

CABE-RY: A PAM-flexible dual-mutation base editor for reliable modeling of multi-nucleotide variants

Wanyu Tao,^{1,2,6} Qing Liu,^{3,6} Shisheng Huang,^{1,2} Xin Wang,^{1,2} Shiyuan Qu,^{1,2} Junfan Guo,^{1,2} Danfeng Ou,⁵ Guanglei Li,^{1,2} Yu Zhang,^{1,2} Xiangmin Xu,^{3,4} and Xingxu Huang^{1,2}

¹School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China; ²University of Chinese Academy of Sciences, Beijing 100049, China; ³Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong 510515, China; ⁴Guangdong Genetic Testing Engineering Research Center, Guangzhou, Guangdong 510515, China; ⁵International Academy of Optoelectronics at Zhaoqing, South China Normal University, Zhaoqing 526238, China

Multi-nucleotide variants (MNVs) represent an important type of genetic variation and have biological and clinical significance. To simulate MNVs, we designed four dual-mutation base editors combining hA3A(Y130F), TadA8e(V106W), and protospacer adjacent motif (PAM)-flexible SpRY and selected cytosine and adenine base editor-SpRY (CABE-RY), which had the best editing performance, for further study. Characterization and comparison showed that CABE-RY had a smaller DNA editing window and lower RNA off-target edits than the corresponding single base editors. Thus, we have established a versatile tool to efficiently simulate MNVs over the genome, which could be very useful for functional studies on MNVs in humans.

INTRODUCTION

Sequencing technologies have rapidly advanced our understanding of human genetic variants. Many disease-associated variants have been identified in patients. Single-nucleotide variants (SNVs), which represent one type of variant, are usually included in genetic evaluations with phenotypic information.¹ Other kinds of genetic variants may also have influences on diseases. Therefore, precisely elucidating how SNVs contribute to causal relationships between genotypes and phenotypes remains a major challenge.

Multi-nucleotide variants (MNVs) represent another variant type that is considered to be related to disease. For example, Kaplanis et al.² found significant enrichment of MNVs in genes associated with diagnosing developmental disorders (DDD), showing that MNVs are on average more harmful than SNVs. Usually, there are two or more nearby variants on the same haplotype in an individual.^{2,3} When nearby variants are within the same codon, the amino acid changes are different than if the separate SNVs are annotated independently. Due to neglect of MNVs, most existing variant callers mistake MNVs for SNVs with incorrect amino acid change predictions,^{4,5} which probably hampers scientific research and clinical practice.

Recently, the widespread application of CRISPR-based gene-editing technologies has revolutionized genetics and disease research. CRISPR-based base editors, which are derived by fusing deoxynucleoside deaminase to the nickase Cas9 and are recruited by a guide RNA to their target DNA region, efficiently make C-to-T or A-to-G nucleotide changes, improving simulation of SNVs to enable exploration of the relationship between genotype and phenotype. To further expand the editing capabilities of base editors, dual-mutation base editors, which combine cytosine base editors (CBEs) and adenine base editors (ABEs), have been successfully developed.^{6–10} Theoretically, dual-mutation base editors make simultaneous A/C conversions, the most frequent adjacent dinucleotide MNVs,² more conveniently than single-mutation base editors. However, existing dual-mutation base editors have narrow application ranges due to protospacer adjacent motif (PAM) restriction. Although prime editors (PEs) can theoretically simulate most variants, including MNVs,¹¹ the efficiency of PEs is low, especially *in vivo*.¹² Therefore, better dual-mutation base editors that can simulate MNVs need to be developed.

SpRY is a mutant of SpCas9 that is highly PAM-compatible.¹³ The near-PAMless SpCas9 shows powerful advantages in PAM recognition. However, although it exhibits robust activity on sites with NRN PAMs (where R is A or G), it exhibits low activity on those with NYN PAMs (where Y is C or T). Considering that one of the limitations of using existing dual-mutation base editors is that they require target recognition of the NGG motif in the genome, in

Received 17 January 2021; accepted 17 July 2021;
<https://doi.org/10.1016/j.omtn.2021.07.016>.

⁶These authors contributed equally

Correspondence: Xiangmin Xu, Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong 510515, China.

E-mail: xixm@smu.edu.cn

Correspondence: Xingxu Huang, School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China.

E-mail: huangxx@shanghaitech.edu.cn



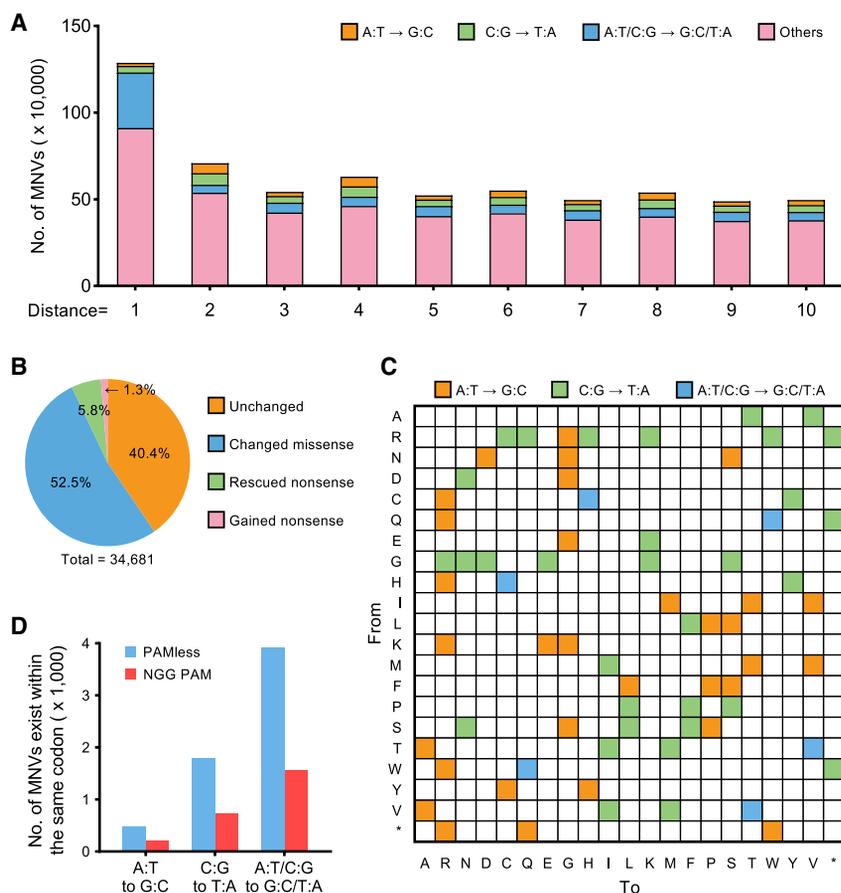


Figure 1. Necessity of a dual-mutation base editor with minimal PAM restriction for simulating MNVs

(A) Substitution pattern of MNVs from gnomAD. The x coordinate is the distance between two SNVs. (B) Differences in functional impacts on the protein between MNVs and individual SNPs. (C) Heatmap showing the amino acid changes that can be simulated by the dual-mutation base editor. (D) The PAM-flexible dual-mutation base editor has a broadened scope to target MNVs within the same codon. A broadened editing window (positions 3–10) was used to calculate the targeting scope.

of MNVs rescued nonsense mutations caused by the individual mutations, and 1.3% of MNVs produced nonsense mutations that were not caused by the individual mutations (Figure 1B). This difference may affect understanding of functional research or clinical diagnosis of genetic diseases. The dual-mutation base editor allows us to better understand the functional impacts of MNVs by simulating these MNVs.

Ideally, a dual-mutation base editor should be able to effectively simulate 70 amino acid substitutions (Figure 1C), including six that cannot be achieved by existing CBEs and ABEs (Figure S1B). However, the previously reported dual-mutation base editors can only target sites with NGG PAMs, which greatly limits their application. Therefore, we speculated that the

this study, we used SpRY to construct a dual-mutation base editor with minimal PAM restriction. To this end, we fused hA3A(Y130F), which has efficient C-to-T editing, including in G/C rich regions,¹⁴ and TadA8e(V106W), which has the highest efficiency of A-to-G editing among editors developed to date,¹⁵ to the N terminus of SpRY to produce a dual-mutation base editor named cytosine and adenine base editor-SpRY (CABE-RY). With this tool, we effectively performed simultaneous A/C conversion and successfully simulated MNVs.

RESULTS

Necessity of a dual-mutation base editor with minimal PAM restriction for simulating MNVs

An increasing number of MNVs have been discovered and deposited in the Genome Aggregation Database (gnomAD).¹⁶ To better understand the biology of MNVs, we analyzed the MNV mutation type in gnomAD. The results showed that there was more simultaneous A/C conversion than AA or CC conversion, particularly in two adjacent mutations (Figure 1A). When two nearby mutations are located in the same codon, they may have a different functional impact on the protein than the individual mutations (Figure S1A). Indeed, among a total of 31,575 MNVs, 52.5% of them resulted in different missense mutations from those caused by the individual mutations, while 5.8%

near-PAMless engineered CRISPR-Cas9 variant SpRY could be used to construct a PAM-flexible dual-mutation base editor without strict PAM restriction. We analyzed the editing scopes between a conventional NGG PAM editor and the PAM-flexible editor. Considering that dual-mutation editors usually have a broader editing window than conventional editors,⁹ we calculated the editable MNVs with an editing window spanning positions 3–10 (versus the common window spanning positions 4–8), which presumably increased the number of targetable MNVs. As expected, the PAM-flexible dual-mutation base editor was found to target ~2.7 times the number of MNVs and specific codons than the conventional NGG PAM editor (Figure 1D; Figure S1C), suggesting that the PAM-flexible dual-mutation base editor is a potential tool with an expansive editing scope for modeling MNVs.

Establishment of a PAM-flexible dual-mutation base editor, CABE-RY

To create a PAM-flexible dual-mutation base editor, we combined SpRY with cytosine deaminase and adenine deaminase from the previously described hA3A(Y130F), which has efficient C-to-T editing, including in G/C rich regions,¹⁴ and TadA8e(V106W), which has the highest efficiency of A-to-G editing among editors developed thus far.¹⁵ Four constructs were generated based on

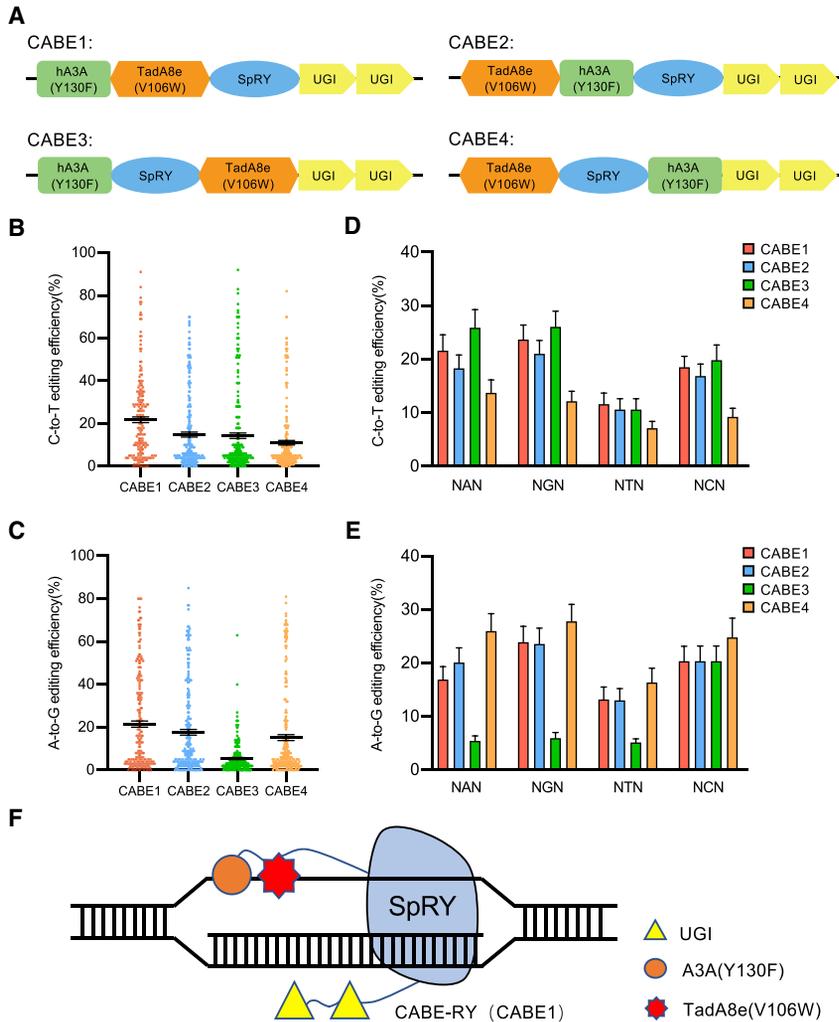


Figure 2. Establishment of a PAM-flexible dual-mutation base editor, CABB-RY, in HEK293T cells

(A) Schematic diagram for construction of four dual-mutation base editors, CABB-1, CABB-2, CABB-3, and CABB-4. (B and C) Comparison of the C-to-T (B) and A-to-G (C) editing efficiencies of the four CABB editors. The horizontal black lines represent the mean editing efficiencies at 48 endogenous sites for each CABB editor. Each dot represents the editing efficiency of each edited base ($n = 3$ independent replicates). The error bars represent the standard error of the mean (SEM) values. (D and E) Mean C-to-T (D) and A-to-G (E) editing efficiency plots for the four CABB editors with different PAMs. The data are presented for each edited base at 48 endogenous sites ($n = 3$ independent replicates). The error bars represent the SEM values. (F) Schematic overview of CABB-RY. hA3A(Y130F)-TadA8e(V106W) is linked to the N terminus of SpRY.

Determination of the on-target DNA editing efficiency of CABB-RY

Next, amplicon deep sequencing was used for the target sites to thoroughly characterize CABB-RY performance. Related simultaneous A/C conversions of some sites are shown in Figure 3A. The C-to-T editing window (positions 4–14) of CABB-RY was narrower than that of hA3A(Y130F)-RY (positions 1–14) (Figure S2A). Similarly, the A-to-G editing window (positions 3–9) was slightly smaller than that of ABE8e(V106W)-RY (positions 3–11) (Figure 3B; Figure S2B). Analysis of the 48 targets containing both As and Cs in the editing window revealed that the C-to-T editing efficiency was 46.9%, while the A-to-G editing efficiency was 48.3%. Then, we examined the C-to-T and A-to-G editing activities

of CABB-RY in different PAMs. For NYN PAMs, we observed that the mean activities of both C-to-T and A-to-G editing were lower than those for NRN PAMs (36.7% versus 58.0% for C-to-T; 38.8% versus 59.8% for A-to-G) (Figure 3C; Figures S3A–S3C). This is consistent with the characteristics of SpRY.¹³ In addition, CABB-RY had a low insertion or deletion (indel) frequency of 1.60% (Figure S3D).

We further compared the editing efficiency between CABB-RY and coexpressed ABE8e(V106W)-RY and hA3A(Y130F)-RY in HEK293T cells (Figures S4A–S4C) and found that CABB-RY showed a slightly higher editing efficiency than the two coexpressed base editors (54.7% versus 47.8%). Additionally, robust base editing of CABB-RY was also observed in murine N2a cells (29.4%) (Figure S4D), suggesting that CABB-RY works universally for different species.

Determination of the off-target RNA editing efficiency of CABB-RY

One of the concerns regarding base editors is RNA off-target mutagenesis.¹⁷ To detect transcriptome-wide RNA off-target effects,

the locations of SpRY and the two deaminases (Figure 2A; Sequence S1).

Then, we compared the on-target editing efficiency of the four constructs across 48 endogenous sites. Each single-guide RNA (sgRNA), together with four individual constructs, was cotransfected into HEK293T cells. The editing efficiency was analyzed by Sanger sequencing, and the results showed that all 4 editors induced simultaneous A/C conversions. The mean efficiency of CABB-1 was higher than that of CABB-2 (21.5% versus 17.71% for A-to-G; 21.8% versus 14.88% for C-to-T). The mean A-to-G editing efficiency of CABB-3 was far lower than that of CABB-1 (5.4% versus 21.5%), while the mean C-to-T editing efficiency of CABB-4 was far lower than that of CABB-1 (11.1% versus 21.8%) (Figures 2B and 2C). We further analyzed the C-to-T and A-to-G editing activities of CABB-1 at sites of different PAMs (Figures 2D and 2E). Similarly, the CABB-1 had the best performance. Therefore, we chose CABB-1, in which hA3A(Y130F)-TadA8e(V106W) was fused to the N terminus of SpRY, for further study and named it CABB-RY (Figure 2F).

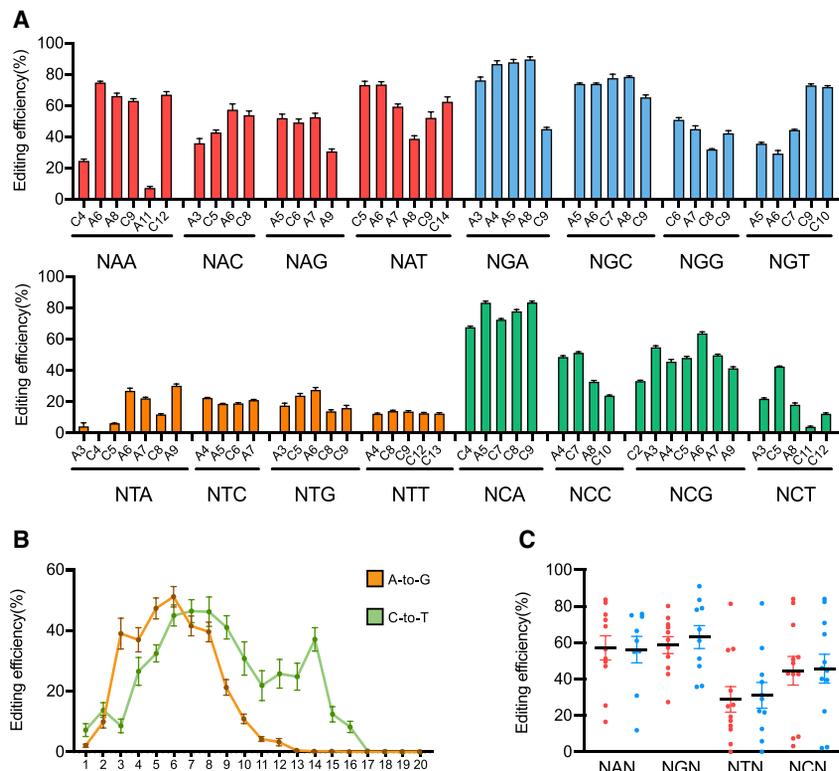


Figure 3. Determination of the on-target DNA editing efficiency of CBE-RY in HEK293T cells

(A) Relative C-to-G and A-to-T base editing efficiencies of CBE-RY with different PAMs at 16 human genomic target DNA sites ($n = 3$ independent replicates). The error bars represent the SEM values. (B) Sequencing analysis of the base editing window of CBE-RY across NNN PAMs in HEK293T cells. The error bars represent the SEM values. The data are presented for each edited base separately at 48 endogenous sites ($n = 3$ independent replicates). (C) Aggregate distribution of C-to-T (red) and A-to-G (blue) edits made across the editing window with CBE-RY. The error bars represent the SEM values. The data are presented for each edited base in the editing window at 48 endogenous sites.

ABE8e(V106W)-RY or hA3A(Y130F)-RY in HEK293T cells. Compared with those of the single base editors, the editing efficiencies for C-to-T and A-to-G editing of CBE-RY were slightly lower (25.8% versus 31.4% for C-to-T, 38.1% versus 43.9% for A-to-G) (Figure 5A). However, CBE-RY created more mutation types and amino acid types than the single base editors (Figures S6 and S7) and showed high simultaneous A/C conversion efficiency (31.8%) in terms of simulating MNVs, which is unfeasible for a single base editor. It is worth noting that CBE-RY induced bystander edits like those induced by ABE and CBE. To better understand how bystander edits confound MNV simulation using CBE-RY, we first analyzed the potential bystanders flanking two adjacent MNVs with gnomAD (Figure S8) and found that 18.9% of MNVs may not be affected by bystanders, while 14.4% of MNVs may be affected by bystanders. Thus, CBE-RY is still a potentially useful tool for modeling MNVs.

we performed RNA sequencing (RNA-seq) on HEK293T cells coexpressing CBE-RY, ABE8e(V106W)-RY, hA3A(Y130F)-RY, and GFP (negative control) with a sgRNA targeting NAT SITE3 (Table S1), which had high DNA on-target editing efficiency (Figure 4A). As expected, ABE8e(V106W)-RY introduced thousands of RNA edits in the transcriptome, while hA3A(Y130F)-RY had a slight effect (Figure 4B; Figure S5A). Interestingly, the number of RNA off-target edits of CBE-RY was one-third that of ABE8e(V106W)-RY (Figure 4B; Figure S5A). In addition, these editors did not cause significant differences in gene expression at the whole-transcriptome level (Figures S5B–S5D).

Then, we further analyzed the specific types of RNA off-target edits and found that the main off-target edit type for CBE-RY and ABE8e(V106W)-RY was A-to-I RNA editing, while that of hA3A(Y130F)-RY was C-to-U RNA editing (Figures 4C and 4D). Interestingly, CBE-RY showed lower A-to-I RNA editing than the other editors (nearly 3,000) and undetectable C-to-U RNA editing (comparable to the editing of the GFP control) (Figures 4C and 4D). A possible explanation of the performance may be the embedding of the deaminase domains in nCas9 for minimization of off-target effects of base editors;¹⁸ in addition, the combination of hA3A(Y130F) and TadaA8e(V106W) may create possible steric hindrance to reduce off-target edits.

Successful simulation of MNVs by CBE-RY in HEK293T cells

Given these findings, we tested ten sites in gnomAD and directly compared the editing efficiency of CBE-RY with that of

PEs can theoretically simulate most variants, including MNVs, with no bystander edits. Therefore, we compared the simulation efficiency between PE3 and CBE-RY. A pegRNA with a 13 nt prime binding site (PBS) length and a 12 nt reverse transcriptase (RT) template length was designed as suggested.¹⁹ As expected, both CBE-RY and PE3 successfully induced MNVs, but CBE-RY exhibited significantly higher simulation efficiency than PE3 (31.8% versus 2.8%; Figure 5B). Notably, perfect simulation of MNVs without bystander edits was successfully achieved in 8.7% of cases on average (Figure 5C). These results demonstrate that CBE-RY is a good tool for simulating MNVs.

To further explore the accuracy of MNV simulation by CBE-RY, we analyzed the proportion of perfectly simulated MNVs. The results showed that the influence of bystander edits was site dependent. Some sites retained a relatively high percentage of perfect MNV simulations (35.47% for the WDR90 site, 13.47% for the SYNMM), while some sites had significantly reduced proportions of perfect MNV

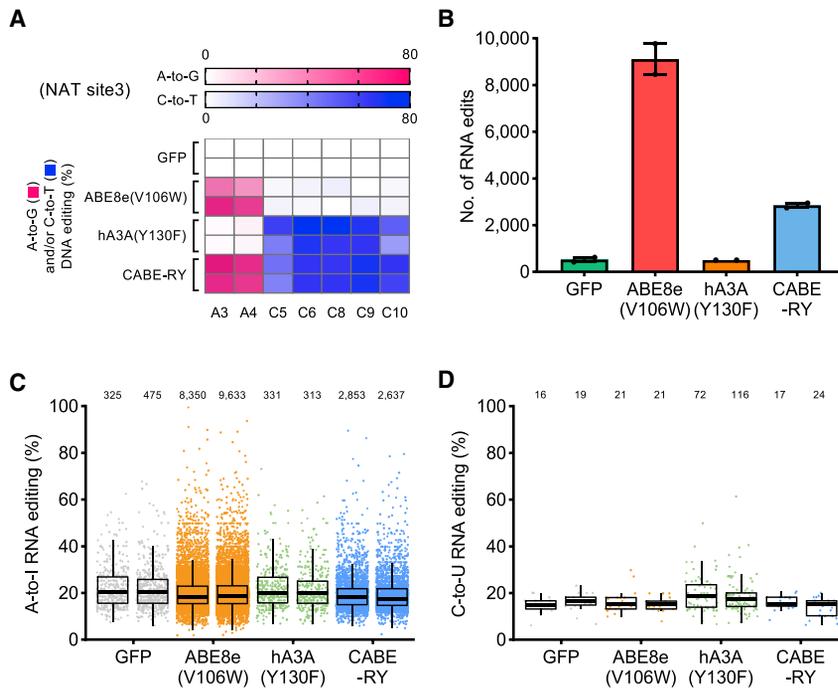


Figure 4. Determination of the off-target RNA editing efficiency of CABE-RY

(A) Heatmaps showing the on-target editing frequencies of GFP (control), ABE8e(V106W), hA3A(Y130F), and CABE-RY in NAT SITE3. A-to-G edits are indicated in pink, and C-to-T edits are indicated in blue. (B) Total number of RNA off-target edits detected in RNA-seq experiments for GFP (control), ABE8e(V106W), hA3A(Y130F), or CABE-RY. The error bars represent the SEM values. (C and D) Jitter plots showing the efficiencies of A-to-I (C) or C-to-U (D) RNA off-target editing. The data shown are from two independent replicates. The number of edits is represented on the top.

simulations (Figure S6). Taken together, the results indicate that CABE-RY can efficiently simulate MNVs and is a versatile tool for dissection of the functions of MNVs.

DISCUSSION

The data in gnomAD reveal that simultaneous A/C conversion has a higher mutation frequency than AA or CC conversion, particularly in two adjacent mutations for MNVs, and this mutation type can be simulated by reported dual-mutation base editors.^{6–10} However, the published dual-mutation base editors are restricted to sites with NGG PAMs, which greatly limits their application. Here, we constructed a dual-mutation base editor, CABE-RY, by combining hA3A(Y130F), Tada8e(V106W), and SpRY. This editor was found to effectively induce simultaneous A/C conversion over the whole genome with a sgRNA after successful verification. Characterized by the presence of SpRY, CABE-RY edits many more sites than existing dual-mutation editors with minimal PAM restriction.

We further characterized the editing of CABE-RY and revealed that CABE-RY efficiently edited DNA (46.9% efficiency for C-to-T edits and 48.3% efficiency for A-to-G edits). Interestingly, CABE-RY induced less RNA off-target mutagenesis, including less A-to-I RNA editing and undetectable C-to-U RNA editing, than the other editors tested. To illustrate that CABE-RY is an appropriate tool to simulate MNVs, we compared the simulation efficiency between PE3 and CABE-RY. CABE-RY exhibited higher simulation efficiency than PE3 (31.8% versus 2.8%), even when we focused only on perfect MNV (8.7% versus 2.0%). Here, we have confirmed that CABE-RY can efficiently generate simultaneous A/C conversions. Nevertheless, several concerns remain. For example, self-targeting of sgRNA is un-

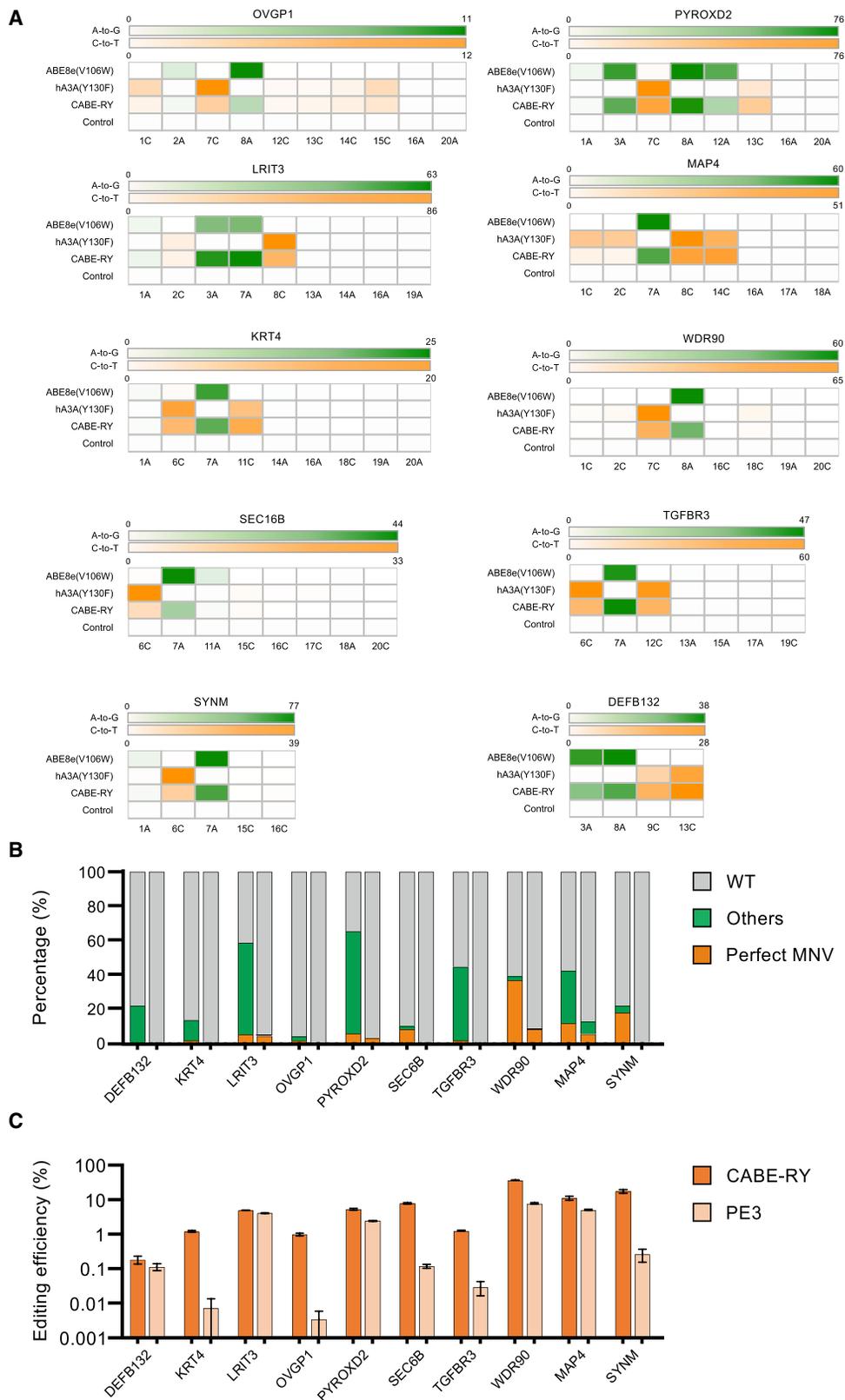
avoidable, and it may lead to wrong sgRNA generation, resulting in reduction of editing efficiency and even unpredictable targeting. Meanwhile, CABE-RY exhibits robust activities with NRN PAMs but lower activities with NYN PAMs, which limits the use of CABE-RY over the genome. Considering other SpCas9 variants may have stronger affinity with NYN PAMs, such as FnCas9RHA on YG PAMs,²⁰ M44 on TTTN PAMs,²¹ AsCas12a-K949A on TTTV PAMs, and²² Nme2Cas9 on NNNNCC PAMs,²³ we can choose these SpCas9 variants instead of SpRY, which may improve editing efficiency at specific sites. Nevertheless, to create a fully PAMless Cas9 in the future is desired. In fact, ortholog mining and protein engineering have produced many PAM variants. The combination of these strategies may further relieve PAM restrictions and create a truly PAMless Cas9.²⁴

In addition, we also compared CABE-RY editing frequencies with those of single base editors. Our results demonstrated that compared with a single base editor, CABE-RY has several advantages. First, CABE-RY can efficiently generate simultaneous A/C conversions to simulate MNVs. Second, CABE-RY can generate more mutation types than single base editors; thus, it can effectively simulate 70 amino acid substitutions, including six that cannot be achieved by existing single base editors, which is helpful for understanding disease-related mutations. For example, the unique p.Cys91His mutation has been discovered in Niemann-Pick disease and p.Gln326Trp in intervertebral disc disease.^{25,26} In addition, it's reported that simultaneous A/C conversion induced higher hemoglobin subunit gamma (HBG) reactivation.⁹ In conclusion, CABE-RY is a versatile tool for the field of gene editing and is especially useful for simulating MNVs.

MATERIALS AND METHODS

Cell culture

HEK293T cells were purchased from the American Type Culture Collection (ATCC). The cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. We maintained the cell lines at 37°C in a 5% CO₂ cell culture incubator.



(legend on next page)

Plasmid construction

To construct base editor expression plasmids, human codon-optimized DNA sequences of hA3A(Y130F), TadA8e(V106W), and UGI were synthesized by GenScript and cloned into the pCMV-SpRY(D10A) backbone containing C-terminal-fused EGFP and blasticidin (BSD). To construct sgRNA expression vectors, synthesized oligos were annealed and ligated into a BsaI-digested sgRNA expression vector (pGL3-U6-sgRNA-mCherry plasmid). To construct pegRNA expression vectors, synthesized oligos were annealed and ligated into pegRNA expression vector (pGL3-U6-pegRNA-mCherry plasmid). Sequences of sgRNA and pegRNA constructs used in this work are listed in [Tables S1](#) and [S2](#).

Analysis of on-target editing

To evaluate editing efficiency, HEK293T cells were seeded into 24-well plates 1 day before the analysis. The base editor expression plasmids (1,000 ng) and corresponding sgRNA plasmids (500 ng) were cotransfected using Lipofectamine 2000 (Life Technologies) as the manufacturer's protocol recommended. Seventy-two hours after transfection, ~10,000 cells with dual fluorescence signals (GFP and mCherry) were collected by fluorescence-activated cell sorting (FACS) to improve efficiency ([Figure S9](#)), and these cells were harvested for genomic DNA extraction using QuickExtract DNA Extraction Solution (Lucigen) according to the manufacturer's protocols. The genomic regions encompassing the target sites were amplified from the genomic DNA with Phanta Max Super-Fidelity DNA polymerase (Vazyme, P505-03). The primers used are listed in [Table S3](#). The PCR products were analyzed by Sanger sequencing or high-throughput sequencing as indicated. For Sanger sequencing, the chromatograms were quantified using EditR. For high-throughput sequencing, the PCR products were sequenced on an Illumina HiSeq X Ten (2 × 150 paired-end) at the Novogene Bioinformatics Institute (Beijing, China). The sequencing data were analyzed using CRISPResso2.

Analysis of RNA off-target editing

HEK293T cells were seeded into 6-cm dishes and transfected with 4 μg of CABE-RY, hA3A(Y130F)-RY, ABE8e(V106W)-RY, and GFP (control) plasmids and 2 μg of sgRNA expression vector using Lipofectamine 2000 at ~70% confluency. Two days after transfection, the top 15% of the GFP signal-positive cells were harvested by FACS. RNA was immediately extracted using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. The RNA samples were subjected to deep sequencing (~20 million reads per sample) on an Illumina HiSeq X Ten platform (2 × 150 paired-end) at the Novogene Bioinformatics Institute (Beijing, China). The clean data

were first mapped to the human reference genome (version: hg38) with annotations from GENCODE version 30 by STAR software (version 2.5.1). After removing duplicates, GATK HaplotypeCaller (version 4.1.2) was used to identify and filter the edits. All edits were verified, and the efficiency was calculated using the bam-readcount program with the parameters -q 20 -b 30. Importantly, for the reference allele in the wild-type sample, the depth of a given edit had to be least 10×, and all the edits had to be present in at least 99% of reads.

Statistics and reproducibility

All statistical analyses mentioned above were performed on data from at least 3 biologically independent experiments (n = 3). The data shown in this research were statistically analyzed by unpaired two-tailed Student's t test using GraphPad Software (GraphPad Prism 8). A p value smaller than 0.05 was considered to indicate statistical significance.

Data availability

The high-throughput sequencing data have deposited. The accession code is NCBI Sequence Read Archive database: PRJNA688630. All other data are available upon reasonable request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omtn.2021.07.016>.

ACKNOWLEDGMENTS

We thank members of Huang lab and Xu lab for helpful discussions. We also thank the Molecular and Cell Biology Core Facility (MCBCF) at the School of Life Science and Technology, ShanghaiTech University, for providing technical support. This work is supported by the National Science Foundation of China (81830004) and the Leading Talents of Guangdong Province Program (608285568031).

AUTHOR CONTRIBUTIONS

X.H. and X.X. conceived, designed, and supervised the project. W.T. and Q.L. performed most experiments with the help of other authors. S.H. analyzed and interpreted the data. W.T., Q.L., and X.H. wrote the paper with inputs from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Hay, E.H.A., Utsunomiya, Y.T., Xu, L., Zhou, Y., Neves, H.H.R., Carvalheiro, R., Bickhart, D.M., Ma, L., Garcia, J.F., and Liu, G.E. (2018). Genomic predictions

Figure 5. Evaluation of MNV simulation by CABE-RY in HEK293T cells

(A) Heatmaps showing the A-to-G (green) and C-to-T (orange) editing frequencies of ABE8e(V106W), hA3A(Y130F), and CABE-RY at ten sites (n = 3 independent replicates). The number and the letter at the bottom represent the related position and the respective base in the protospacer sequence. (B) Proportion of mutation types edited by CABE-RY (left) and PE3 (right) in different sites. Data are analyzed by deep sequencing and summarized from three independent experiments. Perfect MNV (orange) represents the mutation type that just alters the target MNV site without bystanders. Others (green) represent the mutation types that just alters the target MNV site within bystanders. WT (gray) represents unedited type. (C) Comparison of simulated perfect MNV efficiency between CABE-RY and PE3 on 10 endogenous sites. The error bars represent the SEM values.

- combining SNP markers and copy number variations in Nellore cattle. *BMC Genomics* 19, 441.
2. Kaplanis, J., Akawi, N., Gallone, G., McRae, J.F., Prigmore, E., Wright, C.F., Fitzpatrick, D.R., Firth, H.V., Barrett, J.C., and Hurles, M.E.; Deciphering Developmental Disorders study (2019). Exome-wide assessment of the functional impact and pathogenicity of multinucleotide mutations. *Genome Res.* 29, 1047–1056.
 3. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.
 4. McKenna, A., Hanna, M., Banks, A., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
 5. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
 6. Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., Qiu, J.L., Chen, Y.H., Li, J., and Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat. Biotechnol.* 38, 875–882.
 7. Grünewald, J., Zhou, R., Lareau, C.A., Garcia, S.P., Iyer, S., Miller, B.R., Langner, L.M., Hsu, J.Y., Aryee, M.J., and Joung, J.K. (2020). A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat. Biotechnol.* 38, 861–864.
 8. Xie, J., Huang, X., Wang, X., Gou, S., Liang, Y., Chen, F., Li, N., Ouyang, Z., Zhang, Q., Ge, W., et al. (2020). ACBE, a new base editor for simultaneous C-to-T and A-to-G substitutions in mammalian systems. *BMC Biol.* 18, 131.
 9. Zhang, X., Zhu, B., Chen, L., Xie, L., Yu, W., Wang, Y., Li, L., Yin, S., Yang, L., Hu, H., et al. (2020). Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* 38, 856–860.
 10. Sakata, R.C., Ishiguro, S., Mori, H., Tanaka, M., Tatsuno, K., Ueda, H., Yamamoto, S., Seki, M., Masuyama, N., Nishida, K., et al. (2020). Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat. Biotechnol.* 38, 865–869.
 11. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157.
 12. Liu, Y., Li, X., He, S., Huang, S., Li, C., Chen, Y., Liu, Z., Huang, X., and Wang, X. (2020). Efficient generation of mouse models with the prime editing system. *Cell Discov.* 6, 27.
 13. Walton, R.T., Christie, K.A., Whittaker, M.N., and Kleinstiver, B.P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 368, 290–296.
 14. Wang, X., Li, J., Wang, Y., Yang, B., Wei, J., Wu, J., Wang, R., Huang, X., Chen, J., and Yang, L. (2018). Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat. Biotechnol.* 36, 946–949.
 15. Richter, M.F., Zhao, K.T., Eton, E., Lapinaite, A., Newby, G.A., Thuronyi, B.W., Wilson, C., Koblan, L.W., Zeng, J., Bauer, D.E., et al. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* 38, 883–891.
 16. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443.
 17. Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., Han, L., Wei, Y., Hu, X., Zeng, R., et al. (2019). Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* 571, 275–278.
 18. Liu, Y., Zhou, C., Huang, S., Dang, L., Wei, Y., He, J., Zhou, Y., Mao, S., Tao, W., Zhang, Y., et al. (2020). A Cas-embedding strategy for minimizing off-target effects of DNA base editors. *Nat. Commun.* 11, 6073.
 19. Kim, H.K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S., and Kim, H.H. (2021). Predicting the efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol.* 39, 198–206.
 20. Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F., et al. (2016). Structure and Engineering of Francisella novicida Cas9. *Cell* 164, 950–961.
 21. Liu, R.M., Liang, L.L., Freed, E., Chang, H., Oh, E., Liu, Z.Y., Garst, A., Eckert, C.A., and Gill, R.T. (2019). Synthetic chimeric nucleases function for efficient genome editing. *Nat. Commun.* 10, 5524.
 22. Gao, L., Cox, D.B.T., Yan, W.X., Manteiga, J.C., Schneider, M.W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., and Zhang, F. (2017). Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* 35, 789–792.
 23. Edraki, A., Mir, A., Ibraheim, R., Gainetdinov, I., Yoon, Y., Song, C.Q., Cao, Y., Gallant, J., Xue, W., Rivera-Pérez, J.A., and Sontheimer, E.J. (2019). A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. *Mol. Cell* 73, 714–726.e4.
 24. Collias, D., and Beisel, C.L. (2021). CRISPR technologies and the search for the PAM-free nuclease. *Nat. Commun.* 12, 555.
 25. Hollak, C.E., de Sonnaville, E.S., Cassiman, D., Linthorst, G.E., Groener, J.E., Morava, E., Wevers, R.A., Mannens, M., Aerts, J.M., Meersseman, W., et al. (2012). Acid sphingomyelinase (Asm) deficiency patients in The Netherlands and Belgium: disease spectrum and natural course in attenuated patients. *Mol. Genet. Metab.* 107, 526–533.
 26. Annunen, S., Paasilta, P., Lohiniva, J., Perälä, M., Pihlajamaa, T., Karppinen, J., Tervonen, O., Kröger, H., Lähde, S., Vanharanta, H., et al. (1999). An allele of COL9A2 associated with intervertebral disc disease. *Science* 285, 409–412.