mutations can be determined by co-expressing *Staphylococcus aureus* and *Streptococcus pyogenes* Cas9 orthologs (CESSCO)[27] (Fig. 4b). In CESSCO, the tBE or a canonical BE (e.g., BE3 or hA3A–BE3, the positive control[27]) is cotransfected into cells with *S. aureus* Cas9 nickase D10A (nSaCas9) and an sgRNA for SaCas9 (Sa–sgRNA), which are orthogonal to the SpCas9 system, but without the sgRNA for SpCas9-derived BEs. nSaCas9 and its corresponding Sa–sgRNA form an R-loop at the Sa–sgRNA-binding site and the ssDNA region of the R-loop (i.e., the nontarget strand for Sa–sgRNA) can serve as the substrate for cytidine deaminases. As the cytidine deaminase in canonical BEs is constitutively active, it can cause gRNA-independent deamination of the cytosine in the ssDNA region of the R-loop. Meanwhile, nSaCas9 cleaves the target strand at the Sa–sgRNA-binding site, which triggers endogenous mismatch repair to excise the nicked target strand and resynthesize it, therefore improving the sensitivity of CESSCO to detect gRNA-independent OT mutations[27,47]. Eventually, the C-to-T mutations induced by canonical BEs will be installed in the Sa–gRNA-binding site in the genomic DNA of transfected cells, representing gRNA-independent OT mutations (Fig. 4b). In contrast, tBE–V5–mA3 remains inactive at gRNA-independent OT sites and therefore does not trigger gRNA-independent OT mutations (Fig. 4b).

Determination of gRNA-independent genome-wide OT mutations (Steps 58–64)

As gRNA-independent genome-wide OT mutations are induced randomly in each cell, genomic DNA from a single-cell clone is suggested for whole-genome sequencing (WGS) (Fig. 4c). We recommend using cell lines with low expression levels of the endogenous APOBEC/AID family of cytidine deaminase, e.g., the *APOBEC3*-knockout (*APOBEC3*-KO) 293FT cell line[27,60], to reduce the background number of single-nucleotide variants (SNVs). A classical BE (e.g., BE3 or hA3A–BE3, the positive control[9,10,60]), Cas9 (the negative control) and tBE are used for transfection and then single-cell clones are isolated from the transfected cells. Genomic DNA is extracted from the cells cultured from single-cell clones and editing efficiency is initially determined at the on-target site. Then, the genomic DNA from the single-cell clones that are edited biallelically is subjected to WGS and the gRNA-independent genome-wide OT mutations can be analyzed with the base/prime editor-induced DNA off-target site identification unified toolkit (BEIDOU)[27,60] (Fig. 4d).

Determination of gRNA-independent transcriptome-wide OT mutations (Steps 65–70)

To determine gRNA-independent transcriptome-wide OT mutations, the tBE and a canonical BE (e.g., BE3 or hA3A–BE3, the positive control[11,12,60]) are used to transfect a common cell line with high transfection efficiency (such as 293FT). As the levels of transcriptome-wide OT mutations are correlated with the expression levels of BEs, transcriptomic RNA is extracted from the transfected cells when the expression levels of BEs peak (usually 48 h after plasmid transfection).

The transcriptomic RNA is then subjected to RNA sequencing (RNA-seq) and gRNA-independent transcriptome-wide OT mutations can be identified by using the RNA-editing analysis pipeline to decode all twelve types of RNA-editing event (RADAR)[36] (Fig. 4e).

Design of AAV-packaging vectors (Steps 71–87)
If in vivo base editing is needed (e.g., modeling human diseases[21,61] or treating a disease in an animal model[39,62]), the screening of sgRNA–hsgRNA pairs should be first performed in a cell line with high transfection efficiency (e.g., 293FT), to examine editing efficiency, and/or in a target-tissue-derived cell line, to detect the cellular effect of base editing. As AAVs can carry cargos of ~4.5 kb in length, a dual-AAV system is used to package the tBE. The nSpCas9 used in the tBE system along with its promoter and polyA signal can be packaged into one AAV (Fig. 2b, pAAV–nSpCas9) and the other components of the tBE system, i.e., the sgRNA, hsgRNA, TEVn, N22p–TEVc and effector protein tBE–V5–mA3, can be packaged into another AAV (Fig. 2b, pAAV–TargetSite–tBE–V5–mA3). All of the components of the tBE are then cloned into two AAV-packaging vectors (Fig. 2b) after the appropriate serotype of AAV is selected based on the target organs or tissues (e.g., AAV8 for the liver[30]). By coexpressing the Rep–Cap plasmid (Fig. 2b, pAAV–DJ/8, containing the genes to construct AAV protein shell and determine AAV serotype) and Helper plasmid (pHelper, containing the adenovirus E4, E2A and VA genes to assist AAV packing) with each of the two tBE-packaging plasmids (Fig. 2b, pAAV–nSpCas9 or pAAV–TargetSite–tBE–V5–mA3), two types of AAV can be produced for in vivo base editing.

Determination of in vivo editing effects (Steps 105–111)
After the tBE system has been stably expressed in the target organs or tissues through AAV delivery, the effect of in vivo base editing can be determined. The target issues can be extracted and homogenized and then the corresponding DNA, RNA and protein samples can be prepared accordingly. The genomic DNA can be subjected to amplicon sequencing to determine on-target editing and gRNA-dependent DNA OT mutations, with samples from untreated animals serving as negative controls. Genomic DNA can also be subjected to WGS to determine gRNA-independent genome-wide OT mutations, with samples from nontarget organs or tissues of the same treated animal serving as negative controls to reduce background SNVs, as the SNVs among the genomes of different organs or tissues in the same animal are much less than the SNVs among the genomes of different animals in a species. The RNA samples can be subjected to RNA-seq analysis to determine gRNA-independent transcriptome-wide OT mutations, with samples from untreated animals serving as negative controls. Finally, the protein samples can be subjected to western blotting or enzyme-linked immunoassay (ELISA) to determine the efficacy of in vivo base editing, with samples from untreated animals serving as negative controls.

Materials

Reagents

- Plasmid pUC57–sgRNA–MS2–U6 (Addgene, cat. no. 171694)
- Plasmid pU6–ccdB–boxB–tBE–V5–mA3 (Addgene, cat. no. 171693)
- Plasmid pU6–SasgRNA scaffold–CMV–nSaCas9 (Addgene, cat. no. 171695)
- Plasmid pUC57kan–T7–gRNA–U6 V2 (Addgene, cat. no. 115520)
- Plasmid pGL3–U6–XsgRNA–ccdB–EF1a–Puromycin (Addgene, cat. no. 115519)
- Plasmid pEF1a–nSpCas9 (Addgene, cat. no. 171691)
- Plasmid pST1374–NLS–flag-linker–Cas9 (Addgene, cat. no. 44758)
- Plasmid pEF1a–nSpCas9–2a–eGFP (Molecular Cloud, cat. no. MC_0101389)
- Plasmid pAAV–MCS (Cell Biolabs, Inc., cat. no. VPK-410)
- AAV–DJ/8 Packaging System (Plasmid pAAV–DJ/8, Plasmid pHelper) (Cell Biolabs, Inc., cat. no. VPK-400-DJ-8)
- All primers used in this protocol (Sangon Biotech) are listed in Table 1.

Table 1 | Example primer sequences

Step	Primer name	Sequence (5′–3′)	Purpose	Primer usage
3,40,41, 46,57	hgDNA_FOR	N_{20-30}	Forward primer to amplify genomic DNA for amplicon sequencing	The primer located 200–300 bp upstream to your target site or gRNA-dependent OT site for amplicon sequencing. We recommend using the Primer-BLAST of NCBI to design your primer. The size of PCR product is generally 400–600 bp. The G/C content of primer is generally between 40% and 60%
3,40,41, 46,57	hgDNA_REV	N_{20-30}	Reverse primer to amplify genomic DNA for amplicon sequencing	The primer located 200–300 bp downstream to your target site or gRNA-dependent OT site for amplicon sequencing. We recommend using the Primer-BLAST of NCBI to design your primer. The size of PCR product is generally 400–600 bp. The G/C content of primer is generally between 40% and 60%
7,8,9	hsgRNA_MS2_FOR	ATGCGTCTC*AACCG* (**N**$_{10}$) GTTTGAGAGCTAGGCC AACATGA	Inserting both hsgRNA–MS2 and sgRNA–boxB simultaneously into the tBE–V5–mA3-expressing plasmid pU6–ccdB–boxB–tBE–V5–mA3	This primer contains an Esp3I recognition site CGTCTC and an Esp3I cleavage site *A/ACCG*. The bold sequence **N**$_{10}$ is the spacer region of hsgRNA–MS2 and the sequence downstream of **N**$_{10}$ is cognate to the plasmid pUC57–sgRNA–MS2–U6. When designing your own primer, please replace the **N**$_{10}$ with the spacer region of your hsgRNA–MS2
7,8,9	sgRNA_boxB_REV	ATGCGTCTC*GAAAC*(**N**$_{20}$) CGGTGTTTCGTCCTTT CCACAAG	Inserting both hsgRNA–MS2 and sgRNA–boxB simultaneously into the tBE–V5–mA3-expressing plasmid pU6–ccdB–boxB–tBE–V5–mA3	This primer contains an Esp3I recognition site CGTCTC and an Esp3I cleavage site *G/AAAC*. The bold sequence **N**$_{20}$ is the reverse complementary sequence of the spacer region of sgRNA–boxB and the sequence downstream of **N**$_{20}$ is cognate to the plasmid pUC57–sgRNA–MS2–U6. When designing your own primer, please replace the **N**$_{20}$ with the reverse complementary sequence of the spacer region of your sgRNA–boxB
48	Sa–sgRNA_FOR	*ACCG*(**N**$_{20}$)	Inserting Sa–sgRNA into the vector pU6–SasgRNA scaffold–CMV–nSaCas9	Forward primer with an overhang *ACCG*, which is ligatable to the BsaI–HFv2 linearized vector (pU6–SasgRNA scaffold–CMV–nSaCas9) after annealing to Sa–sgRNA_REV. The bold sequence **N**$_{20}$ is same as the spacer region of Sa-sgRNA. When designing your own primer, please replace the **N**$_{20}$ with the spacer region of your Sa–sgRNA
48	Sa–sgRNA_REV	*AAAC*(**N**$_{20}$)	Inserting Sa–sgRNA into the vector pU6–SasgRNA scaffold–CMV–nSaCas9	Reverse primer with an overhang *AAAC*, which is ligatable to the BsaI–HFv2 linearized vector (pU6–SasgRNA scaffold–CMV–nSaCas9) after annealing to Sa–sgRNA_FOR. The bold sequence **N**$_{20}$ is the reverse complementary sequence of the spacer region of Sa-sgRNA. When designing your own primer, please replace the **N**$_{20}$ with the reverse complementary sequence of the spacer region of your Sa–sgRNA
1 in Box 2	sghA3A_FOR	ATGCGTCTC*AACCG* **CTAGTGGGTGG AATCCGGAA**GTTTT AGAGCTAGAAATAGCAAG	Inserting both sghA3A and sghA3H simultaneously into the vector (pGL3–U6–XsgRNA–ccdB–EF1a–Puromycin) for knocking out human *APOBEC3* (*hA3*) cluster	This primer contains an Esp3I recognition site CGTCTC and an Esp3I cleavage site *A/ACCG*. The bold sequence is the spacer region of sghA3A and the sequence downstream of bold sequence is cognate to the vector pUC57kan–T7–gRNA–U6 V2 The sghA3A targeted site located upstream (~2.6 kb to the transcriptional start site) of human *APOBEC3A*
1 in Box 2	sghA3H_REV	ATGCGTCTC*GAAAC* **AGGTATTAAGCTCG GTACAG**CGGTGTTT CGTCCTTTCCACAAG	Inserting both sghA3A and sghA3H simultaneously into the vector (pGL3–U6–XsgRNA–ccdB–EF1a–Puromycin) for knocking out *hA3* cluster	This primer contains an Esp3I recognition site CGTCTC and an Esp3I cleavage site *G/AAAC*. The bold sequence is the reverse complementary sequence of the spacer region of sghA3H and the sequence downstream of bold sequence is cognate to the vector pUC57kan–T7–gRNA–U6 V2. The sghA3H targeted site located downstream (~12.8 kb to the transcriptional start site) of human *APOBEC3H*
16 in Box 2	hA3A_up_FOR	GTACCATGAATTGC AAGTGTGT	Forward primer to identify whether *hA3* cluster is knocked out in 293FT cells	The forward primer located upstream (~3 kb to the transcriptional start site) of human *APOBEC3A*. The size of PCR products is ~600 bp
16 in Box 2	hA3H_dn_REV	CTGTGTTAGCCA GGATCATCT	Reverse primer to identify whether *hA3* cluster is knocked out in 293FT cells	The reverse primer located downstream (~13 kb to the transcriptional start site) of human *APOBEC3H*. The size of PCR products is ~600 bp
71	ITR_U6_F	ctaggggttcctgcggcc gcacgcgtGAGCGGCC GCCCCCTTCACC	Constructing the AA-packaging plasmid pAAV–TargetSite–tBE–V5–mA3	The sequence in lower case is cognate to the vector pAAV–MCS. The homologous sequence of primer and template is generally 18–25 bp
71	ITR_sgRNA-2_R	cttggaaatccccgtgagtc AAGCTTGCATGCAGG CCTCTG	Constructing the AAV-packaging plasmid pAAV–TargetSite–tBE–V5–mA3	The sequence in lower case is cognate to the DNA fragment (CMVmini). The homologous sequence of primer and template is generally 18–25 bp
72	CMVmini_F	GACTCACGGGGA TTTCCAAGTCTC	Constructing the AAV-packaging plasmid pAAV–TargetSite–tBE–V5–mA3	The sequence is cognate to the vector pST1374–NLS–flag-linker–Cas9. The homologous sequence of primer and template is generally 18–25 bp

Table 1 (continued) | Example primer sequences

Step	Primer name	Sequence (5′–3′)	Purpose	Primer usage
72,81	CMVmini_R	GATCTGACGGTT CACTAAACGAGCTCT GCTTAT	Constructing the AAV-packaging plasmid pAAV–TargetSite–tBE–V5–mA3 and pAAV–nSpCas9	The sequence is cognate to the vector pST1374–NLS–flag-linker–Cas9. The homologous sequence of primer and template is generally 18–25 bp
74	ITR_SV40NLS-MCP_F	cgtttagtgaaccgtcagatc ACCGGTGCCACCATGCC AAAGAAGAAGAGAAAAGT GATGGCCTCTAACTTCACCCA	Constructing the AAV-packaging plasmid pAAV–TargetSite–tBE–V5–mA3	The sequence in lower case is cognate to the DNA fragment (CMVmini). The homologous sequence of primer and template is generally 18–25 bp
74,83	ITR_bGH_R	ggggttcctgcggccgct cggtccgCCATAGAG CCCACCGCATCCCC	Constructing the AAV-packaging plasmid pAAV–TargetSite–tBE–V5–mA3 and pAAV–nSpCas9	The sequence in lower case is cognate to the plasmid pAAV–MCS. The homologous sequence of primer and template is generally 18–25 bp
81	ITR_CMVmini_F	ctaggggttcctgcgg ccgcacgcgtGACTCA CGGGGATTTCCAAGTCTC	Constructing the AAV-packaging plasmid pAAV–nSpCas9	The sequence in lower case is cognate to the plasmid pAAV–MCS. The homologous sequence of primer and template is generally 18–25 bp
83	ITR_SpCas9_F	cgtttagtgaaccgtcagatc ACCGGTGCCACCATGGA CAAGAAGTACAGCATCGGC	Constructing the AAV-packaging plasmid pAAV–nSpCas9	The sequence in lower case is cognate to the DNA fragment (ITR–CMVmini). The homologous sequence of primer and template is generally 18–25 bp
101	ITR-Fwd	GGAACCCCTAGTGATGGAGTT	Forward primer to test the AAV titert by using qPCR	The sequence is cognate to ITR
101	ITR-Rev	CGGCCTCAGTGAGCGA	Reverse primer to test the AAV titer by using qPCR	The sequence is cognate to ITR

- PrimeSTAR HS DNA Polymerase (Takara, cat. no. R010B)
- ChamQ Universal SYBR qPCR Master Mix (Vazyme, cat. no. DC711-02)
- Agarose (Thermo Fisher, cat. no. 75510019)
- T4 DNA ligase (NEB, cat. no. M0202M)
- T4 DNA ligase reaction buffer (NEB, cat. no. B0202S)
- NEBuilder HiFi DNA Assembly Master Mix (NEB, cat. no. E2621X)
- BsaI–HFv2 (NEB, cat. no. R3733V)
- Esp3I (NEB, cat. no. R0734L)
- MluI–HF (NEB, cat. no. R3198L)
- RsrII (NEB, cat. no. R0501S)
- DNase I (NEB, cat. no. M0303S)
- rCutsmart buffer (NEB, cat. no. B6004S)
- DNase I reaction buffer (NEB, cat. no. B0303S)
- 10× DNA loading buffer (Vazyme, cat. no. P022-01)
- 100 bp DNA ladder (Vazyme, cat. no. MD104-01)
- DL 15000 DNA marker (Vazyme, cat. no. MD103-01)
- Ultra GelRed (Vazyme, cat. no. GR501-01)
- Ethidium bromide (Sigma, cat. no. 1239-45-8)
  ▲ **CAUTION** Ethidium bromide is a mutagen that can be absorbed through mouth and skin. Wear goggles, masks and gloves when handling ethidium bromide.
- Ampicillin (ABCONE, cat.no. A08483-100G)
- LB medium powder (Sangon Biotech, cat. no. A507002-0250)
- LB agar powder (Sangon Biotech, cat. no. A507003-0250)
- NucleoBond Xtra Midi (Macherey-Nagel, cat. no. 740410.50)
- TIANpure midi plasmid kit (TIANGEN, cat. no. DP107)
- FastPure Gel DNA extraction mini kit (Vazyme, cat. no. DC301-01)
- QuickExtract DNA extraction solution (Epicentre, cat. no. QE09050)
- Lipofectamine LTX reagent (Thermo Fisher, cat. no. 15338100)
- EZ Cell Transfection reagent (Life-iLab, cat. no. AC04L092)
- FastPure DNA isolation kit (Vazyme, cat. no. DC102-01)
- TransZol Up Plus RNA kit (TransGen, cat. no. ER501-01)
- Dulbecco's modified Eagle medium (DMEM; Gibco/Thermo Fisher, cat. no. 10566)
- Opti-MEM (Thermo Fisher, cat. no. 31985070)

- FBS (Gibco/Thermo Fisher, cat. no. 16000-044)
- Penicillin/streptomycin (Gibco/Thermo Fisher, cat. no. 15140122)
- Puromycin (InvivoGen, cat. no. ant-pr-1)
- Blasticidin (InvivoGen, cat. no. ant-bl-1)
- 0.25% (wt/vol) Trypsin–ethylenediaminetetraacetic acid (EDTA; Gibco/Thermo Fisher, cat. no. 25200072)
- Mouse PCSK9 ELISA kit (R&D Systems, cat. no. MPC-900)
- Total Cholesterol kit (Shanghai ShenSuo UNF Medical Diagnostic Articles, cat. no. 1040280)
- E.Z.N.A. Tissue DNA kit (Omega Bio-Tek, cat. no. D3396-01)
- HiPure Total RNA Plus Mini kit (Magen, cat. no. R4121-02)
- NEBNext Ultra II FS DNA Library prep kit for Illumina (NEB, cat. no. E7103L)
- Qubit dsDNA Quantification Assay Kits (Thermo Fisher, cat. no. Q32854)
- TruSeq Stranded Total RNA Library prep kit with Ribo-Zero Globin High Throughput (96 samples, 96 indexes) (Illumina, cat. no. RS-122-2503)
- AAVpro Concentrator (Takara, cat. no. 6674)
- AAVpro Purification kit maxi (All Serotypes) (Takara, cat. no. 6666)
- Isoflurane (RWD Life Science, cat. no. R510-22-10)
  ▲ **CAUTION** Isoflurane is a general anesthetic. Wear a face mask, protective goggles, gloves and laboratory coat when handling isoflurane. Handle isoflurane in a hood.

Biological materials

- Trans5α Chemically Competent Cells (TransGen, cat. no. CD201-01)
- 293FT cell line (Thermo Fisher, cat. no. R70007; RRID: CVCL_6911)
- HEK293T/17 cell line (ATCC, cat. no. CRL-11268; RRID: CVCL_1926)
- N2A cell line (ATCC, cat. no. CCL-131; RRID: CVCL_0470)
  ▲ **CAUTION** The cell lines used in your research should be regularly checked to ensure that they are authentic and are not infected with mycoplasma.
- C57BL/6 mice (female, 6–8 weeks old) purchased from Shanghai Jihui Experimental Animal Breeding. Mice were given free access to food and water and maintained under a 12 h–12 h light–dark cycle with controlled temperature (20–25 °C) and humidity (50 ± 10%).
  ▲ **CAUTION** All the animal experimental procedures are approved by Animal Ethics Committee at ShanghaiTech University and are consistent with governmental regulations for the care and use of animals.

Equipment

- Mini-centrifuges (SCILOGEX, cat. no. D1008E)
- Benchtop centrifuge (Eppendorf, cat. no. 5810R)
- Multipurpose centrifuges (Thermo Fisher, cat. no. 75002446)
- Veriti 96-Well Fast Thermal Cycler (Life Technologies, cat. no. 4375305)
- Floored large refrigerated shaking incubator (Shanghai Zhichu Instrument, cat. no. ZQLY-300S)
- Electric Thermostatic Incubator (Shanghai Jinghong Instrument, cat. no. DNP-9272)
- Blue LED transilluminator (Bio-Friend, cat. no. BD-BGC1)
- Medical icebox (Panasonic, cat. no. MPR-710)
- Low temperature freezer (−20 °C) (Panasonic, cat. no. MDF-539)
- Heating magnetic whisk (SCILOGEX, cat. no. MS7-H-Pro)
- Water bath (Crystal Technology & Industries, cat. no. SYG-1220)
- Thermo Scientific Heracell VIOSCO2 (Thermo Fisher, cat. no. 51030286)
- Biosafety cabinets (Thermo Fisher, cat. no. 1389)
- FACSAria III (BD)
- NanoDrop 2000 (Thermo Fisher)
- AB Quanut Studio 7 Real-Time PCR System (Life Technologies)
- 1.5 mL Eppendorf tube (Pullen, cat. no. PL03001)
- 150 mm tissue-culture (TC)-treated culture dish (Corning, cat. no. 430599)
- CellBIND Surface 100 mm culture dish (Corning, cat. no. 3296)

- 6-Well plates (Corning, cat. no. 3516)
- 24-Well plates (Corning, cat. no. 3524)
- 96-Well plates (Corning, cat. no. 3917)
- 5 mL round-bottom polystyrene test tube (BD Falcon, cat. no. 35054)
- 40 µm cell strainer for use with 50 mL conical tubes, blue, sterile, individually packaged, 50/case (BD Falcon, cat. no. 352340)
- 8-Tube strip (Pullen, cat. no. PC02003)
- PCR tubes, eight tubes per strip, 125 strips per unit (Crystalgen (GY), cat. no. L-2081)
- PCR tubes, domed eight caps per strip, 125 strips per unit (Crystalgen (GY), cat. no. L-2082)
- 15 mL centrifuge tube (PULLEN, cat. no. PL01005)
- 50 mL centrifuge tube (PULLEN, cat. no. PL01006)
- 1 mL sterile insulin syringe with 0.33 × 13 mm-sized needle (KDL, U-100 type 7)
- R500 Small Animal Anesthesia Machine (RWD Life Science)
- General surgery instrument kit for mouse (RWD Life Science, cat. no. SP0009-M)
- Multimode plate reader (TECAN spark)
- Frozen tissue grinder (Shanghai Jing Xin, cat. no. JXFSTPRP-24)

Software

- Calculating frequencies of base substitution and indels (CFBI) v.1.0.0 (https://github.com/YangLab/CFBI)
- BEIDOU v.1.0.0 (https://github.com/YangLab/BEIDOU)
- RADAR v.1.0.0 (https://github.com/YangLab/RADAR)
- FastQC v.0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
- Trimmomatic v.0.38 (http://www.usadellab.org/cms/?page=trimmomatic)
- BWA v.0.7.17 (http://bio-bwa.sourceforge.net/)
- Samtools v.1.9 (https://github.com/samtools/samtools)
- HISAT2 v.2.1.0 (https://daehwankimlab.github.io/hisat2/)
- Picard v.2.21.2 (https://broadinstitute.github.io/picard/)
- GATK v.4.1.3 (https://gatk.broadinstitute.org)
- BLAT v.364 (https://genome.ucsc.edu/cgi-bin/hgBlat)
- LoFreq v.2.1.3.1 (http://csb5.github.io/lofreq/)
- Strelka2 v.2.9.10 (https://github.com/Illumina/strelka)
- Scalpel v.0.5.4 (https://scalpel.sourceforge.net/)
- R v.3.5.1 (https://www.r-project.org)

Reagent setup

TAE buffer

The 50× TAE stock solution contains 2 mol/L Tris, 0.1 mol/L EDTA disodium salt and 57.1 mL/L acetic acid. Add double-distilled water (ddH_2O) to a final volume of 1 L and pH of 8.0, place the solution on the magnetic stirrer and stir until it is completely dissolved. The 50× TAE stock solution can be stored at room temperature (~16–25 °C) for several years, dilute it with ddH_2O to 0.5× TAE buffer when using.

DMEM-1

DMEM with 10% (vol/vol) FBS and 1% (vol/vol) penicillin/streptomycin. Store at 4 °C for several weeks.

DMEM-2

DMEM with 10% (vol/vol) FBS. Store at 4 °C for several weeks.

DMEM-3

DMEM-1 with the puromycin at a final concentration of 4 µg/mL. Store at 4 °C for several weeks.

DMEM-4

DMEM-1 with the puromycin at a final concentration of 10 µg/mL. Store at 4 °C for several weeks.

0.5 M EDTA buffer
Weigh 186.1 g of $Na_2EDTA{\cdot}2H_2O$ in a 1 L beaker. Add ~800 mL of ddH_2O and stir thoroughly. Adjust pH to 8.0 with NaOH (~20 g NaOH). Add ddH_2O and adjust the volume of the solution to 1 L. Autoclave and store at room temperature for several years.

Procedure

Design of sgRNA–boxB–hsgRNA–MS2 pairs

● **TIMING 2 days**

1. Download target gene sequence from the database. We recommend downloading the corresponding sequences from a commonly used database (e.g., National Center for Biotechnology Information (NCBI) or University of California, Santa Cruz (UCSC) genome browser).
2. Design and order a pair of polymerase chain reaction (PCR) primers located 200–300 bp upstream and downstream of the target site separately, and the primers should be complementary to opposite DNA strands. Dilute the ordered primer dry powder or stock solution with nuclease-free H_2O to a final concentration of 10 μM.
   ▲ **CRITICAL STEP** Using nuclease-free H_2O to dilute PCR templates and primers can minimize the influence on the PCR reaction. Primers diluted with nuclease-free H_2O can be stored at 4 °C for a few weeks or at −20 °C for a longer time.
3. Extract genomic DNA from the untreated cells or tissues planned for your experiments using the E.Z.N.A. Tissue DNA Kit following the manufacturer's instructions. Prepare the PCR1 mixture according to the following table:

Component	Volume	Final concentration
Genomic DNA (Step 1)	1 μL	
hgDNA_FOR (Table 1), 10 μM	1.5 μL	0.3 μM
hgDNA_REV (Table 1), 10 μM	1.5 μL	0.3 μM
PrimeSTAR Buffer (5×)	10 μL	
dNTP Mix (2.5 mM)	4 μL	0.2 mM
PrimeSTAR HS DNA Polymerase	1 μL	
Nuclease-free H_2O	31 μL	
Total	50 μL	

4. Perform the PCR1 reaction under the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	94 °C, 5 min		
2–36	98 °C, 15 s	68 °C, 15 s	72 °C, 30 s
37			72 °C, 3 min

   ▲ **CRITICAL STEP** Any manufacturer's qualified high-fidelity DNA polymerase can be used, but please modify the PCR mixture and procedure according to manufacturer's instructions.
5. The PCR products were subject to Sanger sequencing to determine the actual genotype of the untreated cells or tissues for your experiments.
   ▲ **CRITICAL STEP** It is important to identify whether the genomes of your cells or experimental animals contain any SNV at the target site. SNV may affect editing efficiency if mismatches occur between the actual target sequence and the spacer region of sgRNA–boxB (or hsgRNA–MS2) designed according to the sequences downloaded from the database.

6. Design the pairs of sgRNA–boxB–hsgRNA–MS2. According to the location of the target cytosine, search for proper PAMs around to ensure the target cytosine located in the editing window of tBE.
   ▲ **CRITICAL STEP** Design all suitable sgRNA–boxB and then all suitable hsgRNA–MS2 as described in the 'Experimental design' section to screen for the pair of sgRNA–boxB and hsgRNA–MS2 that induces the highest editing efficiency at the target cytosine (Fig. 2a).
7. Design the PCR primers for constructing the gRNA-containing plasmids, each of which expresses both hsgRNA–MS2 and sgRNA–boxB. Replace the N_{10} of hsgRNA_MS2_FOR (Table 1) with the spacer region of the designed hsgRNA–MS2 and the N_{20} of sgRNA_boxB_REV (Table 1) with the reverse complementary sequence of the spacer region of the designed sgRNA–boxB.
   ▲ **CRITICAL STEP** Pay particular attention when designing the primer sgRNA_boxB_REV as the reverse complementary sequence of the designed sgRNA–boxB should be used to replace the N_{20}.

Construction of plasmids each of which expresses a pair of designed hsgRNA–MS2 and sgRNA–boxB

● **TIMING 2–3 d**

8. Preparation before PCR. Dilute the PCR template (pUC57–sgRNA–MS2–U6) with nuclease-free H_2O to a final concentration of 30–50 ng/μL. Meanwhile, dilute the primer (hsgRNA_MS2_FOR/sgRNA_boxB_REV, Step 7) dry powder or stock solution with nuclease-free H_2O to a final concentration of 10 μM (Extended Data Fig. 1a).
9. Use a high-fidelity DNA polymerase to generate the DNA fragment (hsgRNA–MS2–U6–sgRNA) by PCR (Extended Data Fig. 1a). Prepare the PCR2 mixture according to the following table:

Component	Volume	Final concentration
pUC57–sgRNA–MS2–U6 (30–50 ng/μL)	1 μL	
hsgRNA_MS2_FOR (Table 1), 10 μM	1.5 μL	0.3 μM
sgRNA_boxB_REV (Table 1), 10 μM	1.5 μL	0.3 μM
PrimeSTAR Buffer (5×)	10 μL	
dNTP Mix (2.5 mM)	4 μL	0.2 mM
PrimeSTAR HS DNA Polymerase	1 μL	
Nuclease-free H_2O	31 μL	
Total	50 μL	

10. Perform the PCR2 reaction under the cycling conditions as described in Step 4.
11. PCR product purification. Purify the PCR product by using agarose gel (2% (wt/vol), supplemented with 1:10,000 (vol/vol) GelRed or ethidium bromide) electrophoresis. Add 6 μL 10× DNA loading to the 50 μL of PCR product and then load the mixture into a well of the agarose gel. Load 3 μL 100 bp DNA ladder in a flanking well of the same gel. Run the gel in 0.5× TAE buffer at 5 V/cm for 20 min. The size of the target DNA fragment is ~500 bp. Excise the target DNA band with a blue-light gel cutter and extract the DNA fragment by using a FastPure Gel DNA extraction mini kit, following the manufacturer's instructions.
    ▲ **CRITICAL STEP** Excise the gel block as thin as possible to get a good purity of DNA fragment.
12. After extracting the DNA fragment, measure the concentration of extracted DNA fragment (hsgRNA–MS2–U6–sgRNA) on a NanoDrop 2000.
13. The DNA fragment obtained in Step 12 is flanked by two Esp3I sites and correspondingly the vector pU6–ccdB–boxB–tBE–V5–mA3 also contains two Esp3I sites. Digest the DNA fragment and the vector with Esp3I and ligate them with Golden Gate assembly (Extended Data Fig. 1a), using the following mixture:

Component	Amount	Volume
pU6–ccdB–boxB–tBE–V5–mA3	50 ng	
The fragment of hsgRNA–MS2–U6–sgRNA (Step 12)	75 ng	
Esp3I		0.5 µL
T4 DNA ligase		0.5 µL
T4 DNA ligase reaction buffer (10×)		1 µL
Nuclease-free H_2O		Add to 10 µL
Total		10 µL

14. Perform the Golden Gate assembly 1 reaction under the following cycling conditions:

Cycle number	T4 enzyme linkage	Enzyme reaction	Denature
1		37 °C, 5 min	
2–11	16 °C, 10 min	37 °C, 15 min	
12			80 °C, 15 min

15. Add the entire Golden Gate assembly mixture from Step 14 to 50 µL Trans5α Chemically Competent Cells and incubate the Eppendorf tube containing the competent cells on ice for 20 min (Extended Data Fig. 1a).
▲ **CRITICAL STEP** Although most of the commonly used chemical-competent *Escherichia coli* cells can be used in this step, competent cells resistant to the toxicity of *ccdB* products can not be used, such as DB3.1. Because nonlinearized vectors may still exist after the Esp3I restriction in Step 14, using competent cells with *ccdB* resistance will allow the survival of cells that contain the original vectors.
16. Heat shock the competent cells in a 42 °C water bath for 90 s and then immediately keep them on ice for 1 min.
17. Add 500 µL LB medium (without any antibiotics) into the tube, put the tube in a thermostatic shaker and shake the tube at 37 °C at 580*g* for 45 min.
18. Centrifuge the recovered bacteria (1,500*g*) at room temperature for 5 min and then remove most of the supernatant by pipetting, leaving ~50 µL LB medium to resuspend the bacteria. Evenly apply the bacterial solution to an LB agar plate with ampicillin (100 µg/mL) using a glass rod or beads and invert the plate and incubate at 37 °C overnight (Extended Data Fig. 1a).
19. After overnight incubation, single clones are visible on the LB agar plate. Pick up three individual clones and culture each in 20 mL LB medium with ampicillin (100 µg/mL) individually. Shake them in a 37 °C shaker incubator at 580*g* for 12–14 h (Extended Data Fig. 1a).
◆ **TROUBLESHOOTING**
20. Use TIANpure midi plasmid kit to extract the plasmids from the clones in Step 19, following the manufacturer's instructions, and then send these plasmids for Sanger sequencing to confirm that the sequence of the constructed pTargetSite–tBE–V5–mA3 is correct (Extended Data Fig. 1a).
21. Use the same procedure (Steps 8–20) to construct all the plasmids each of which expresses a pair of hsgRNA–MS2 and sgRNA–boxB designed in Steps 6–7 (Fig. 2a and Extended Data Fig. 1b).
■ **PAUSE POINT** The constructed plasmid DNA can be stored at 4 °C for a few weeks or at −20 °C for a longer time.

Determination of on-target editing efficiency in cell lines

● **TIMING 3–4 days**

tBE plasmid transfection

▲ **CRITICAL STEP** In a typical experiment, we recommend using some easily transfected cell lines such as 293FT or N2A cells. The following experimental steps are performed in N2A cells by liposome transfection as an example. N2A cells are maintained in DMEM-1 and regularly tested to exclude mycoplasma contamination.

22. One day before transfection, select the N2A cells with 80–90% confluence and ensure that a monolayer of cells is distributed in the cell culture dish via microscopic examination.
23. Remove DMEM-1 by pipetting, then add 3–5 mL PBS along the sidewall of the dish to cover the dish bottom for washing off the residual medium.
    ▲ **CRITICAL STEP** PBS should not be added directly to the cells as it may detach the cells from the bottom of the dish.
24. Remove the PBS by pipetting, add 1 mL 0.25% (wt/vol) Trypsin–EDTA, carefully move the dish so that the Trypsin–EDTA covers the cells, place the dish in a cell culture incubator and then keep it at 37 °C for 1 min.
25. Add 3 mL of prewarmed DMEM-1 to neutralize the Trypsin–EDTA and use a pipette to blow the dissociated cells off from the bottom of the dish. Gently blow several times to disperse the cells and then transfer the dissociated cells into a 15 mL centrifuge tube.
26. Centrifuge at 300*g* for 5 min at room temperature and then remove the supernatant by pipetting.
27. Add 4 mL prewarmed DMEM-2 and gently blow the cell pellet with a pipette to resuspend it into homogeneous cell solution.
28. Use a hemocytometer or cell counter to determine the density of the cell solution.
29. After determining the density of cell solution, dilute the cell solution to 1.6×10^5 cell/mL with DMEM-2 and then seed the cells into 24-well plates at 0.5 mL/well.
30. Move the 24-well plate in a 'cross-like' way so that the cells are spread evenly in the wells. Then put the plate into a cell culture incubator and incubate it for 24 h at 37 °C with 5% (vol/vol) CO_2.
31. The day of transfection, check the seeded cells under a microscope. A monolayer of cells should attach uniformly to the bottom of wells with a confluence of 60–70%.
    ▲ **CRITICAL STEP** Put the cells back into the incubator until the transfection mixture is ready.
    ◆ **TROUBLESHOOTING**
32. Prepare transfection mixture 1 as follows:

Component	Amount	Volume
pTargetSite–tBE–V5–mA3 (Step 20)	500 ng	1 µL
pEF1a–nSpCas9	500 ng	1 µL
Lipofectamine PLUS		1 µL
Lipofectamine LTX		3 µL
Opti-MEM		250 µL

    Take a 1.5 mL sterile Eppendorf tube and add 250 µL serum-free Opti-MEM. Next, add the plasmids, then add 1 µL Lipofectamine PLUS and mix them by vortex or a pipette. After spinning all components to the bottom of the tube, add 3 µL Lipofectamine LTX, mix them again gently with a pipette and leave this transfection mixture at room temperature for 5 min.
    ▲ **CRITICAL STEP** Avoid using vortex to mix transfection reagents after adding Lipofectamine LTX.
33. Take out the 24-well plate (Step 31) from the incubator, pipette the transfection mixture 1 into each well gently and evenly, then move the plate in a 'cross-like' way and put the 24-well plate back into the incubator again.
34. Enrich transfected cells with antibiotics (recommended) or using fluorescence-activated cell sorting (FACS) (optional, Box 1). Here we use puromycin as the example antibiotics to enrich transfected cells. Twenty-four hours after transfection, take out the 24-well plate (Step 33) from the incubator and remove the culture medium by pipetting.Then add 500 µL of prewarmed DMEM-3 per well along the sidewall of each well. Place the 24-well plate back into the incubator.

Collect cells to obtain genomic DNA

35. Culture the cells in a cell culture incubator at 37 °C with 5% (vol/vol) CO_2 for another 48 h, remove the medium from the 24-well plate by pipetting, add 1 mL PBS per well along the sidewall to prevent blowing off the cells and then remove the PBS by pipetting.
    ▲ **CRITICAL STEP** This step can remove the floating dead cells and residual DMEM-1.

BOX 1

Using FACS to enrich transfected cells

Procedure

1. Enrich the transfected cells with FACS. Forty hours after transfection, remove DMEM-2 from the six-well plate by pipetting, add 1 mL PBS per well along the sidewall of well to avoid blowing off the cells and then remove the PBS by pipetting.
2. Add 300 µL of 0.25% (wt/vol) Trypsin–EDTA directly into each well of the six-well plate and incubate the cells at 37 °C for 1 min.
3. Stop the Trypsin–EDTA digestion by adding 900 µL per well of DMEM-2 to blow the cells away from the plate bottom with a pipette and then collect the dissociated cells into 1.5 mL Eppendorf tubes.
4. Centrifuge the Eppendorf tubes at 300*g* for 5 min and then remove the supernatant by pipetting.
5. Add 400 µL PBS per Eppendorf tube, pipette the pellet gently but thoroughly to disperse the cells and finally transfer the resuspended cells to the special tubes for FACS.
   ▲ **CRITICAL STEP** The resuspended cells should be passed through a cell strainer to remove large clumps of cells and then are transferred into FACS tubes. The FACS tubes that contain resuspended cells should be kept on ice during the entire process of FACS experiments, which is critical for improving cell viability.
6. Collect the cells with top 15% fluorescence intensity by FACSAria III.
   ▲ **CRITICAL STEP** If the editing efficiency is modest at the designed target site, collect the cells with top 5–10% fluorescence intensity.

36. Add 100 µL per well 0.25% (wt/vol) Trypsin–EDTA to the 24-well plate, gently move the 24-well plate so that the Trypsin–EDTA can cover the cells and incubate the plate at 37 °C for 1 min. Add 300 µL per well DMEM-1 to stop the Trypsin–EDTA digestion, blow the cells away from the bottom of the well with a pipette and collect the dissociated cell into 1.5 mL Eppendorf tubes individually.
37. Place the Eppendorf tubes in a centrifuge and centrifuge at 300*g* for 5 min at room temperature, then remove the supernatant by pipetting and discard.
38. Resuspend the cell pellet with 20 µL PBS per sample, then add 20 µL QuickExtractT DNA extraction solution and mix well with a pipette.
    ▲ **CRITICAL STEP** If bulky cell pellets are still visible, we suggest mixing the cells with more PBS and DNA extraction solution with a 1:1 (vol/vol) ratio. Do not vortex in this step as vortexing may stack the dispersed cells again and reduce the effect of cell lysis.
39. Pipette the mixture (40 µL per tube) into an eight-tube strip. Place the eight-tube strip into a thermal cycler and lyse the cells according to the following program:

Cycle number	Cell lysis	Inactivate the proteinase K
1	65 °C, 30 min	
2		98 °C, 2 min

■ **PAUSE POINT** The extracted genomic DNA can be stored at 4 °C for a few weeks or at −20 °C for a longer time.

Determine the on-target editing efficiency

40. Primer design. Design and order two PCR primers (hgDNA_FOR and hgDNA_REV, Table 1) located ~200–300 bp upstream and downstream of the target site.
41. DNA library preparation. Use a high-fidelity DNA polymerase to amplify the sequence of target site from the genomic DNA obtained from Step 39. Prepare PCR3 mixture according to the following table:

Component	Volume	Final concentration
Genomic DNA (Step 39)	1 µL	
hgDNA_FOR (Table 1), 10 µM	1.5 µL	0.3 µM
hgDNA_REV (Table 1), 10 µM	1.5 µL	0.3 µM
PrimeSTAR buffer (5×)	10 µL	
dNTP mix (2.5 mM)	4 µL	0.2 mM
PrimeSTAR HS DNA polymerase	1 µL	
Nuclease-free H_2O	31 µL	
Total	50 µL	

42. Perform PCR3 reaction under the cycling conditions as described in Step 4 and repeat Steps 11–12 to purify and quantify the DNA.
The size of PCR products should be close to the theoretical size (~500 bp).
▲ **CRITICAL STEP** If the volume of PCR product is too large to load in a gel well, use a DNA extraction kit to concentrate it.
▲ **CRITICAL STEP** Ensure that the final concentration of the purified DNA fragment is higher than 15 ng/μL and the total DNA amount is more than 200 ng for each sample.
43. Prepare the DNA libraries by using the NEBNext Ultra II FS DNA library prep kit for Illumina, following the manufacturer's instructions.
44. DNA amplicon sequencing. After the quantification of the amplicon-containing library by using the Qubit dsDNA Quantification Assay Kits, sequence the library using the Illumina HiSeq X Ten system. Evaluate raw read qualities using FastQC (v.0.11.5, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Refer to Step 112 for more details.
45. Compare the on-target editing efficiencies induced by different pTargetSite–tBE–V5–mA3 plasmids that contain different hsgRNA–MS2–sgRNA–boxB pairs and then select the pTargetSite–tBE–V5–mA3 that contains the pair of hsgRNA–MS2–sgRNA–boxB with the highest editing efficiency for the following experiments.
◆ **TROUBLESHOOTING**

Determination of OT mutations

● **TIMING 2–5 weeks**

Determine gRNA-dependent DNA OT mutations

46. Choose a published software (e.g., Cas-OFFinder[59]) to predict possible gRNA-dependent OT sites according to the spacer regions of the sgRNA–boxB and hsgRNA–MS2. Find their locations in the genome and design PCR primers (hgDNA_FOR and hgDNA_REV, Table 1) to amplify the corresponding DNA fragments for deep sequencing and data analysis.
▲ **CRITICAL STEP** The prediction of gRNA-dependent OT sites in this protocol is performed by Cas-OFFinder[59] but it can also be performed by alternative programs (e.g., CRISPOR[63], Elevation[64] and CRISPRme[65]).
47. Repeat Steps 40–44 for PCR, DNA extraction, sequencing and data analysis to determine gRNA-dependent OT mutations.
◆ **TROUBLESHOOTING**

Determine gRNA-independent DNA OT mutations (CESSCO)

48. Generate the nSaCas9–Sa–sgRNA construct. Resuspend the dry powder of DNA oligonucleotides Sa–sgRNA_FOR and Sa–sgRNA_REV (Table 1) with nuclease-free H_2O to a final concentration of 100 μM and prepare the annealing mixture as follows:

Component	Volume	Final concentration
Sa–sgRNA_FOR (Table 1), 100 μM	1 μL	10 μM
Sa–sgRNA_REV (Table 1), 100 μM	1 μL	10 μM
Nuclease-free H_2O	8 μL	
Total	10 μL	

49. Perform annealing in a thermal cycler as follows:

Step number	Step description	Duration
1	95 °C	3 min
2	Cool to 25 °C with a rate of 0.1 °C/s	

50. Linearize the pU6–SasgRNA scaffold–CMV–nSaCas9 with BsaI–HFv2. Prepare the restriction mixture as follows then incubate the mixture at 37 °C for 2–3 h.

Component	Amount	Volume
pU6–SasgRNA scaffold–CMV–nSaCas9	2 μg	
BsaI–HFv2		2 μL
rCutsmart buffer (10×)		3 μL
Nuclease-free H_2O		Add to 30 μL
Total		30 μL

51. Purify the linearized vector by using 1% (wt/vol) agarose gel electrophoresis. Load the linearized vector into a gel well and run the gel in 0.5× TAE buffer at 5 V/cm for 20 min.
52. Extract the linearized vector with a FastPure Gel DNA extraction mini kit as described in Step 11. Measure the concentration of the linearized vector on a NanoDrop 2000.
53. Use T4 DNA ligase to ligate the annealing product of Step 49 to the linearized vector of Step 52. Prepare the ligation mixture as follows:

Component	Amount	Volume
The linearized pU6–SasgRNA scaffold–CMV–nSaCas9 (Step 52)	50 ng	
Annealling product (Step 49)		1 μL
T4 DNA ligase		0.5 μL
T4 DNA ligase reaction buffer (10×)		1 μL
Nuclease-free H_2O		Add to 10 μL
Total		10 μL

54. Keep the ligation mixture at room temperature for 1 h and transform the ligation mixture into competent cells as described in Steps 15–20. Confirm that the sequence of the constructed nSaCas9–Sa–sgRNA-expressing plasmid (pSa–sg–nSaCas9) is correct by Sanger sequencing.
55. Transfect 293FT cells simultaneously with two orthogonal Cas9 systems. Transfection procedures are same as described in Steps 22–34. The transfection mixture 2 is shown in the following table:

Component	Amount	Volume
pSa–sg–nSaCas9 (Step 54)	500 ng	1 μL
pTargetSite–tBE–V5–mA3 (Step 45)	500 ng	1 μL
pEF1a–nSpCas9	500 ng	1 μL
Lipofectamine PLUS		1.5 μL
Lipofectamine LTX		4.5 μL
Opti-MEM		250 μL

56. Collect transfected cells and extract genomic DNA as described in Steps 35–39.
57. Design and order PCR primers (hgDNA_FOR and hgDNA_REV, Table 1) for the Sa–sgRNA target site and perform amplicon sequencing to determine gRNA-independent DNA OT mutations as described in Steps 40–44.
    ◆ **TROUBLESHOOTING**

Determine gRNA-independent genome-wide OT mutations

58. Transfect tBE plasmids into *APOBEC3*-KO 293FT cells following Steps 22–34. One day before transfection, seed *APOBEC3*-KO 293FT cells into a 24-well plate. Refer to Steps 22–30 for more details.
59. On the day of transfection, prepare the transfection mixture. Refer to Steps 31–33 for more details.
60. 24 h after transfection, enrich the transfected cells with antibiotics. Refer to Step 34 for more details.
61. 48 h after cell enrichment, isolate single-cell clones into 96-well plates by limiting dilution (refer to step 12 in Box 2) and culture them in DMEM-1.

BOX 2

Establishing the *APOBEC3*-KO cell line

Procedure

1. As the *APOBEC3*-KO 293FT cell line can be provided on request. If you are not making *APOBEC3*-KO cell line yourself, skip to Step 69. Generate a construct containing sghA3A and sghA3H sgRNAs to knock out the *APOBEC3* cluster. Use the plasmid pUC57kan–T7–gRNA–U6 V2 as the template to amplify a DNA fragment (sghA3A–scaffold–U6–sghA3H) with PCR. Prepare the PCR9 mixture as follows:

Component	Volume	Final concentration
pUC57kan–T7–gRNA–U6 V2 (30–50 ng/μL)	1 μL	
sghA3A_FOR (Table 1), 10 μM	1.5 μL	0.3 μM
sghA3H_REV (Table 1), 10 μM	1.5 μL	0.3 μM
PrimeSTAR buffer (5×)	10 μL	
dNTP Mix (2.5 mM)	4 μL	0.2 mM
PrimeSTAR HS DNA polymerase	1 μL	
Nuclease-free H_2O	31 μL	
Total	50 μL	

2. Perform the PCR9 reaction under the cycling conditions described in Step 4.
3. Purify the amplified DNA fragment by using 2% (wt/vol) agarose gel electrophoresis as described in Step 11 and measure the concentration of DNA fragment (sghA3A–scaffold–U6–sghA3H) with a NanoDrop 2000.
4. Digest the DNA fragment (step 1 in Box 2) and the vector with Esp3I and ligate them with Golden Gate assembly 2. The mixture for Golden Gate assembly 2 is as follows:

Component	Amount	Volume
pGL3–U6–XsgRNA–ccdB–EF1a–Puromycin	50 ng	
The fragment of sghA3A–scaffold–U6–sghA3H (step 1 in Box 2)	75 ng	
Esp3I		0.5 μL
T4 DNA ligase		0.5 μL
T4 DNA ligase reaction buffer (10×)		1 μL
Nuclease-free H_2O		Add to 10 μL
Total		10 μL

5. Perform the Golden Gate assembly 2 reaction under the cycling conditions as described in Step 14.
6. Transform the Golden Gate assembly mixture into competent cells as described in Steps 15–20. Confirm that the sequence of constructed plasmid (pU6–sghA3A–U6–sghA3H) for knocking out the *APOBEC3* cluster is correct by Sanger sequencing.
7. Seed 293FT cells for transfection according to Steps 22–31.
8. Prepare the cell transfection mixture 6 as follows:

Component	Amount	Volume
pU6–sghA3A–U6–sghA3H (step 6 in Box 2)	500 ng	1 μL
pST1374–NLS–flag-linker–Cas9	500 ng	1 μL
Lipofectamine PLUS		1 μL
Lipofectamine LTX		3 μL
Opti-MEM		250 μL

9. Transfect 293FT cells according to Steps 32–33.
10. 24 h after transfection, replace the medium of the transfected cells with 500 μL prewarmed DMEM-4 per well.
11. 48 h after medium replacement, isolate single-cell clones into 96-well plates by limiting dilution.
12. Digest the cells as described in Steps 22–28 and determine the cell density. Prepare 1 mL DMEM-1 containing 1.06×10^4 cells in a fresh 1.5 mL sterile EP. Transfer 100 μL cell suspension from the first 1.5 mL sterile Eppendorf tube into a second 1.5 mL sterile Eppendorf tube by pipetting and add then 900 μL DMEM-1. Transfer 100 μL of the cell suspension from the second 1.5 mL sterile Eppendorf tube into a 50 mL centrifuge tube by pipetting and then add 15.8 mL DMEM-1, which leads to the final cell density of 1 cell per 150 μL. Thoroughly blow the cell suspension in the 50 mL centrifuge tube and transfer the cell suspension to a 96-well plate (150 μl per well) by pipetting. Culture the clones in DMEM-1 in a cell culture incubator at 37 °C with 5% (vol/vol) CO_2 for 3 weeks.
▲ **CRITICAL STEP** Add medium into the 96-well plate after culturing for 10 d.
13. After culturing for 3 weeks, remove DMEM-1 by pipetting, then add 100 μL PBS along the sidewall of the dish to cover the dish bottom for washing off the residual medium.
14. Remove the PBS by pipetting, add 10 μL 0.25% (wt/vol) Trypsin–EDTA, carefully move the dish so that the Trypsin–EDTA covers the cells, place the dish in a cell culture incubator and then keep it at 37 °C for 1 min.
15. Add 100 μL of prewarmed DMEM-1 to neutralize the Trypsin–EDTA and use a pipette to blow the dissociated cells off from the bottom of the dish. Gently blow several times to disperse the cells and transfer 80 μL of the cells into 48-well plates for culturing.
16. Transfer the other 20 μL of the cells into an eight-tube strip and extract the genomic DNA as described in Steps 38–39, and then use PCR to identify the genotype of each clone. Prepare PCR10 as follows:

Component	Volume	Final concentration
Genomic DNA (step 16 in Box 2)	1 μL	
hA3A_up_FOR (Table 1), 10 μM	1.5 μL	0.3 μM
hA3H_dn_REV (Table 1), 10 μM	1.5 μL	0.3 μM
PrimeSTAR buffer (5×)	10 μL	
dNTP mix (2.5 mM)	4 μL	0.2 mM
PrimeSTAR HS DNA polymerase	1 μL	
Nuclease-free H_2O	31 μL	
Total	50 μL	

(continued from previous page)

17. Perform PCR10 reaction under the cycling conditions as described in Step 4.
18. The PCR product is evaluated with 2% (wt/vol) agarose gel electrophoresis. The theoretical size of PCR products should be close to 500 bp, which is resulted from the joining of two DNA ends caused by Cas9 cleavage at the target sites of sghA3A and sghA3H in the *APOBEC3*-knockout (*APOBEC3*-KO) clones. Send the PCR9 product for Sanger sequencing to confirm the sequence of junction site.
19. Culture the potential *APOBEC3*-KO single-cell clones for another 2 or 3 weeks and extract the cellular mRNA using TransZol Up Plus RNA kit following the instructions from 1×10^5 cells for each clone. Compare the mRNA expression levels of endogenous *APOBEC3*s in the expanded clones with those in the wild-type 293FT cells with RT–qPCR[27] (Supplementary Table 1). After confirming that no endogenous *APOBEC3* is expressed at mRNA level, the *APOBEC3*-KO 293FT cell line is established successfully.

▲ **CRITICAL STEP** Please store multiple tubes of *APOBEC3*-KO 293FT cells in liquid nitrogen for future experiments for several years.

62. After culturing for another 2 or 3 weeks, extract the genomic DNA from a portion of cells for each clone with FastPure DNA isolation kit.
63. Amplify the on-target site sequence and identify the clones in which the on-target site is edited biallelic by Sanger sequencing.
64. WGS. Extract >1 μg genomic DNA from each biallelically edited clone using E.Z.N.A. Tissue DNA kit, prepare the DNA-seq libraries with NEBNext Ultra II FS DNA library prep kit for Illumina and sequence the libraries with Illumina HiSeq X Ten system. The analysis of gRNA-independent genome-wide OT mutations is described in Steps 113–114.

Determine gRNA-independent transcriptome-wide OT mutations

65. Seed N2A cells with a density of 5×10^5 cells per well into six-well plates 1 day before transfection and culture the cells in 2 mL DMEM-2 per well overnight. Refer to Steps 22–31 for more details.
66. Transfect tBE plasmids into N2A cells. The transfection mixture 3 is shown in the following table:

Component	Amount	Volume
pTargetSite–tBE–V5–mA3 (Step 20)	1,000 ng	2 μL
pEF1a–nSpCas9–2a–eGFP	1,000 ng	2 μL
Lipofectamine PLUS		2 μL
Lipofectamine LTX		6 μL
Opti-MEM		250 μL

Refer to Steps 32–33 for more details.

67. Enrich the transfected cells with FACS. Refer to Box 1.

◆ **TROUBLESHOOTING**

68. Extract total RNA and perform RNA-seq. Use the TransZol Up Plus RNA kit to extract the total RNA according to the manufacturer's instructions. The extracted RNA should be more than 1.5 μg for the next step.

◆ **TROUBLESHOOTING**

69. Use the TruSeq Stranded Total RNA with Ribo-Zero Globin to prepare the indexed RNA libraries and sequence the libraries with the Illumina HiSeq X Ten system.
70. The analysis of gRNA-independent transcriptome-wide OT mutations is described in Steps 115–117.

Production of AAVs containing the tBE system

● **TIMING 1 week**

Construct the AAV-packaging plasmid corresponding to pTargetSite–tBE–V5–mA3

71. Use a high-fidelity DNA polymerase to generate DNA fragments. Prepare PCR4 mixture to generate the DNA fragment (U6–hsgRNA–MS2–U6–sgRNA–boxB) as follows:

Component	Volume	Final concentration
pTargetSite–tBEV5–mA3 (Step 20) (30–50 ng/μL)	1 μL	
ITR_U6_F (Table 1), 10 μM	1.5 μL	0.3 μM
ITR_sgRNA-2_R (Table 1), 10 μM	1.5 μL	0.3 μM
PrimeSTAR buffer (5×)	10 μL	
dNTP mix (2.5 mM)	4 μL	0.2 mM
PrimeSTAR HS DNA polymerase	1 μL	
Nuclease-free H_2O	31 μL	
Total	50 μL	

72. Prepare the PCR5 mixture to generate the DNA fragment (CMVmini) according to the following table:

Component	Volume	Final concentration
pST1374–NLS–flag-linker–Cas9 (30–50 ng/μL)	1 μL	
CMVmini_F (Table 1), 10 μM	1.5 μL	0.3 μM
CMVmini_R (Table 1), 10 μM	1.5 μL	0.3 μM
PrimeSTAR buffer (5×)	10 μL	
dNTP mix (2.5 mM)	4 μL	0.2 mM
PrimeSTAR HS DNA polymerase	1 μL	
Nuclease-free H_2O	31 μL	
Total	50 μL	

73. Perform PCR4 and PCR5 reactions under the cycling conditions as described in Step 4.
74. Prepare the PCR6 mixture to generate the DNA fragment (tBE–V5–mA3–bGH) according to the following table:

Component	Volume	Final concentration
pTargetSite–tBE–V5–mA3 (Step 20) (30–50 ng/μL)	1 μL	
ITR_SV40NLS-MCP_F (Table 1), 10 μM	1.5 μL	0.3 μM
ITR_bGH_R (Table 1), 10 μM	1.5 μL	0.3 μM
PrimeSTAR buffer (5×)	10 μL	
dNTP mix (2.5 mM)	4 μL	0.2 mM
PrimeSTAR HS DNA polymerase	1 μL	
Nuclease-free H_2O	31 μL	
Total	50 μL	

75. Perform the PCR6 reaction under the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	94 °C, 5 min		
2–36	98 °C, 15 s	68 °C, 15 s	72 °C, 2 min
37			72 °C, 3 min

76. Purify the products of PCR4, PCR5 and PCR6 with 1% (wt/vol) agarose gel electrophoresis as described in Step 11.
77. Linearize pAAV–MCS with MluI–HF and RsrII. Prepare the restriction mixture according to the following table and then incubate the mixture in a water bath at 37 °C for 4 h.

Component	Amount	Volume
pAAV–MCS	3 μg	
MluI–HF		1 μL
RsrII		1 μL
rCutsmart buffer (10×)		5 μL
Nuclease-free H_2O		Add to 50 μL
Total		50 μL

78. Purify the MluI–HF/RsrII-restricted pAAV–MCS vector with 1% (wt/vol) agarose gel electrophoresis as described in Step 11.
79. Use NEBuilder HiFi DNA assembly master mix to assemble the DNA fragments of Step 76 into the linearized vector of Step 78. Prepare the HiFi DNA assembly mixture 1 according to the following table:

Component	Amount	Volume
U6–hsgRNA–MS2–U6–sgRNA–boxB (Step 76)	9 ng	
CMVmini (Step 76)	2 ng	
tBE–V5–mA3–bGH (Step 76)	37 ng	
Linearized pAAV–MCS vector (Step 78)	50 ng	
NEBuilder HiFi DNA assembly master mix		5 μL
Nuclease-free H_2O		Add to 10 μL
Total		10 μL

80. Keep the HiFi DNA assembly mixture 1 at 50 °C for 30 min and then transform the mixture into competent cells as described in Steps 15–20. Confirm that the sequence of constructed pAAV–TargetSite–tBE–V5–mA3 is correct by Sanger sequencing.

Construct the AAV-packaging plasmid corresponding to pEF1a–nSpCas9

81. Use a high-fidelity DNA polymerase to generate DNA fragments. Prepare PCR7 mixture to generate the DNA fragment (ITR–CMVmini) according to the following table:

Component	Volume	Final concentration
pST1374–NLS–flag-linker–Cas9 (30–50 ng/μL)	1 μL	
ITR_CMVmini_F (Table 1), 10 μM	1.5 μL	0.3 μM
CMVmini_R (Table 1), 10 μM	1.5 μL	0.3 μM
PrimeSTAR buffer (5×)	10 μL	
dNTP mix (2.5 mM)	4 μL	0.2 mM
PrimeSTAR HS DNA polymerase	1 μL	
Nuclease-free H_2O	31 μL	
Total	50 μL	

82. Perform the PCR7 reaction under the cycling conditions as described in Step 4.
83. Prepare the PCR8 mixture to generate the DNA fragment (nSpCas9–bGH) according to the following table:

Component	Volume	Final concentration
pEF1α–nSpCas9 (30–50 ng/μL)	1 μL	
ITR_SpCas9_F (Table 1), 10 μM	1.5 μL	0.3 μM
ITR_bGH_R (Table 1), 10 μM	1.5 μL	0.3 μM

Component	Volume	Final concentration
PrimeSTAR buffer (5×)	10 µL	
dNTP mix (2.5 mM)	4 µL	0.2 mM
PrimeSTAR HS DNA polymerase	1 µL	
Nuclease-free H_2O	31 µL	
Total	50 µL	

84. Perform the PCR8 reaction under the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	94 °C, 5 min		
2–36	98 °C, 15 s	68 °C, 15 s	72 °C, 3 min
37			72 °C, 3 min

85. Purify the products of PCR7 and PCR8 with 1% (wt/vol) agarose gel electrophoresis as described in Step 11.
86. Use NEBuilder HiFi DNA assembly master mix to assemble these DNA fragments of Step 85 into the linearized vector of Step 78. Prepare the HiFi DNA assembly mixture 2 according to the following table:

Component	Amount	Volume
ITR–CMVmini (Step 85)	9 ng	
nSpCas9–bGH (Step 85)	2 ng	
Linearized pAAV–MCS vector (Step 78)	50 ng	
NEBuilder HiFi DNA assembly master mix		5 µL
Nuclease-free H_2O		Add to 10 µL
Total		10 µL

87. Keep the HiFi DNA assembly mixture 2 at 50 °C for 30 min and then transform the mixture into competent cells as described in Step 15–20. Confirm that the sequence of constructed pAAV–nSpCas9 is correct by Sanger sequencing.

AAV packaging

88. Seed HEK293T/17 cells with 70% confluence in 150 mm TC-treated culture dishes with 20 mL DMEM-1 and culture the cells in a cell culture incubator at 37 °C with 5% (vol/vol) CO_2 overnight. Use 20 dishes of HEK293T/17 cells to package one type of AAV.
89. 2 h before transfection, replace the medium with 20 mL prewarmed DMEM-2.
90. Transfect HEK293T/17 cells with pAAV–DJ/8, pHelper and pAAV–nSpCas9 (Step 87). Prepare transfection mixture 4 for producing AAV–DJ/8–nSpCas9 according to the following table:

Component	Amount (per dish)	Volume
pHelper	24 µg	
pAAV–DJ/8	15.2 µg	
pEF1a–nSpCas9 (Step 87)	15.2 µg	
EZ cell transfection reagent		163.2 µL
Opti-MEM		1,000 µL

91. Transfect HEK293T/17 cells with pAAV–DJ/8, pHelper and pAAV–TargetSite–tBE–V5–mA3 (Step 80). We use the pAAV–TargetSite–tBE–V5–mA3 that contains hsgRNA–MS2 and sgRNA–boxB targeting mouse *PCSK9* (pAAV–PCSK9–tBE–V5–mA3) as the example for the following experiments. Prepare transfection mixture 5 for producing AAV–DJ/8–PCSK9–tBE–V5–mA3 according to the following table:

Component	Amount (per dish)	Volume
pHelper	24 μg	
pAAV–DJ/8	15.2 μg	
pAAV–TargetSite–tBE–V5–mA3 (Step 80)	15.1 μg	
EZ cell transfection reagent		162.9 μL
Opti-MEM		1,000 μL

92. 24 h after transfection, refresh the transfected cells with prewarmed DMEM-1.
93. After another 48 h, add 0.5 M EDTA (pH 8.0) into the cell culture medium with a 1:80 (vol/vol) ratio and mix well. Incubate the cells at room temperature for 10 min.
94. Collect both the medium and the transfected cells into 50 mL centrifuge tubes, centrifuge at 300*g* for 5 min at room temperature and collect the medium and the transfected cells separately.
▲ **CRITICAL STEP** Do not digest the cells with 0.25% (wt/vol) Trypsin–EDTA as it would reduce the AAV activity. Blow off the cells directly with the cell culture medium.
95. Concentrate and purify AAV from the cell culture medium by using AAVpro concentrator according to the manufacturer's instructions.
96. Purify AAV from the transfected cells by using AAVpro purification kit maxi (all serotypes) according to the manufacturer's instructions.
▲ **CRITICAL STEP** The AAVs are purified with AAVpro purification kit maxi in this protocol but AAVs can also be purified by iodixanol gradient ultracentrifugation[66].
97. Mix AAV obtained from the medium and the transfected cells into the same centrifuge tube.
■ **PAUSE POINT** AAV can be stored at −80 °C for a few months or in liquid nitrogen for a longer time.

AAV titration

98. Take 5 μL of the produced AAV–DJ/8–nSpCas9 and AAV–DJ/8–PCSK9–tBE–V5–mA3 separately and treat them with DNase I at 37 °C for 30 min to remove plasmid DNA according to the following table:

Component	Volume
AAV sample	5 μL
DNase I	1 μL
DNase I reaction buffer (10×)	5 μL
Nuclease-free H_2O	39 μL
Total	50 μL

99. Based on the molecular mass of pAAV–SpCas9 (or pAAV–PCSK9–tBE–V5–mA3), dilute it to a concentration of 2×10^9 molecule/μL in nuclease-free H_2O.
100. Make six serial ten-fold gradient dilutions of 2×10^9 molecule/μL pAAV–nSpCas9 (or pAAV–PCSK9–tBE–V5–mA3) in nuclease-free H_2O to create a plasmid standard curve, as follows:

Dilution Series	Sample	Nuclease-free water	Dilution factor	Total dilution
Dil. 1	10 μL DNase I treated sample	90 μL	10×	100×[a]
Dil. 2	10 μL Dil. 1	90 μL	10×	1,000×
Dil. 3	20 μL Dil. 2	80 μL	5×	5,000×
Dil. 4	20 μL Dil. 3	80 μL	5×	25,000×
Dil. 5	20 μL Dil. 4	80 μL	5×	125,000×
Dil. 6	20 μL Dil. 5	80 μL	5×	625,000×
Dil. 7	20 μL Dil. 6	80 μL	5×	3,125,000×

[a]The AAV samples have been diluted tenfold in the DNase I treatment.

101. Perform qPCR to obtain the data for creating a standard curve with the diluted plasmids. Prepare qPCR mixtures by using ChamQ Universal SYBR qPCR master mix in duplicates according to the following table:

Component	Volume	Final concentration
2 × ChamQ Universal SYBR qPCR master mix	10 μL	
ITR-Fwd (10 μM)	0.4 μL	0.2 μM
ITR-Rev (10 μM)	0.4 μL	0.2 μM
Standard curve plasmids or AAV samples	5.0 μL	
Nuclease-free H_2O	4.2 μL	
Total	20 μL	

102. Perform qPCR reaction under the following cycling conditions:

Cycle number	Denature	Anneal & Extend	Maintain
1	95 °C, 30 s		
2–41	95 °C, 10 s	60 °C, 15 s	
42	95 °C, 15 s	60 °C, 60 s	95 °C, 15 s

103. Calculate the physical titers of AAV samples. Use the qPCR data from the plasmids to create a standard curve. Substitute the Ct values of the AAV samples at different dilutions into the standard curve to calculate the physical titer of the two types of produced AAV (viral genome (vg)/mL).
◆ **TROUBLESHOOTING**

AAV injection and in vivo editing

● **TIMING 4 weeks**

AAV injection

104. Inject both AAV–DJ/8–nSpCas9 and AAV–DJ/8–PCSK9–tBE–V5–mA3 from Step 97 with a ratio of 1:1 (vg/vg) intravenously through the lateral tail vein into mice according to the following table:

AAV dose (AAV–DJ/8–nSpCas9 + AAV–DJ/8–PCSK9–tBE–V5–mA3, vg per mouse)	Volume	Analysis time after injection
$8 \times 10^{10} + 8 \times 10^{10}$	200 + 200 μL	2 and 4 weeks
$4 \times 10^{11} + 4 \times 10^{11}$	200 + 200 μL	2 and 4 weeks
$2 \times 10^{12} + 2 \times 10^{12}$	200 + 200 μL	2 and 4 weeks

▲ **CRITICAL STEP** Thaw AAV on ice (or 4 °C) before using and avoid repeated freeze–thaw cycles. We recommend using the mice with a body weight of ~20 g.

Determine tBE editing effects in vivo

105. Two and four weeks after injection, collect blood through the orbital or tail vein and separate the serum by centrifuging at 4 °C and 2,000*g* for 20 min. After centrifugation, collect serum and perform the planned assays immediately, or aliquot and store serum samples at ≤−20 °C.
▲ **CRITICAL STEP** Animals should be fasted for 4 h before blood collection. Serum samples should be aliquoted for storage and avoid repeated freeze–thaw cycles.
◆ **TROUBLESHOOTING**
106. Measure PCSK9 protein levels in the serum by using the Mouse PCSK9 ELISA kit, according to the manufacturer's instructions.
107. Measure the total low-density lipoprotein cholesterol levels in the serum by using the Total Cholesterol kit, according to the manufacturer's instructions.

108. Extract the genomic DNA and total RNA from mouse tissues. Four weeks after AAV injection, kill the mice in a chamber with a CO_2 fill rate of 30% (vol/vol) to 70% (vol/vol) of the chamber volume per min. Then dissect the killed mice by using the general surgery instrument kit for mouse at room temperature, extract 20 mg of the target and nontarget organ samples (e.g., liver and kidney) and place these tissues in a frozen tissue grinder to grind the samples into powder. Extract genomic DNA and total RNA using E.Z.N.A. Tissue DNA kit and HiPure Total RNA plus mini kit, respectively, following the manufacturer's instructions.
◆ **TROUBLESHOOTING**
109. Determine on-target editing efficiency in vivo. Subject the genomic DNA obtained in Step 108 for amplicon sequencing to determine the editing efficiency at the on-target site. Refer to Steps 40–45 for more details.
110. Determine gRNA-independent genome-wide OT mutations in vivo. Prepare WGS samples of mouse tissues (e.g., liver and kidney) and then analyze WGS data. Refer to Step 64 for more details.
111. Determine gRNA-independent transcriptome-wide OT mutations in vivo. Prepare RNA-seq samples of mouse tissues (e.g., the livers of AAV-injected and uninjected mice) and then analyze the RNA-seq data. Refer to Steps 68–70 for more details.

Data analysis

● **TIMING 1 week**

Analyze base substitution frequency

112. Select base substitutions at each position of the examined on-target sites of tBE (Steps 40–45), gRNA-dependent OT sites (Steps 46–47) and gRNA-independent DNA OT sites (nSaCas9 on-target site, Steps 48–57) with at least 1,000 independent reads mapped. Calculate base substitution frequencies by dividing base substitution reads by total reads using the CFBI pipeline (https://github.com/YangLab/CFBI; v.1.0.0).
▲ **CRITICAL STEP** The base editing efficiency in this protocol is calculated with CFBI but it can also be calculated with alternative tools (e.g., BE-Analyzer[67] and CRISPResso2 (ref. 68)).

Determine gRNA-independent genome-wide OT mutations with BEIDOU

▲ **CRITICAL STEP** Determine genome-wide OT mutations from WGS data by the computational pipeline BEIDOU (https://github.com/YangLab/BEIDOU). The workflow of BEIDOU is as follows.

113. DNA-sequencing reads mapping. After sequencing by the Illumina Hiseq X Ten system, clean and map raw reads saved in FASTQ format to reference genome. In brief, perform quality control by FastQC (v.0.11.5, parameters: default) and trim DNA-sequencing reads by Trimmomatic (v.0.38; parameters: ILLUMINACLIP:TruSeq3-PE-2.fa: 2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36)[69] and remove the reads with low quality. Map clean reads to the reference genome using the BWA-MEM algorithm (v.0.7.17-r1188; default parameters). Select reads with a mapping quality score ≥30 and sort BAM files by samtools (v.1.9; parameters: -bh -F 4 -q 30). After marking duplicate reads identified by Picard (v.2.21.2; parameters: REMOVE_DUPLICATES = false) in the BAM file, correct systematic bias using a two-stage process by GATK (BaseRecalibrator and ApplyBQSR; default parameters). The output file is composed of the aligned results in BAM format.
114. Variants calling and filtering. To identify genome-wide de novo variants with high confidence, conduct SNV calling using three algorithms with the germline calling methods, GATK (v.4.1.3), LoFreq (v.2.1.3.1, default parameters) and Strelka2 (v.2.9.10; default parameters) and conduct indel calling by using GATK (v.4.1.3), Strelka2 (v.2.9.10; default parameters) and Scalpel (v.0.5.4; parameters: -single -window 600). For GATK, determine genome-wide de novo variants by three GATK commands: HaplotypeCaller (parameters: default), VariantRecalibrator (parameters: '- resource: hapmap,

known = false, training = true, truth = true, prior = 15.0 hapmap_3.3.hg38.vcf.gz -resource: omni, known = false, training = true, truth = false, prior = 12.0 1000G_omni2.5.hg38.vcf.gz -resource:1000G, known = false, training = true, truth = false, prior = 10.0 1000G_phase1.snps.high_confidence.hg38.vcf.gz -resource: dbsnp, known = true, training = false, truth = false, prior = 2.0 dbsnp_146.hg 38.vcf.gz -an QD -an MQ -an MQRankSum -an ReadPosRankSum -an FS -an SOR -an DP -max-gaussians 4' for SNVs; '-resource: mills, known = true, training = true, truth = true, prior = 12.0 Mills_and_1000G_gold_standard.ind els.hg38.vcf.gz -an QD -an MQRankSum -an ReadPosRankSum -an FS -an SOR -an DP -max-gaussians 4 - mode INDEL' for indels) and ApplyVQSR (parameters: '-mode SNP -ts-filter-level 95' for SNVs; '-mode INDEL -ts-filter-level 95' for indels). Of note, consider the overlaps of SNVs/indels called by these three algorithms as reliable variants by the BEIDOU toolkit. Save the information of variants in VCF file format. Furthermore, to obtain de novo SNVs/indels, filter out the background variants, including: (1) SNVs/ indels with allele frequencies <10% or depth fewer than ten reads, (2) SNVs/indels in nontransfected cells of this study or dbSNP (v.151; http://www.ncbi.nlm.nih.gov/SNP/) database and (3) SNVs/indels overlapped with the UCSC repeat regions. Focus analyses only on SNVs/indels from canonical (chr 1-22, X, Y and M) chromosomes. The output of BEIDOU is novel SNVs/indels information in TXT file format.

Determine gRNA-independent transcriptome-wide OT mutations with RADAR

▲ **CRITICAL** Detect and visualize all possible 12 types of RNA-editing event from RNA-seq datasets with the computational pipeline RNA-editing analysis-pipeline to decode all 12 types of RNA-editing event (RADAR; https://github.com/YangLab/RADAR). The workflow of RADAR is as follows.

115. RNA-seq reads mapping. After sequencing by the Illumina Hiseq X Ten system, clean and map raw reads saved in FASTQ format to reference genome by two-step alignment. In brief, perform quality control by FastQC (v.0.11.5, parameters: default), trim sequencing reads by Trimmomatic (v.0.38, parameters: TruSeq3-PE-2.fa:2:30:10 TRAILING:25 MINLEN:30) to remove reads with low quality. Then, map clean reads to ribosomal DNA (rDNA) sequences by the BWA-MEM algorithm (v.0.7.17a, parameters: default) to remove the reads mapped to redundant rRNAs. To capture more candidate RNA-editing events, apply a two-round unique mapping strategy to align reads to the reference genome, sequentially by HISAT2 (v.2.1.0, parameters: -rna-strandness RF -no-mixed -secondary -no-temp-splicesite -known-splicesite-infile -no-softclip -score-min L,-16,0 -mp 7,7 -rfg 0,7 -rdg 0,7 -max-seeds 20 -k 10 -dta) and by BWA-MEM (v.0.7.17a, parameters: default). Select and combine unique mapped reads by HISAT2 and BWA-MEM with up to six mismatches for subsequent analysis. After marking duplicate reads identified by Picard (v.2.21.2, parameters: CREATE_INDEX = true VALIDATION_STRINGENCY = SILENT) in the BAM file, split uniquely-mapped reads that span exon–exon junctions into segments by the GATK (v.4.1.2.0) command: SplitNCigarReads (parameters: default). Recalibrate base quality scores of all uniquely mapped reads by two GATK (v.4.1.3) commands: BaseRecalibrator (parameters: default) and ApplyBQSR (parameters: default). The output file is composed of the aligned results in BAM format.
116. Variant calling and classification. Determine variants at RNA level from the BAM file by the GATK (v.4.1.3) command HaplotypeCaller (parameters: -minimum-mapping-quality 0 -stand_call_conf 0 -don't-use-soft-clipped-bases true). After filtering out RNA variants overlapped with single nucleotide polymorphisms (SNPs) from dbSNP version 151 (https://www.ncbi.nlm.nih.gov/SNP/), the 1000 Genomes Project (https://www.internationalgenome.org/) and the University of Washington Exome Sequencing Project (https://evs.gs.washington.edu/EVS/), remove the low-quality RNA variants with mapped read numbers <2, hits per billion-mapped-bases (HPB) <3 or editing ratio <0.05. Classify RNA variants into three groups according to their genomic locations as previously described[70], including in Alu, non-Alu repetitive or nonrepetitive regions. Filter RNA variants in non-Alu repetitive and nonrepetitive regions with a series of stringent cutoffs to reduce false positive. In brief, remove RNA variants in simple repeats, in mononucleotide

microsatellites ≥5 bp or within 4 bp of splice junctions. In addition, remove RNA variants within highly similar regions using BLAST-like alignment tool (BLAT, v.364, parameters: -repMatch = 2253 -stepSize = 5)[71]. Finally, remove RNA variants within bidirectional transcription regions. Eventually determine all 12 types of RNA-editing event in Alu, non-Alu repetitive and nonrepetitive regions according to the strands of overlapped genes (Human: hg38 knownGene.txt updated at 28 June 2015).

117. Results output. List all possible RNA-editing events from each given RNA-seq dataset in an Excel file. Plot the numbers of all 12 types of RNA-editing event by histograms according to their genomic locations in Alu, repetitive non-Alu and nonrepetitive regions. Illustrate RNA-editing ratios of selected types of RNA-editing event, such as C-to-U or A-to-G by manhattan plots.

Troubleshooting

Troubleshooting advice can be found in Table 2.

Table 2 | Troubleshooting table

Step	Problem	Possible reason	Solution
19	No clones grown on LB agar plates	Inappropriate amount of DNA fragment or linearized vector	Remeasure the concentrations of the DNA fragment and the linearized vector, reprepare them if necessary
31	Cells are less than or greater than 60–70% confluence	Cell density is not precisely determined when the cells are seeded	Determine the cell density precisely and re-seed the cells
45	Editing efficiency is low or no editing	SNVs at the target site result in mismatches between the spacer region of designed sgRNA or hsgRNA and the actual genomic DNA	Confirm the actual genomic DNA sequence of the target site and re-design the corresponding sgRNA or hsgRNA
45	Editing efficiency is low or indel frequency is high	The hsgRNA is designed to target the DNA strand opposite to the one that sgRNA targets	Redesign the hsgRNA to ensure that it targets the strand that sgRNA targets
47	No gRNA-dependent DNA OT mutations are induced by the positive control BE	The in silico predicted OT sites are not the authentic OT sites	Determine the mutations at the potential gRNA-dependent OT sited using alternative prediction programs
57	No gRNA-independent DNA OT mutations are induced by the positive control BE	The R-loop region generated by nSaCas9 and Sa–sgRNA is not bound by the positive control BE	Reperform the CESSCO assay by using different Sa–sgRNAs
67	Sorted cells are too few	Impure plasmids that result in low transfection efficiency	Reprepare the plasmid DNA to improve its purity
68,108	The concentration of total RNA is too low	RNA is degraded	Extract RNA in an RNase-free environment
103	The titer of AAV is too low	The length of cargo gene is close to the upper limit of AAV packaging capability	Increase the number of cells used to package AAV, or use a 'cell factory' system consisting of multiple large culture dishes stacked on top of each other to package AAV
105	Hemolysis occurs during blood collection	The samples are drawn or centrifuged too violently so that the red blood cells break	Gently draw during the collection of blood samples and also gently centrifuge the blood samples at 4 °C after collection

Timing

Steps 1–21, construction of the plasmids expressing hsgRNA–MS2 and sgRNA–boxB for tBE: 4–5 d
Steps 22–57, transfection of tBE plasmids and determination of the on-target editing efficiency and the OT mutations at gRNA-dependent OT sites and gRNA-independent OT sites (with CESSCO) in cell lines: 2 weeks
Steps 58–70, determination of gRNA-independent genome-wide and transcriptome-wide OT mutations: 5 weeks (can be performed parallelly with the following steps)
Steps 71–103, production of tBE-containing AAV: 1 week
Steps 104–111, AAV injection and in vivo editing sample preparation: 4 weeks
Steps 112–117, analysis of editing effects: 1 week

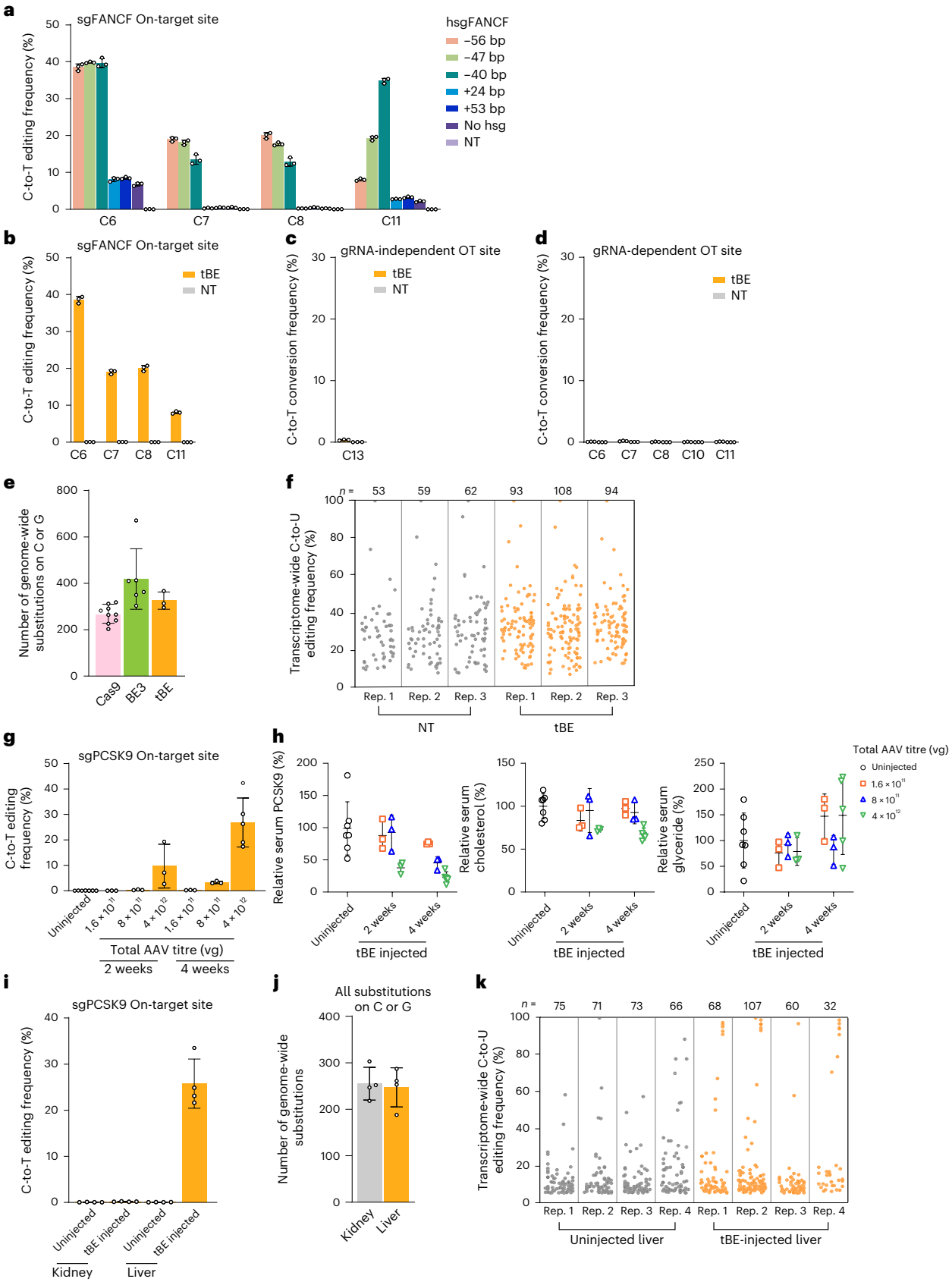

a
sgFANCF On-target site
C-to-T editing frequency (%)
hsgFANCF
−56 bp
−47 bp
−40 bp
+24 bp
+53 bp
No hsg
NT
C6
C7
C8
C11
b
sgFANCF On-target site
tBE
NT
c
gRNA-independent OT site
C13
d
gRNA-dependent OT site
C10
e
Number of genome-wide substitutions on C or G
Cas9
BE3
tBE
f
Transcriptome-wide C-to-U editing frequency (%)
Rep. 1
Rep. 2
Rep. 3
NT
tBE
g
sgPCSK9 On-target site
C-to-T editing frequency (%)
Uninjected
Total AAV titre (vg)
2 weeks
4 weeks
h
Relative serum PCSK9 (%)
Relative serum cholesterol (%)
Relative serum glyceride (%)
tBE injected
i
sgPCSK9 On-target site
Kidney
Liver
j
All substitutions on C or G
k
Uninjected liver
tBE-injected liver

Fig. 5 | Results of tBE-mediated editing in cells and mice. a, Comparison of the editing efficiencies induced by sgFANCF paired with different hsgFANCFs. The on-target C-to-T base editing frequencies induced by the tBE and sgFANCF paired with the indicated hsgFANCFs located at different positions relative to sgFANCF are shown. C6, C7, C8 and C11 represent the cytosines in the protospacer region for sgFANCF, counting the PAM distal position as 1. The pair of sgFANCF–hsgFANCF_-56_bp, which induced the highest editing efficiency at the target cytosine (C7), was selected for the subsequent analyses. NT, nontransfected cells. **b**, Determination of on-target editing efficiency. The C-to-T editing frequencies at the on-target site in the 293FT cells transfected with the tBE system containing sgFANCF–hsgFANCF_-56_bp are shown. C6, C7, C8 and C11 represent the cytosines in the protospacer region for sgFANCF, counting the PAM distal position as 1. **c**, Determination of gRNA-independent DNA OT mutations. The C-to-T conversion frequencies at the gRNA-independent OT site generated by a representative Sa–sgRNA and nSaCas9 in the 293FT cells transfected with the tBE system containing sgFANCF–hsgFANCF_-56_bp are shown. C13 represents the cytosine in the protospacer region for the Sa–sgRNA, counting the PAM distal position as 1. **d**, Determination of gRNA-dependent DNA OT mutations. The C-to-T conversion frequencies at a representative gRNA-dependent OT site predicted by Cas-OFFinder in the 293FT cells transfected with the tBE system containing sgFANCF–hsgFANCF_-56_bp are shown. C6, C7, C8 and C10 represent the cytosines in the mismatched protospacer region for sgFANCF, counting the PAM distal position as 1. For **a**–**d**, data are mean ± s.d. from three independent experiments. **e**, Determination of gRNA-independent genome-wide OT mutations. The number of gRNA-independent genome-wide substitutions of C or G in the *APOBEC3*-KO single-cell clones edited with Cas9, BE3 or tBE are shown. Data are mean ± s.d. from $n = 9$ (Cas9), $n = 6$ (BE3) and $n = 3$ (tBE) single-cell colonies. Rep., replicate. **f**, Determination of gRNA-independent transcriptome-wide OT mutations. Manhattan plots of gRNA-independent transcriptome-wide C-to-U mutations in the 293FT cells transfected with the tBE are shown. **g**, Determination of in vivo editing efficiencies induced by tBE through the delivery of AAVs with different titers and expression time periods. The C-to-T editing frequencies at the *PCSK9* on-target site in the livers of mice injected with AAV–DJ/8–nSpCas9 and AAV–DJ/8–PCSK9–tBE–V5–mA3 are shown. **h**, Determination of in vivo editing effects. The levels of serum PCSK9 protein, cholesterol and glyceride in uninjected mice and the mice injected with AAV–DJ/8–nSpCas9 and AAV–DJ/8–PCSK9–tBE–V5–mA3 are shown. For **g** and **h**, data are mean ± s.d. from $n = 7$ (uninjected), $n = 3$ (1.6×10^{11} vg, 2 weeks), $n = 3$ (8×10^{11} vg, 2 weeks), $n = 3$ (4×10^{12} vg, 2 weeks), $n = 3$ (1.6×10^{11} vg, 4 weeks), $n = 3$ (8×10^{11} vg, 4 weeks) or $n = 5$ (4×10^{12} vg, 4 weeks) mice. **i**, Comparison of in vivo editing efficiencies in the AAV-nontarget and AAV-target organs. The C-to-T editing frequencies in the nontarget (kidney) and target (liver) organs of the mice injected with injected with AAV–DJ/8–nSpCas9 and AAV–DJ/8–PCSK9–tBE–V5–mA3 are shown. **j**, Determination of gRNA-independent genome-wide OT mutations in vivo. The number of gRNA-independent genome-wide substitutions of C or G in the nontarget (kidney) and target (liver) organs of mice injected with AAV–DJ/8–nSpCas9 and AAV–DJ/8–PCSK9–tBE–V5–mA3 are shown. For **i** and **j**, data are mean ± s.d. from $n = 4$ mice. **k**, Determination of gRNA-independent transcriptome-wide OT mutations in vivo. Manhattan plots of gRNA-independent transcriptome-wide C-to-U mutations in the livers of mice injected with injected with AAV–DJ/8–nSpCas9 and AAV–DJ/8–PCSK9–tBE–V5–mA3 are shown. For **g**–**k**, the data are from the animal experiments that were approved by the Animal Care and Ethical Committee at Wuhan University. Figure adapted from ref. 27, Springer Nature Limited.

Anticipated results

Previously, we have used tBE to induce base editing in mammalian cells and mice and found that the evaluation of several sgRNA–hsgRNA pairs can allow efficient tBE-induced C-to-T editing with no observable OT mutations[27]. Here, we select some results from our previous study[27], including screening sgRNA–hsgRNA pairs (Fig. 5a, the pair of sgFANCF and hsgFANCF 47 bp is selected for Fig. 5b–d), the on-target editing efficiency induced by the selected sgRNA–hsgRNA pair (Fig. 5b), the gRNA-independent OT DNA mutations determined by CESSCO (Fig. 5c) and the gRNA-dependent OT DNA mutations at in silico-predicted OT sites (Fig. 5d). By sequencing the genome of the edited single-cell clones and the transcriptome of the edited cell population, gRNA-independent genome-wide OT mutations (Fig. 5e) and gRNA-independent transcriptome-wide OT mutations (Fig. 5f) can also be determined.

Also, through packaging the tBE into a dual-AAV system, efficient in vivo C-to-T base editing (Fig. 5g) is induced and the efficacy of in vivo editing can be assessed (Fig. 5h). By comparing the editing frequencies (Fig. 5i) and genome-wide SNVs (Fig. 5j) of the target and nontarget organs or tissues (e.g., liver and kidney) of the edited mice, the in vivo editing specificity in the genome can be determined. Finally, by comparing transcriptome-wide SNVs (Fig. 5k) of the target organs or tissues between the AAV-injected and uninjected animals, the in vivo editing specificity in the transcriptome can be determined.

Data availability

The data used to generate the example results shown in Fig. 5 were originally published in ref. 27. All sequencing datasets have been deposited in the Gene Expression Omnibus under the accession code GSE164837 and GSE164477, at the NCBI BioProject under the accession code PRJNA692761 and in the National Omics Data Encyclopedia under the accession codes OEP001688, OEP001689 and OEP001690. All other data supporting the finding of this study are available from the corresponding authors on reasonable requests. Source data are provided with this paper.

Code availability

The custom Perl and Shell scripts for CFBI are available at GitHub (https://github.com/YangLab/CFBI). The computational pipeline of BEIDOU to identify high-confidence base substitution or indel events from WGS data is available at GitHub (https://github.com/YangLab/BEIDOU). The workflow of RADAR to detect and visualize all 12 possible types of RNA-editing event from RNA-seq data is available at GitHub (https://github.com/YangLab/RADAR).

Received: 17 December 2022; Accepted: 21 June 2023;
Published online: 04 October 2023

References

1. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
2. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
3. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
4. Yang, B., Yang, L. & Chen, J. Development and application of base editors. *CRISPR J.* **2**, 91–104 (2019).
5. 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).
6. Yang, L. & Chen, J. A tale of two moieties: rapidly evolving CRISPR/Cas-based genome editing. *Trends Biochem. Sci.* **45**, 874–888 (2020).
7. Porto, E. M., Komor, A. C., Slaymaker, I. M. & Yeo, G. W. Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov.* **19**, 839–859 (2020).
8. Nambiar, T. S., Baudrier, L., Billon, P. & Ciccia, A. CRISPR-based genome editing through the lens of DNA repair. *Mol. Cell* **82**, 348–388 (2022).
9. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
10. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
11. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
12. Grunewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
13. Yang, B., Li, X., Lei, L. & Chen, J. APOBEC: from mutator to editor. *J. Genet. Genomics* **44**, 423–437 (2017).
14. Lei, L. et al. APOBEC3 induces mutations during repair of CRISPR–Cas9-generated DNA breaks. *Nat. Struct. Mol. Biol.* **25**, 45–52 (2018).
15. Olson, M. E., Harris, R. S. & Harki, D. A. APOBEC enzymes as targets for virus and cancer therapy. *Cell Chem. Biol.* **25**, 36–49 (2018).
16. Chen, J., Yang, B. & Yang, L. To BE or not to BE, that is the question. *Nat. Biotechnol.* **37**, 520–522 (2019).
17. Rees, H. A. et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
18. Wang, X. et al. Efficient base editing in methylated regions with a human APOBEC3A–Cas9 fusion. *Nat. Biotechnol.* **36**, 946–949 (2018).
19. Billon, P. et al. CRISPR-mediated base editing enables efficient disruption of Eukaryotic genes through induction of STOP codons. *Mol. Cell* **67**, 1068–1079 e1064 (2017).
20. Kuscu, C. et al. CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
21. Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
22. Diorio, C. et al. Cytosine base editing enables quadruple-edited allogeneic CART cells for T-ALL. *Blood* **140**, 619–629 (2022).
23. Wang, L. et al. Reactivation of gamma-globin expression through Cas9 or base editor to treat beta-hemoglobinopathies. *Cell Res.* **30**, 276–278 (2020).
24. Zeng, J. et al. Therapeutic base editing of human hematopoietic stem cells. *Nat. Med.* **26**, 535–541 (2020).
25. Liang, P. et al. Correction of beta-thalassemia mutant by base editor in human embryos. *Protein Cell* **8**, 811–822 (2017).
26. Koblan, L. W. et al. In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. *Nature* **589**, 608–614 (2021).
27. Wang, L. et al. Eliminating base-editor-induced genome-wide and transcriptome-wide off-target mutations. *Nat. Cell Biol.* **23**, 552–563 (2021).
28. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
29. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR–Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
30. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).
31. Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
32. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
33. Wang, L. et al. Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res.* **27**, 1289–1292 (2017).
34. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
35. Li, X. et al. Base editing with a Cpf1-cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).
36. Wang, X. et al. Cas12a base editors induce efficient and specific editing with low DNA damage response. *Cell Rep.* **31**, 107723 (2020).
37. Zhang, X. et al. Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain. *Nat. Cell Biol.* **22**, 740–750 (2020).
38. Levy, J. M. et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat. Biomed. Eng.* **4**, 97–110 (2020).
39. Chadwick, A. C., Wang, X. & Musunuru, K. In vivo base editing of PCSK9 (proprotein convertase subtilisin/kexin type 9) as a therapeutic alternative to genome editing. *Arterioscler. Thromb. Vasc. Biol.* **37**, 1741–1747 (2017).
40. Wu, Y. et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat. Med.* **25**, 776–783 (2019).
41. Kunkel, T. A. & Erie, D. A. Eukaryotic mismatch repair in relation to DNA replication. *Annu. Rev. Genet.* **49**, 291–313 (2015).
42. Liu, Z. et al. Highly efficient RNA-guided base editing in rabbit. *Nat. Commun.* **9**, 2717 (2018).
43. Xie, J. et al. Efficient base editing for multiple genes and loci in pigs using base editors. *Nat. Commun.* **10**, 2852 (2019).
44. Li, G. et al. Base pair editing in goat: nonsense codon introgression into FGF5 results in longer hair. *FEBS J.* **286**, 4675–4692 (2019).
45. Zhang, Y. et al. Programmable base editing of zebrafish genome using a modified CRISPR–Cas9 system. *Nat. Commun.* **8**, 118 (2017).
46. Zuo, E. et al. A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat. Methods* **17**, 600–604 (2020).
47. Jin, S. et al. Rationally designed APOBEC3B cytosine base editors with improved specificity. *Mol. Cell* **79**, 728–740 e726 (2020).
48. Lee, S. et al. Single C-to-T substitution using engineered APOBEC3G-nCas9 base editors with minimum genome- and transcriptome-wide off-target effects. *Sci. Adv.* **6**, eaba1773 (2020).
49. Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
50. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
51. Chen, J. S. et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
52. Nishimasu, H. et al. Engineered CRISPR–Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
53. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR–Cas9 variants. *Science* **368**, 290–296 (2020).

54. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
55. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
56. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
57. Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
58. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
59. 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* **30**, 1473–1475 (2014).
60. Gao, R. et al. Genomic and transcriptomic analyses of prime editing guide RNA-independent off-target effects by prime editors. *CRISPR J.* **5**, 276–293 (2022).
61. Li, J. et al. Efficient base editing in G/C-rich regions to model androgen insensitivity syndrome. *Cell Res.* **29**, 174–176 (2019).
62. Villiger, L. et al. In vivo cytidine base editing of hepatocytes without detectable off-target mutations in RNA and DNA. *Nat. Biomed. Eng.* **5**, 179–189 (2021).
63. Concordet, J. P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).
64. Listgarten, J. et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat. Biomed. Eng.* **2**, 38–47 (2018).
65. Cancellieri, S. et al. Human genetic diversity alters off-target outcomes of therapeutic gene editing. *Nat. Genet.* **55**, 34–43 (2023).
66. Sena-Esteves, M. & Gao, G. Purification of recombinant adeno-associated viruses (rAAVs) by iodixanol gradient centrifugation. *Cold Spring Harb. Protoc.* **2020**, 095612 (2020).
67. Hwang, G. H. et al. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinform.* **19**, 542 (2018).
68. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
69. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
70. Ramaswami, G. et al. Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* **10**, 128–132 (2013).
71. Kent, W J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).

Acknowledgements

This work was supported by 2018YFA0801401 (J.C.) and 2019YFA0802804 (L.Y.) from the National Key R&D Program of China, 81872305 (J.C.) and 31925011 (L.Y.) from the National Natural Science Foundation of China, NK2022010207 (J.C.) from Ministry of Agriculture and Rural Affairs, 21JC1404600 (J.C.), 20PJ1410200 (J.L.) and 23XD1422500 (J.C.) from Shanghai Municipal Science and Technology Commission. We thank Shanghai Frontiers Science Center for Biomacromolecules and Precision Medicine, ShanghaiTech University, Shanghai Clinical Research and Trial Center, Animal Core Facility, ShanghaiTech University and Molecular and Cell Biology Core Facility, School of Life Science and Technology, ShanghaiTech University for providing support.

Author contributions

J.C., L.Y. and J.L. conceived, designed and supervised the project. W.H., B.-Q.G., J.Z., Z.H., J.L., L.Y. and J.C. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41596-023-00877-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-023-00877-w.

Correspondence and requests for materials should be addressed to Jianfeng Li, Li Yang or Jia Chen.

Peer review information *Nature Protocols* thanks Yongsub Kim, Liangxue Lai, Zhanjun Li, Shaohua Yao, Erwei Zuo and the other, anonymous, reviwer(s) for their contribution to the peer review of this work.

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

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

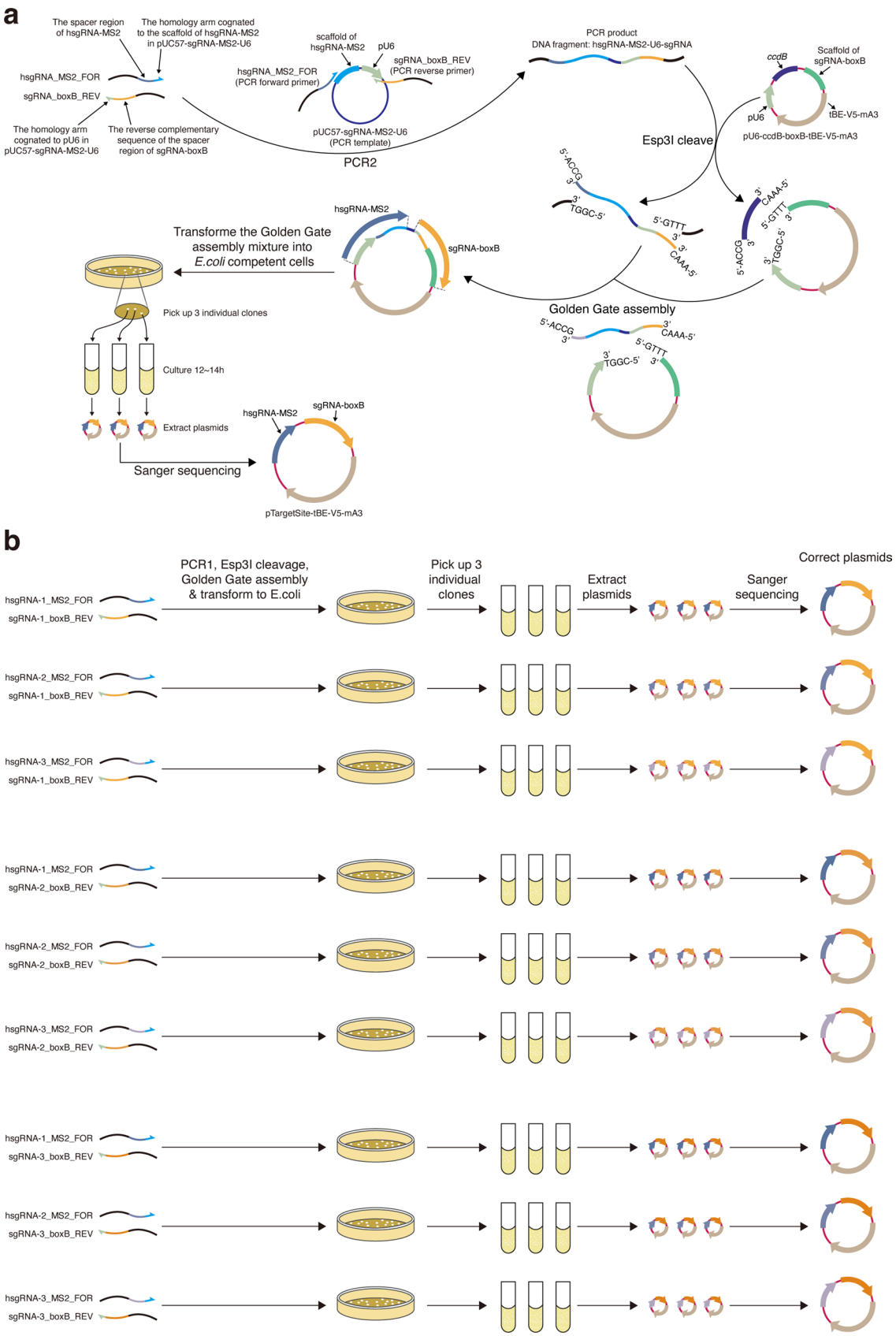
Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Related links

Key reference using this protocol

Wang, L. et al. *Nat. Cell Biol.* **23**, 552-563 (2021): https://doi.org/10.1038/s41556-021-00671-4

© Springer Nature Limited 2023



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Procedure to construct the designed plasmids each of which expresses a pair of hsgRNA-MS2 and sgRNA-boxB. a, The schematic diagram demonstrating the processes of constructing one plasmid expressing a pair of designed hsgRNA-MS2 and sgRNA-boxB. **b**, The schematic diagram demonstrating the construction of plasmids designed for the screening of all combinations of sgRNAs and hsgRNAs against one on-target site.