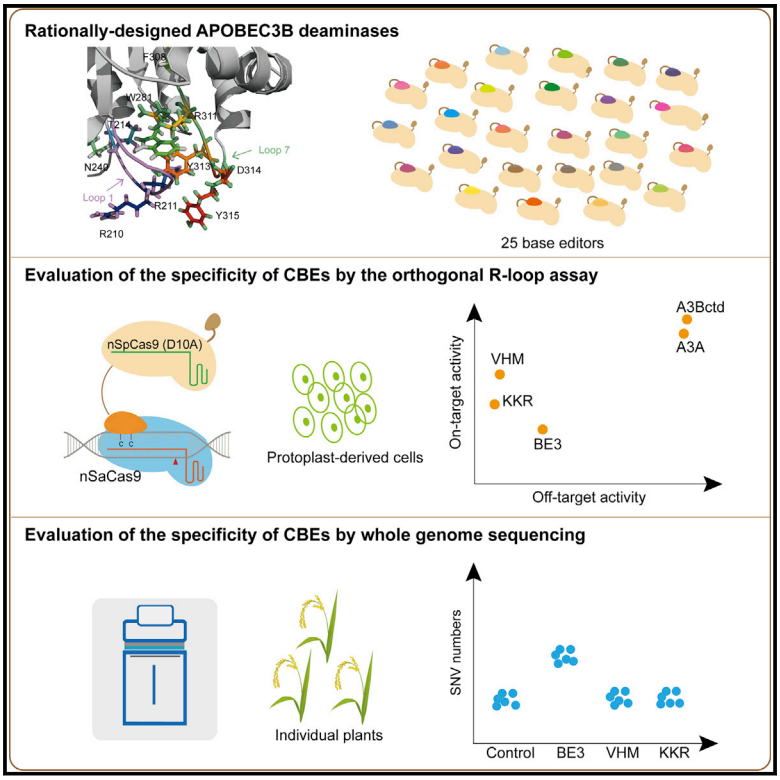
Molecular Cell

Article

Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity

Graphical Abstract



Highlights

- The nSaCas9-mediated R-loop assay enables specificity analysis of CBEs in plants
- A3Bctd, a truncated APOBEC3B deaminase, is rationally designed to generate new CBEs
- The 25 A3Bctd variants are screened with the R-loop assay
- A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 exhibit high specificity and precision in plants

Authors

Shuai Jin, Hongyuan Fei, Zixu Zhu, ..., Yu-Hang Chen, Yanpeng Wang, Caixia Gao

Correspondence

yanpengwang@genetics.ac.cn (Y.W.), cxgao@genetics.ac.cn (C.G.)

In Brief

Jin et al. establish nSaCas9-mediated orthogonal R-loop assay, a rapid, high-throughput, and inexpensive method for assessing CBEs in plants. Using it, they assess the specificity of 25 rationally designed A3Bctd-BE3s, identifying A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 with high specificity and precision.

Jin et al., 2020, Molecular Cell 79, 1–13
September 3, 2020
https://doi.org/10.1016/j.molcel.2020.07.005

CellPress

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

Molecular Cell

CellPress

Article

Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity

Shuai Jin,[1,2,6] Hongyuan Fei,[1,2,6] Zixu Zhu,[1,2,6] Yingfeng Luo,[3,6] Jinxing Liu,[1] Shenghan Gao,[3] Feng Zhang,[4] Yu-Hang Chen,[5] Yanpeng Wang,[1,2,*] and Caixia Gao[1,2,7,*]

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China

[2]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China

[3]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

[4]Department of Plant and Microbial Biology, Center for Plant Precision Genomics, The Microbial and Plant Genomics Institute, University of Minnesota, Minneapolis, MN 55108, USA

[5]State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China

[6]These authors contributed equally

[7]Lead Contact

*Correspondence: yanpengwang@genetics.ac.cn (Y.W.), cxgao@genetics.ac.cn (C.G.)

https://doi.org/10.1016/j.molcel.2020.07.005

SUMMARY

Cytosine base editors (CBEs) generate C-to-T nucleotide substitutions in genomic target sites without inducing double-strand breaks. However, CBEs such as BE3 can cause genome-wide off-target changes via sgRNA-independent DNA deamination. By leveraging the orthogonal R-loops generated by SaCas9 nickase to mimic actively transcribed genomic loci that are more susceptible to cytidine deaminase, we set up a high-throughput assay for assessing sgRNA-independent off-target effects of CBEs in rice protoplasts. The reliability of this assay was confirmed by the whole-genome sequencing (WGS) of 10 base editors in regenerated rice plants. The R-loop assay was used to screen a series of rationally designed A3Bctd-BE3 variants for improved specificity. We obtained 2 efficient CBE variants, A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3, and the WGS analysis revealed that these new CBEs eliminated sgRNA-independent DNA off-target edits in rice plants. Moreover, these 2 base editor variants were more precise at their target sites by producing fewer multiple C edits.

INTRODUCTION

Cytosine and adenine base editors (CBEs and ABEs), which produce highly efficient targeted point mutations in genomic DNA without causing double-stranded DNA breaks, have been used in therapeutic (Chadwick et al., 2017; Liang et al., 2017; Ryu et al., 2018), agricultural (Li et al., 2020; Xue et al., 2018; Zhang et al., 2019), and research settings (Chen et al., 2019; Komor et al., 2017; Rees and Liu, 2018). Current CBEs, such as BE3, which fuse the nickase-type Cas9 (nCas9) protein with a deaminase domain and an uracil glycosylase inhibitor (UGI), catalyze conversions of cytosine to thymine at the guide RNA (gRNA)-targeted sites (Gaudelli et al., 2017; Komor et al., 2016; Nishida et al., 2016). Previous studies using whole-genome sequencing (WGS) showed that BE3 induces off-target C-to-T changes in rice, mouse, and human cells (Jin et al., 2019; Lee et al., 2020; McGrath et al., 2019; Zuo et al., 2019). These mutations were independent of single-guide RNA (sgRNA)-Cas9-programmed DNA binding, and were enriched in transcribed regions of the genome (Jin et al., 2019; Zuo et al., 2019). They are probably due to the high affinity of the cytosine deaminases for single-stranded DNA (ssDNA) (Jin et al., 2019; Lee et al., 2020; Zuo et al., 2019). This high affinity also affects the precision of on-target activity, such that most CBEs generate multiple C mutations if multiple cytosines are present within or around target sites (Cheng et al., 2019; Gehrke et al., 2018; Kim et al., 2017; Lee et al., 2018; Tan et al., 2020). These sgRNA-independent off-target editing and bystander effects have limited the application of CBEs. Recently, several rapid and cost-effective methods for screening the sgRNA-independent deamination activity of different CBEs were developed for *Escherichia coli* and human cells (Doman et al., 2020). Using these methods, several derivatives of the base editor BE4, such as EE-BE4, YE1-BE4, YE2-BE4, and YEE-BE4, were found to display decreased sgRNA-independent off-target activity in human cells (Doman et al., 2020); however, these methods had not been validated in plant cells.

We envisioned the possibility of developing CBEs that were more precise in genome-wide activity by rationally designing

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

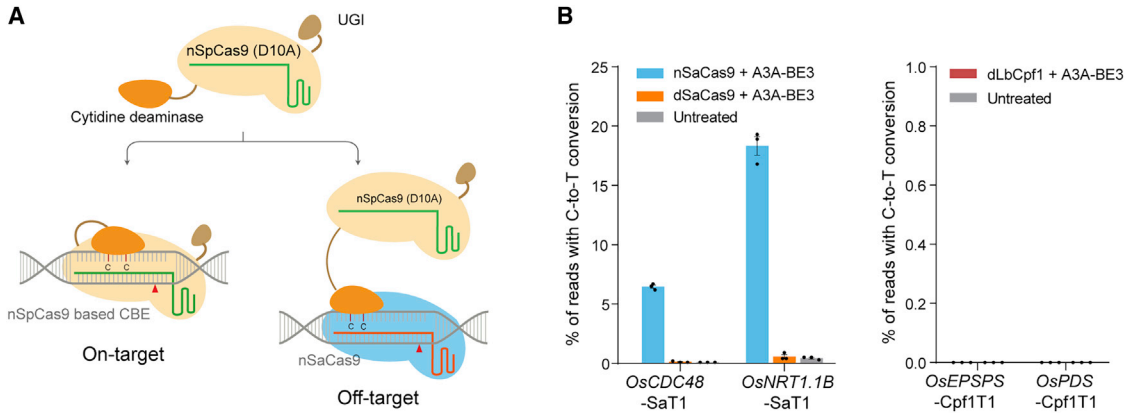


Figure 1. Setting Up the Orthogonal R-Loop Assay for Evaluating the sgRNA-Independent Off-Target Activities of SpCas9-Derived Cytosine Base Editors in Plants

(A) Schematic of the orthogonal R-loop assay.

(B) Frequencies (percentage) of C-to-T conversion induced by A3A-BE3 in nSaCas9-, dSaCas9-, and dLbCpf1- targeted regions. Two sites targeted by dLbCpf1, nSaCas9 and dSaCas9, were used for producing the ssDNA regions (*OsCDC48*-SaT1 and *OsNRT1.1B*-SaT1 by dSaCas9 and nSaCas9, *OsEPSPS*-Cpf1T1 and *OsPDS*-Cpf1T1 by dLbCpf1). Values and error bars indicate the means ± SEMs of 3 independent experiments.

See also Figure S1.

their deaminase domains to have lower affinities for ssDNA. To this end, we optimized and validated the orthogonal R-loop assay for application in plants and used it to evaluate the sgRNA-independent off-target activities of a series of A3Bctd deaminase (a truncated human APOBEC3B cytidine deaminase) variants. In this way, we obtained 2 new CBE variants, A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3, that exhibited efficient on-target activity and markedly reduced sgRNA-independent off-target activity, and were further validated by WGS in plants. In addition, these 2 new CBEs behaved more precisely at their target sites, and mainly produced single and double C edits. We also validated the R-loop assay on 10 CBEs by WGS of regenerated rice plants, and the results indicate that it could be applied to a broad range of new base editors.

RESULTS

Development of the nSaCas9-Mediated Orthogonal R-Loop Assay for Evaluating the Specificities of CBEs in Plant Cells

WGS methods for evaluating genome-wide off-target events via the CRISPR-Cas system or base editors are low throughput, costly, and time-consuming (Iyer et al., 2018; Tang et al., 2018; Willi et al., 2018; Jin et al., 2019; Li et al., 2019; Zuo et al., 2019; Lee et al., 2020). In contrast, the orthogonal R-loop assay reported by Doman et al. (2020) has proven effective in evaluating sgRNA-independent off-target activity in human cells. In the assay, they used the orthogonal CRISPR system, dSaCas9, to create ssDNA regions in human cells that acted as targets for sgRNA-independent deamination changes that could be detected by amplicon deep sequencing in a high-throughput manner (Figure 1A).

We set out to test and optimize the high-throughput orthogonal R-loop assay (Doman et al., 2020) to evaluate the sgRNA-independent off-target activities of CBEs in plant cells. We used 3 Cas proteins, deactivated LbCpf1 (dLbCpf1), deactivated SaCas9 (dSaCas9), and SaCas9 nickase (nSaCas9), to produce ssDNA regions in 4 rice genomic sites (*OsCDC48*-SaT1 and *OsNRT1.1B*-SaT1 in combination with dSaCas9 or nSaCas9, and *OsEPSPS*-Cpf1T1 and *OsPDS*-Cpf1T1 in combination with dLbCpf1) (Figures 1B and S1; Table S1). We chose A3A-BE3, which exhibited high cytosine deaminase activity in previous reports (Li et al., 2020; Wang et al., 2018; Zong et al., 2018), to determine whether it produced C-to-T changes in the 2 ssDNA regions. We co-transformed rice protoplasts with 3 plasmids, A3A-BE3 based on SpCas9, an ssDNA-generating Cas9 protein (i.e., dLbCpf1, dSaCas9, or nSaCas9), and a corresponding gRNA targeting a given genetic site. Two days after transformation, targeted deep sequencing showed that A3A-BE3 had created C-to-T edits in the 2 ssDNA regions produced by nSaCas9, whereas no edits were detected in the ssDNA regions produced by dLbCpf1 and dSaCas9 (Figure 1B). Notably, these observations differed from those obtained in human cells in that in the latter, sgRNA-independent edits were observed in the ssDNA regions generated by dSaCas9 (Doman et al., 2020). This difference was probably due to the fact that rice protoplasts carry out DNA repair but not DNA replication. We inferred that the mutations in the ssDNA regions created by dLbCpf1 or dSaCas9 were mostly repaired by eukaryotic mismatch repair (MMR) or base-excision repair (BER), and were not being retained in the absence of DNA replication. However, with nCas9, which can induce DNA nicks in the opposite strand of the ssDNA with the deaminated cytidine, the unedited G in the nicking strand would simulate newly synthesized DNA to preferentially produce the C-to-T edits. We therefore selected nSaCas9 as the ssDNA producer for evaluating the off-target activities of CBEs in plant protoplasts.

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

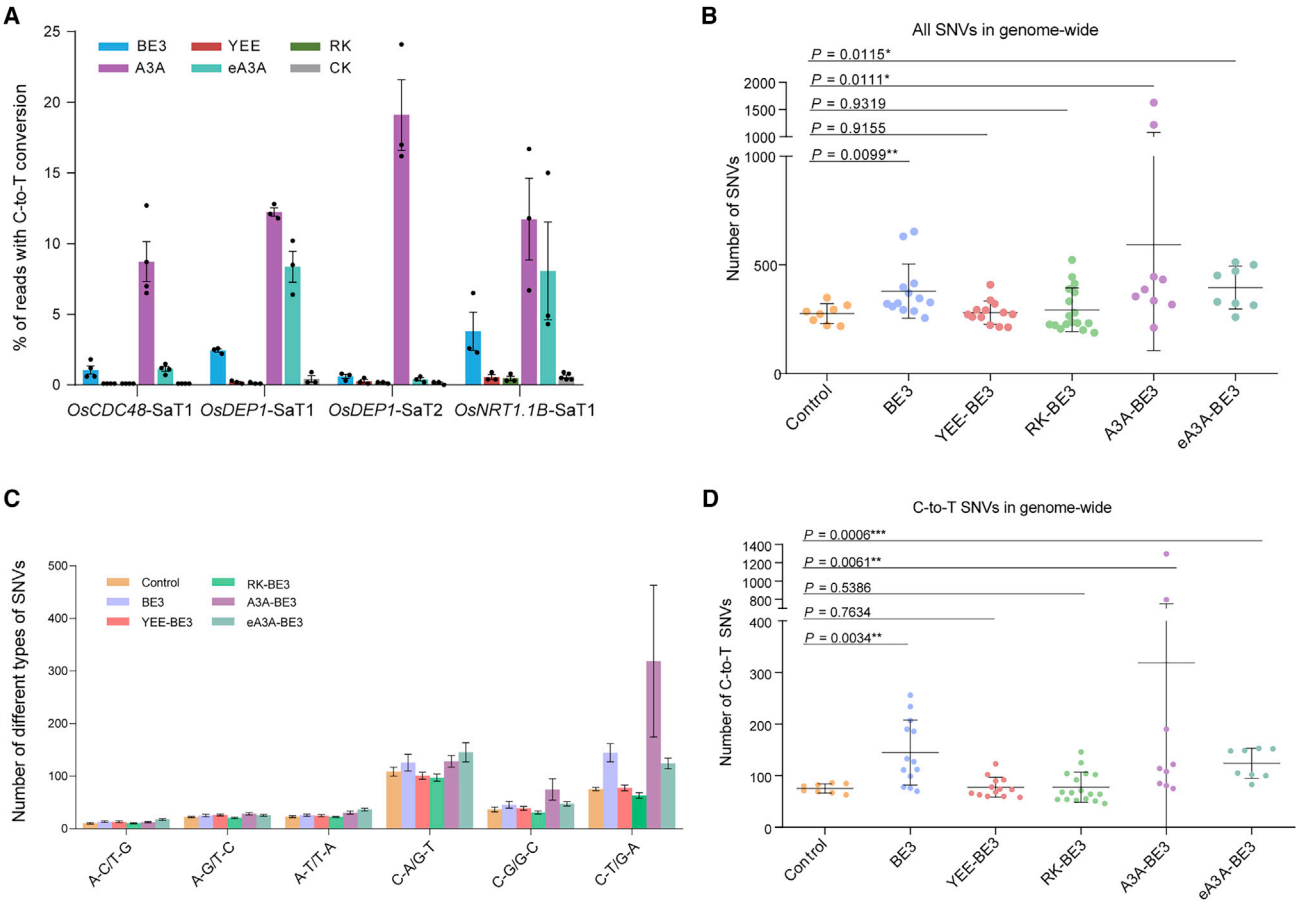


Figure 2. Validating the nSaCas9-Mediated Orthogonal R-Loop Assay in Plants by Whole-Genome Sequencing (WGS)

(A) Frequencies (percentages) of C-to-T conversions detected by the nSaCas9-mediated orthogonal R-loop assay using BE3, YEE-BE3, RK-BE3, A3A-BE3, and eA3A-BE3 in 4 nSaCas9-targeted ssDNA regions (*OsCDC48*-SaT1, *OsDEP1*-SaT2, *OsNRT1.1B*-SaT1, and *OsDEP1*-SaT1). Values and error bars are means ± SEMs of 3–5 independent experiments.

(B)–(D) Numbers of total SNVs (B), different types of SNVs (C), and C-to-T SNVs (D) identified in control and BE3-, RK-BE3-, YEE-BE3-, A3A-BE3-, and eA3A-BE3-treated plants by WGS. The average numbers of total SNVs were 276, 380, 294, 281, 593, and 396, respectively, and the average numbers of C-to-T SNVs were 75, 145, 78, 77, 319, and 124, respectively. The statistical significance of differences was calculated using the 2-sided Mann-Whitney *U* test. Pearson's correlation coefficient (r) and p values (*p < 0.05; **p < 0.01; ***p < 0.001) are shown.

See also Figure S2.

Validation of the nSaCas9-Mediated Orthogonal R-Loop Assay via WGS

To assess the nSaCas9-mediated orthogonal R-loop assay further, we selected 4 rice genomic loci as nSaCas9 targets (*OsCDC48*-SaT1, *OsNRT1.1B*-SaT1, *OsDEP1*-SaT1, and *OsDEP1*-SaT2) (Figure 2A; Table S1) and examined the off-target activities of 5 previously published CBE systems, BE3 (APOBEC1-BE3), YEE-BE3 (rAPOBEC1-W90Y+R126E+R132E-BE3), RK-BE3 (rAPOBEC1-R33A+K34A-BE3), A3A-BE3, and eA3A-BE3 (A3A-N57G-BE3) (Gehrke et al., 2018; Grünewald et al., 2019a, 2019b; Kim et al., 2017; Zong et al., 2017, 2018). Each CBE was co-transformed with nSaCas9 and its nSaCas9 sgRNA into rice protoplasts. Deep sequencing showed that A3A-BE3, eA3A-BE3, and BE3 yielded averages of 13.0% (8.7%–19.1%), 4.5% (0.4%–8.4%), and 2.0% (0.6%–3.8%) off-target edits, respectively, across 4 nSaCas9 sites (Figure 2A). More important, the BE3 variants YEE-BE3 and RK-BE3 gave almost undetectable levels of sgRNA-independent off-target edits (0.3% and 0.2%) (Figure 2A).

To determine whether the results of the nSaCas9-mediated orthogonal R-loop assay were representative of the genome-wide specificities of these CBEs, we compared the nSaCas9-mediated orthogonal R-loop assay results with WGS results. We transformed 5 CBEs (BE3, YEE-BE3, RK-BE3, eA3A-BE3, and A3A-BE3) into rice calli via *Agrobacterium* transformation. In total, 60 T0 regenerated (primary transformant) plants were analyzed by WGS along with 8 plants that went through the transformation process but with no transfer DNA integration, as controls (Table S2). In addition, 12 wild-type (WT) plants were used to filter out background mutations in the rice population (Table S2). The WGS results were analyzed, as reported previously (Jin et al., 2019). The identified changes were confirmed by Sanger sequencing at randomly selected sites with a 100% success rate (Figures S2A and S2B; Table S3). The results

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

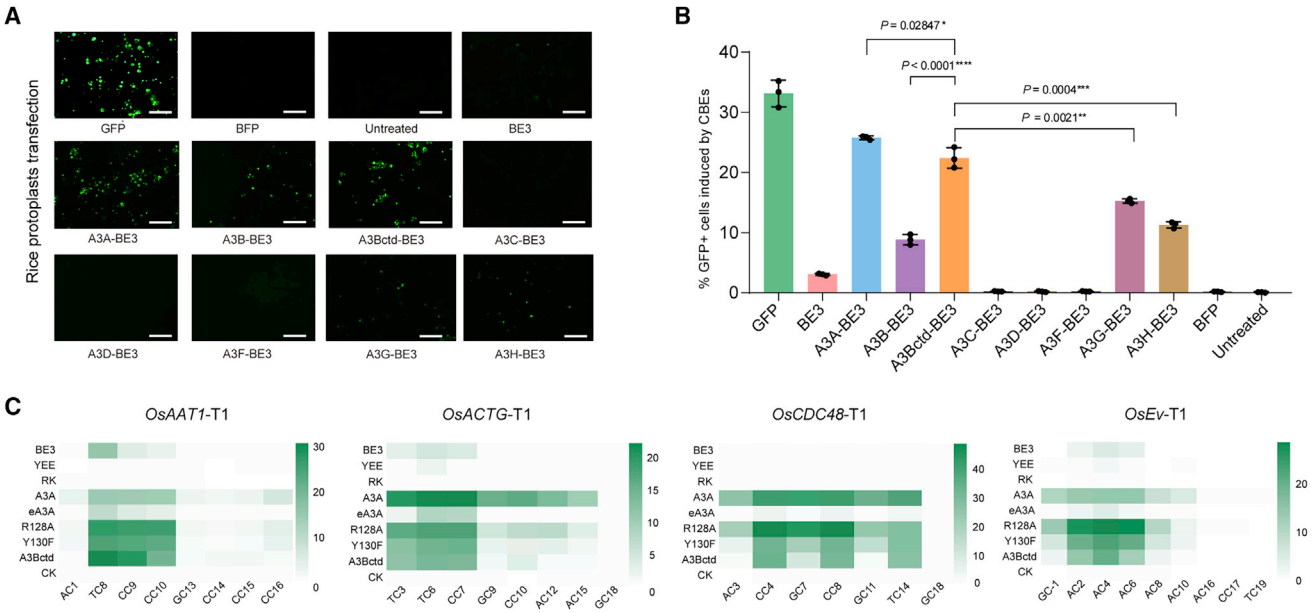


Figure 3. Screening Human APOBEC3-Based Base Editors for the Rationally Designed Platform

(A) Microscopic views of rice BFP-to-GFP conversion using 9 base editors (BE3, A3A-BE3, A3B-BE3, A3Bctd-BE3, A3C-BE3, A3D-BE3, A3F-BE3, A3G-BE3, and A3H-BE3), showing fields of rice protoplasts transformed with the relevant cytidine base editors together with pUbi-BFP and pOsU3-BFP-sgRNA.

(B) Frequencies (percentages) of C-to-T conversions by the relevant cytidine base editors in the target region of the coding sequence of BFP in rice protoplasts, measured by flow cytometry (FCM). Values represent means ± SDs of 3 independent biological replicates (n = 3). ****p < 0.0001. The statistical significance of differences was tested using 2-tailed Student's t tests.

(C) Frequencies (heatmaps) of C-to-T conversions induced by A3Bctd-BE3 and 7 reported base editors at 4 targeted sites in rice protoplasts. The frequencies are mean values of 3–6 independent biological replicates.

showed that the numbers of indels in base editor groups were not significantly different from those detected in the control group (Figure S2C), whereas the numbers of total single-nucleotide variants (SNVs) generated by BE3, A3A-BE3, and eA3A-BE3 groups were significantly higher than those found in the control group (Figure 2B).

Next, we analyzed mutation types of these detected SNVs and found that these increased SNVs were mainly C-to-T transitions (Figure 2C), which is consistent with our previous study (Jin et al., 2019). Therefore, we further analyzed C-to-T SNVs and found that A3A-BE3, BE3, and eA3A-BE3 produced significantly more genome-wide C-to-T SNVs than those found in the control group, whereas, as in the nSaCas9-mediated orthogonal R-loop assay, YEE-BE3 and RK-BE3 did not generate significantly more C-to-T SNVs than were found in the control group (Figures 2A and 2D). The results obtained from the R-loop assay are consistent with the genome-wide specificities of CBEs via WGS, indicating that this new assay provides a simple and high-throughput method for assessing the sgRNA-independent off-target activities of CBEs.

Screening Human APOBEC3-Based Base Editors for the Rationally Designed Platform

To further develop improved base editors with reduced off-targeting of sgRNA-independent deamination activity, we constructed a series of CBEs by replacing the deaminase domain of BE3 separately with 7 members of the human APOBEC3 cytidine deaminase family, including A3Bctd, a truncated APOBEC3B deaminase with enzymatic activity (Hou et al., 2019; Figures 3A and S3A). Use of the BFP-to-GFP reporter system in rice protoplasts showed that A3Bctd-BE3 yielded a high frequency of GFP-expressing cells (22.4%), ~7.2-fold higher than BE3 (3.1%) and similar to A3A-BE3 (25.7%) (Figures 3A and 3B), which is in agreement with the results using the human cell fluorescence reporter system (St Martin et al., 2018).

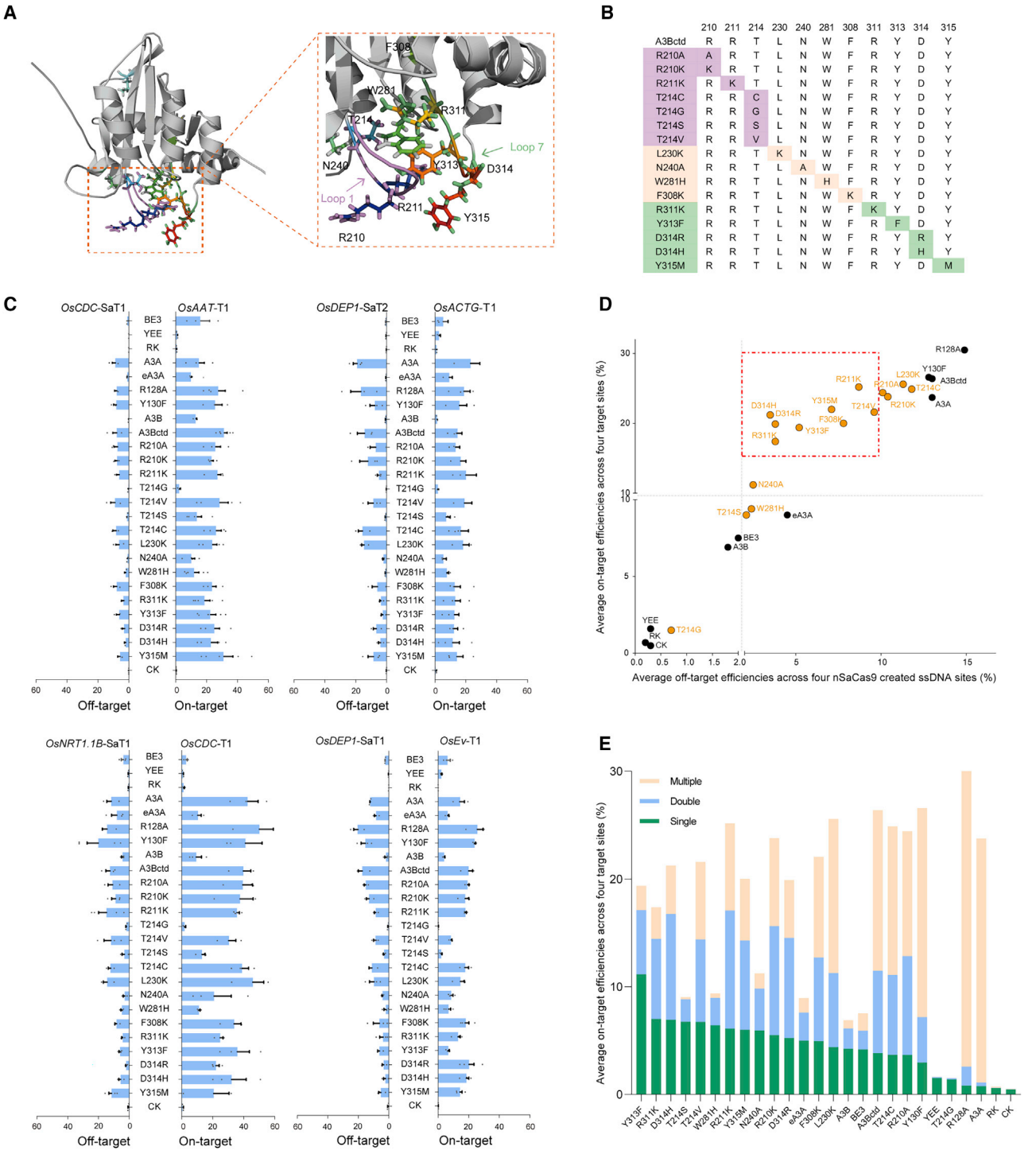
Next, we designed 4 sgRNAs targeting *OsAAT1*-T1, *OsACTG*-T1, *OsCDC48*-T1, and *OsEV*-T1 to examine the plant-editing characteristics of a set of CBEs (Table S1), including A3Bctd-BE3 and several other reported BE3 variants: BE3, YEE-BE3, RK-BE3, A3A-BE3, R128A-BE3 (A3A-R128A-BE3), Y130F-BE3 (A3A-Y130F-BE3), and eA3A-BE3 (Wang et al., 2018; Zhou et al., 2019). Deep sequencing showed that A3A-BE3, R128A-BE3, Y130F-BE3, and A3Bctd-BE3 had much higher C-to-T base editing rates at all target sites than BE3, YEE-BE3, RK-BE3, and eA3A-BE3. Among the highly active CBEs, we found that A3Bctd-BE3 had a more restricted editing window than A3A-BE3, R128A-BE3, and Y130F-BE3 (Figures 3C and S3B), suggesting that A3Bctd may be the best cytidine deaminase on which to base further engineering for improving cytosine base-editing specificity.

Evaluating the Specificities of the Rationally Designed A3Bctd Cytidine Deaminases by the R-Loop Assay

We therefore focused on A3Bctd-BE3-improved specificity. Analysis of the structure of the A3Bctd protein indicated that the amino acid residues in loop 1 (N205-T214) and loop 7

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

CellPress



	210	211	214	230	240	281	308	311	313	314	315
A3Bctd	R	R	T	L	N	W	F	R	Y	D	Y
R210A	A	R	T	L	N	W	F	R	Y	D	Y
R210K	K	R	T	L	N	W	F	R	Y	D	Y
R211K	R	K	T	L	N	W	F	R	Y	D	Y
T214C	R	R	C	L	N	W	F	R	Y	D	Y
T214G	R	R	G	L	N	W	F	R	Y	D	Y
T214S	R	R	S	L	N	W	F	R	Y	D	Y
T214V	R	R	V	L	N	W	F	R	Y	D	Y
L230K	R	R	T	K	N	W	F	R	Y	D	Y
N240A	R	R	T	L	A	W	F	R	Y	D	Y
W281H	R	R	T	L	N	H	F	R	Y	D	Y
F308K	R	R	T	L	N	W	K	R	Y	D	Y
R311K	R	R	T	L	N	W	F	K	Y	D	Y
Y313F	R	R	T	L	N	W	F	R	F	D	Y
D314R	R	R	T	L	N	W	F	R	Y	R	Y
D314H	R	R	T	L	N	W	F	R	Y	H	Y
Y315M	R	R	T	L	N	W	F	R	Y	D	M

Figure 4. Evaluating the Frequencies of On-Target and sgRNA-Independent Off-Target Edits Induced by the Rationally Designed A3Bctd-BE3 Variants

(A) Schematic showing the amino acid positions engineered in A3Bctd based on the previously reported protein structure (PDB: 2NBQ) (Siriwardena et al., 2015).

(B) Single amino acid substitutions in A3Bctd-BE3 variants. Violet indicates the amino acid residues in loop 1, green indicates the amino acid residues loop 7, and yellow indicates the amino acid residues in other regions.

(C) Comparison of the on-target and sgRNA-independent off-target editing frequencies of 16 engineered A3Bctd-BE3 variants. The frequencies of on-target C-to-T conversion at 4 rice loci and sgRNA-independent off-target C-to-T edits in 4 ssDNA regions created by nSaCas9 in rice protoplasts were measured by deep sequencing. The values and error bars indicate the means ± SEMs of 3–6 independent experiments.

(legend continued on next page)

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

(A310-P317) of human A3Bctd play important roles in A3B catalysis activity and ssDNA binding (Byeon et al., 2016; Hou et al., 2019; Shi et al., 2015; Siriwardena et al., 2015; Figure 4A). R211 is the gatekeeper for DNA binding and stacks against Y315 to close the active site in the apo structure (Hou et al., 2019). The R210 variants (R210A and R210K) in loop 1 were able to bind DNA but had decreased binding affinity, as measured in a fluorescence anisotropy-based assay (Hou et al., 2019); D314 is responsible for the substrate preference for thymidine over cytidine (Byeon et al., 2016). Loop 1 (Q213 and T214), loop 3 (K243, C247, G248, F249, Y250, and R252), loop 5 (W281, S282, S286, W287, and G288), and loop 7 (D314, Y315, and D316) residues undergo large ^{1}H-^{15}N chemical shifts upon ssDNA binding, suggesting that their conformation of these loops changes upon substrate binding (Byeon et al., 2016; Hou et al., 2019; Shi et al., 2015; Siriwardena et al., 2015). The L230K (located in β2), N240A, W281H (located in α3), and F308K (located in β4, but close to loop 7), were not located in loop 1 or loop 7, but also performed as controls. We reasoned that reducing the potential non-specific ssDNA-binding activity of A3Bctd deaminase by engineering these key residues may reduce the sgRNA-independent off-target activities of A3Bctd-mediated CBEs, as well as multiple Cs editing at on-target sites. To this end, we generated 16 A3Bctd-BE3 variants bearing single amino acid substitutions at various key positions identified by the structure information mentioned above (Figure 4B).

To analyze the on-target DNA editing efficiencies and sgRNA-independent off-target activities of these A3Bctd-BE3 variants, we transfected rice protoplasts with plasmids encoding them individually together with on-target sgRNAs (for *OsAAT1*-T1, *OsACTG*-T1, *OsCDC48*-T1, or *OsEV*-T1) and the ssDNA-producing construct nSaCas9 and the corresponding off-target sgRNA plasmids (*OsCDC48*-SaT1, *OsDEP1*-SaT1, *OsDEP1*-SaT2, and *OsNRT1.1B*-SaT1) (Table S1). Targeted deep sequencing showed that R211K, T214V, F308K, R311K, Y313F, D314R, D314H, and Y315M variants of A3Bctd-BE3 retained efficient on-target activities (>15% average on-target activity across the 4 on-target sites) and had markedly lower off-target activities (<10% average off-target activity across the 4 off-target sites) at 4 SaCas9 target sites than the original A3Bctd-BE3 (Figures 4C and 4D). A3Bctd-R211K-BE3 produced 1.5-fold fewer sgRNA-independent off-target edits (8.7%) than A3Bctd-BE3 (13.0%) and comparable on-target frequencies (A3Bctd-R211K-BE3: 25.2%, A3Bctd-BE3: 26.4%). Similarly, the A3Bctd-D314R-BE3 and A3Bctd-D314H-BE3 had 3.4- and 3.7-fold lower (3.8% and 3.5%) sgRNA-independent off-target editing activities than A3Bctd-BE3 (13.0%) and retained high on-target activities (19.9% and 21.2%) (Figure 4D). In contrast, variants bearing L230K, N240A, and W281H single amino acid substitutions, which are not located in loops 1 or 7, did not show notable on- and off-target efficiencies (Figures 4A and 4D). These results demonstrated that mutations in A3Bctd loop 1 (R211K and T214V) and loop 7 (R311K, Y313F, D314R, D314H, and Y315M) reduced off-target activity on ssDNA, in agreement with the protein structure information (Figure 4A; Byeon et al., 2016; Hou et al., 2019; Shi et al., 2015; Siriwardena et al., 2015).

Next, we conducted next-generation sequencing analyses to further evaluate the precision of the A3Bctd-BE3 variants on their corresponding targeted sites (Figures 4E and S4A). We divided the editing products of A3Bctd-BE3 variants into 3 categories: single C-to-T reads, double C-to-T reads, and multiple C-to-T reads (≥3 C-to-Ts) (Figures 4E and S4A). Considering all of the sequencing reads across 4 on-target sites, A3A-BE3 and A3Bctd-BE3 yielded mainly multiple C edits, while Y313F, R311K, D314H, T214V, R211K, Y315M, R210K, and D314R produced fewer multiple C-to-T edits but more single C-to-T or double C-to-T edits than the original A3Bctd-BE3, and mutations in these variants were also located in loops 1 and 7 (Figure 4B). In addition, A3B-BE3, A3Bctd-T214S-BE3, A3Bctd-W281H-BE3, A3Bctd-N240A-BE3, A3Bctd-T214G-BE3, YEE-BE3, RK-BE3, and eA3A-BE3 mainly produced single or double C edits, but had low on-target editing efficiencies (Figures 4D, 4E, and S4B). Thus, the rational design approach described in this study allowed us to identify A3Bctd-BE3 variants with reduced sgRNA-independent off-target activities and improved editing precision within on-target sites (Figures 4D and 4E).

Combining the Mutations in A3Bctd to Increase the Specificity and Precision of Base Editing

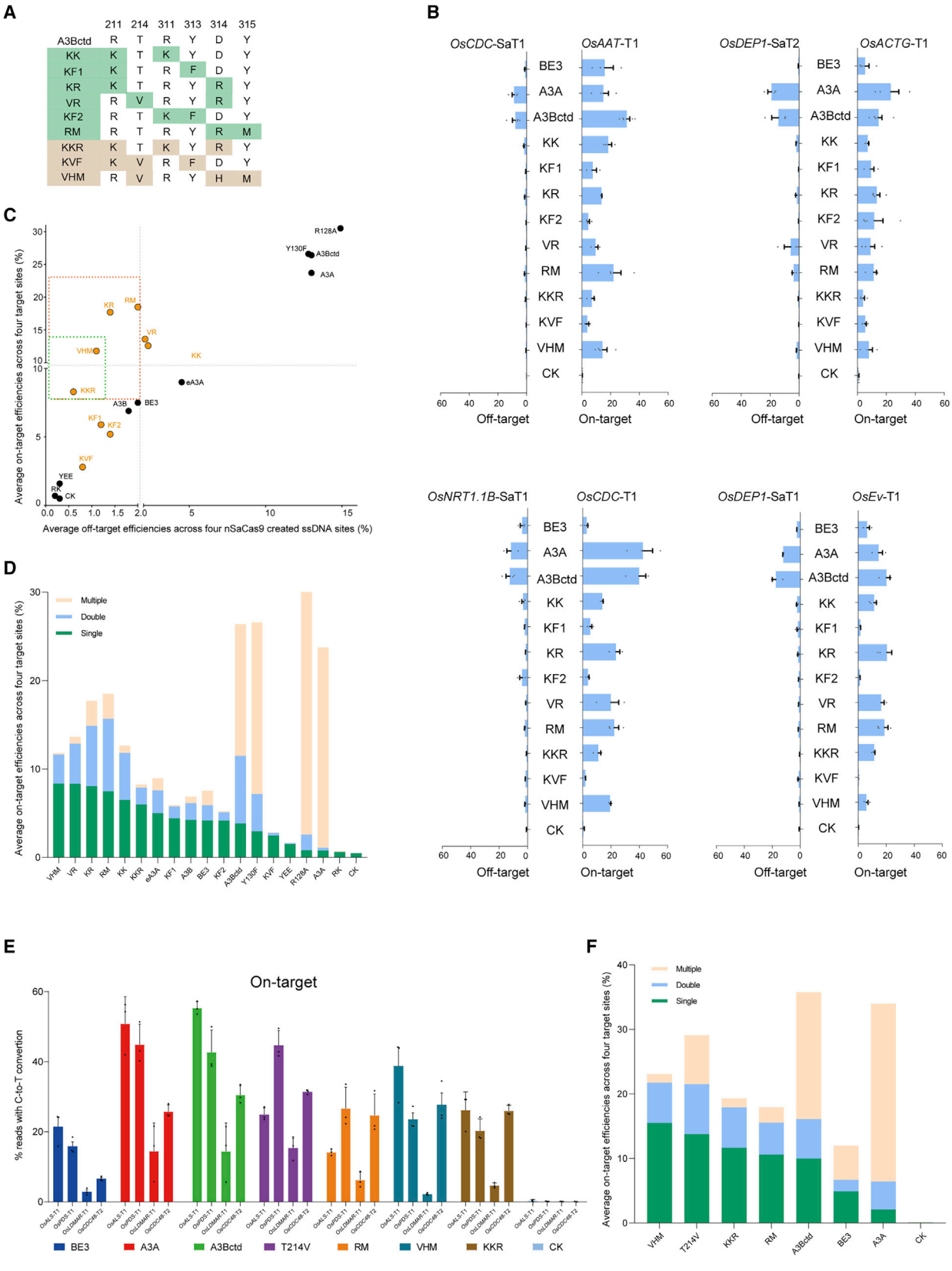
To test whether the specificity of A3Bctd could be improved further, we combined the 7 mutations in loops 1 and 7 associated with efficient on-target editing activity and lower off-target activity to produce 9 new A3Bctd-BE3 variants with double or triple amino acid changes (Figure 5A). The A3Bctd-BE3 variants with KK (12.6%), KR (17.7%), VR (13.6%), RM (18.5%), KKR (8.3%), and VHM (11.8%) mutations had similar or higher C-to-T editing efficiencies at 4 tested on-target sites than the original BE3 (7.5%), ranging from 8.3% to 18.5%, while the variants with KVF (2.8%), KF1 (5.9%), and KF2 (5.2%) mutations had lower C-to-T editing efficiencies ranging from 2.8% to 5.9% (Figures 5B, 5C, and S4B), although all of these base editors showed decreases in editing activities compared to A3Bctd-BE3 (26.4%). In terms of off-target activity, all variants had lower or comparable off-target activities to BE3 at all four off-target sites (Figures 5B and 5C). A3Bctd-BE3 variants with triple mutations VHM (1.1%), KKR (0.6%), or KVF (0.8%) had markedly reduced off-target activities compared with the original A3Bctd-BE3 (13.0%) and BE3 (2.0%). They were nearly as low as those of YEE-BE3 (0.3%) and RK-BE3 (0.2%). Furthermore, VHM and

(D) Average on-target C-to-T conversion frequencies across 4 on-target sites (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1, and *OsEv*-T1) versus the average sgRNA-independent off-target C-to-T conversion frequencies across 4 ssDNA regions (*OsCDC48*-SaT1, *OsDEP1*-SaT2, *OsNRT1.1B*-SaT1, and *OsDEP1*-SaT1) for 24 base editors. The red box shows the A3Bctd-BE3 variants with relatively high on-target efficiencies (>15%) and decreased sgRNA-independent off-target editing (<10%). Base editors created in this study are shown as orange dots and the reported base editors are shown as black dots.

(E) Types of edits created by A3Bctd-BE3 variants bearing single amino acid substitutions across 4 on-target sites (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1, and *OsEv*-T1). Single C-to-T changes are shown in green, double changes in blue, and multiple changes in orange. The frequencies are mean values at 4 sites. See also Figure S4.

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

CellPress



(legend on next page)

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

KKR variants (A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3) retained efficient on-target activities (11.8% and 8.3%, respectively), which are much higher than those of YEE-BE3 (1.6%) and RK-BE3 (0.7%) (Figures 5B and 5C). Thus, A3Bctd-BE3 variants with combinatory mutations exhibited different degrees of reduction of on-target and off-target activities; however, A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 maintained efficient on-target activities and carried out remarkably less off-target editing via sgRNA-independent deamination.

Next, we analyzed deep sequencing reads of 4 on-target sites to assess the editing precision of these combination variants (Figures 5D and S5A). A large proportion of the combined variants, especially A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3, produced almost no multiple C edits at 4 target sites (Figures 5D, S5A, and S5B), while they produced higher average numbers of single and double C-to-T edits than A3Bctd-BE3 across 4 target sites. A3Bctd-VHM-BE3 gave the most single C-to-T changes across all of the examined combined variants. We speculated that this improvement in editing precision has 2 explanations. One explanation of the improved editing precision from these variants could be that these mutations in loops 1 and 7 of A3Bctd decreased their ssDNA-binding affinity, leading to less binding to ssDNA, as shown in the previous study of YEE-BE3 in APOBEC1 (Kim et al., 2017). Alternatively, the improved editing precision could be a result of dinucleotide preference. Consistent with the latter hypothesis, we found the A3Bctd-VHM-BE3 variant displayed the strict sequence motif preference of TC and CC motifs, whereas A3Bctd-KKR-BE3 showed editing activities with more dinucleotide motifs (i.e., TC, CC ACs and GCs) across the 4 endogenous on-target sites (Figure S5C).

We further tested the editing efficiency and specificity of A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 at 4 additional targeted sites in rice protoplasts (*OsALS*-T1, *OsCDC48*-T2, *OsPDS*-T1, and *OsLDMAR*-T1). A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 again had efficient on-target activities (VHM: 38.8%, KKR: 26.2% in *OsALS*-T1; VHM: 27.7%, KKR: 26.0% in *OsCDC48*-T2; VHM: 23.6%, KKR: 20.3% in *OsPDS*-T1; and VHM: 2.2%, KKR: 4.7% in *OsLDMAR*-T1) compared to A3Bctd-BE3 (50.9% in *OsALS*-T1, 25.7% in *OsCDC48*-T2, 44.9% in *OsPDS*-T1, and 14.5% in *OsLDMAR*-T1) and BE3 (21.4% in *OsALS*-T1, 6.6% in *OsCDC48*-T2, 15.8% in *OsPDS*-T1, and 4.1% in *OsLDMAR*-T1) (Figures 5E and S5D). A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 also produced fewer multiple C-to-T edits and more single and double C-to-T edits than A3Bctd-BE3 (Figures 5F and S5E), and the dinucleotide preferences of these 4 sites were the same as the results of the preceding 4 target sites (Figures S5F and S5G). Thus, these combination variants showed efficient on-target activities and remarkably reduced sgRNA-independent off-target editing activities, and generated more precise edits at target sites.

Evaluating the sgRNA-Independent DNA Off-Target Activities of the New Base Editors by WGS

We conducted the WGS analyses to further evaluate the sgRNA-independent DNA off-target edits of 4 A3Bctd variants (A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3) that exhibited different degrees of off-target activities as evaluated by the nSaCas9-mediated orthogonal R-loop assay, along with BE3, A3A-BE3, and A3Bctd-BE3 *in planta* at the whole-genome level. We transformed these CBEs into rice calli via *Agrobacterium* transformation. In total, 51 T0 regenerated plants were analyzed by WGS along with 6 control plants, and 13 WT plants were used to filter out background mutations (Table S2). Identified changes were confirmed by Sanger sequencing at randomly selected sites with a 96% success rate (27/28) (Figures S6A and S6B; Table S3). Consistent with our previous study, the numbers of indels identified by WGS were similar between the base editor groups and the control group (Jin et al., 2019; Figure S6C), and the numbers of SNVs generated by the BE3, A3A-BE3, and A3Bctd-BE3 groups (averages of 475, 820, and 618, respectively) were significantly higher than the numbers generated by the control group (average of 335) (Figure 6A). In contrast, the A3Bctd variants exhibited a remarkable improvement in base editing specificity. Among the variants, A3Bctd-T214V-BE3 also produced fewer total SNVs than the original A3Bctd-BE3 group (averages of 411 and 618, respectively) (Figure 6A), and the variants A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3 showed no elevation in the total number of SNVs compared to the control group (averages of 306, 319, and 301, respectively) (Figure 6A).

Consistent with our previous research, the increased SNVs generated by the tested base editors were mainly C-to-T SNVs

Figure 5. Some of the Newly Engineered Base Editors Exhibit High Specificity and Increased Precision

(A) A3Bctd-BE3 variants with combined double or triple amino acid substitutions.

(B) Comparison of on-target and sgRNA-independent off-target editing frequencies for 9 doubly or triply mutated A3Bctd-BE3 variants. The frequencies of on-target C-to-T changes at 4 rice loci and sgRNA-independent off-target C-to-T changes in 4 ssDNA regions created by nSaCas9 in rice protoplasts were measured by deep sequencing. The values and error bars indicate means ± SEMs of 3–6 independent experiments.

(C) Average on-target C-to-T conversion frequencies across 4 on-target sites (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1, and *OsEv*-T1) versus average off-target C-to-T conversion frequencies across 4 ssDNA regions (*OsCDC48*-SaT1, *OsDEP1*-SaT2, *OsNRT1.1B*-SaT1, and *OsDEP1*-SaT1) for 18 base editors. The orange box shows these combinatory A3Bctd-BE3 variants with efficient on-target activities and decreased or comparable off-target editing compared with BE3. The green box shows two A3Bctd-BE3 variants with efficient on-target activities and lower off-target editing. The base editors created in this study are shown as orange dots and reported base editors are shown as black dots.

(D) Types of edits produced by A3Bctd-BE3 variants bearing double or triple mutations across 4 on-target sites. Single C-to-T changes are shown in green, double changes in blue, and multiple changes in orange. The frequencies are mean values of the 4 sites.

(E) Frequencies (percentages) of C-to-T conversions induced by 4 A3Bctd-BE3 variants (A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3) and BE3, A3A-BE3, and A3Bctd-BE3 across 4 new on-target sites (*OsALS*-T1, *OsCDC48*-T2, *OsPDS*-T1, and *OsLDMAR*-T1) in rice protoplasts measured by deep sequencing. The values and error bars indicate means ± SEMs of 3 independent experiments.

(F) Types of edits induced by 4 key A3Bctd-BE3 variants and BE3, A3A-BE3, and A3Bctd-BE3 across 4 new on-target sites in rice protoplasts measured by deep sequencing. Single C-to-T changes are shown in green, double changes in blue, and multiple changes in orange. The frequencies are mean values across 4 sites. See also Figure S5.

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

CellPress

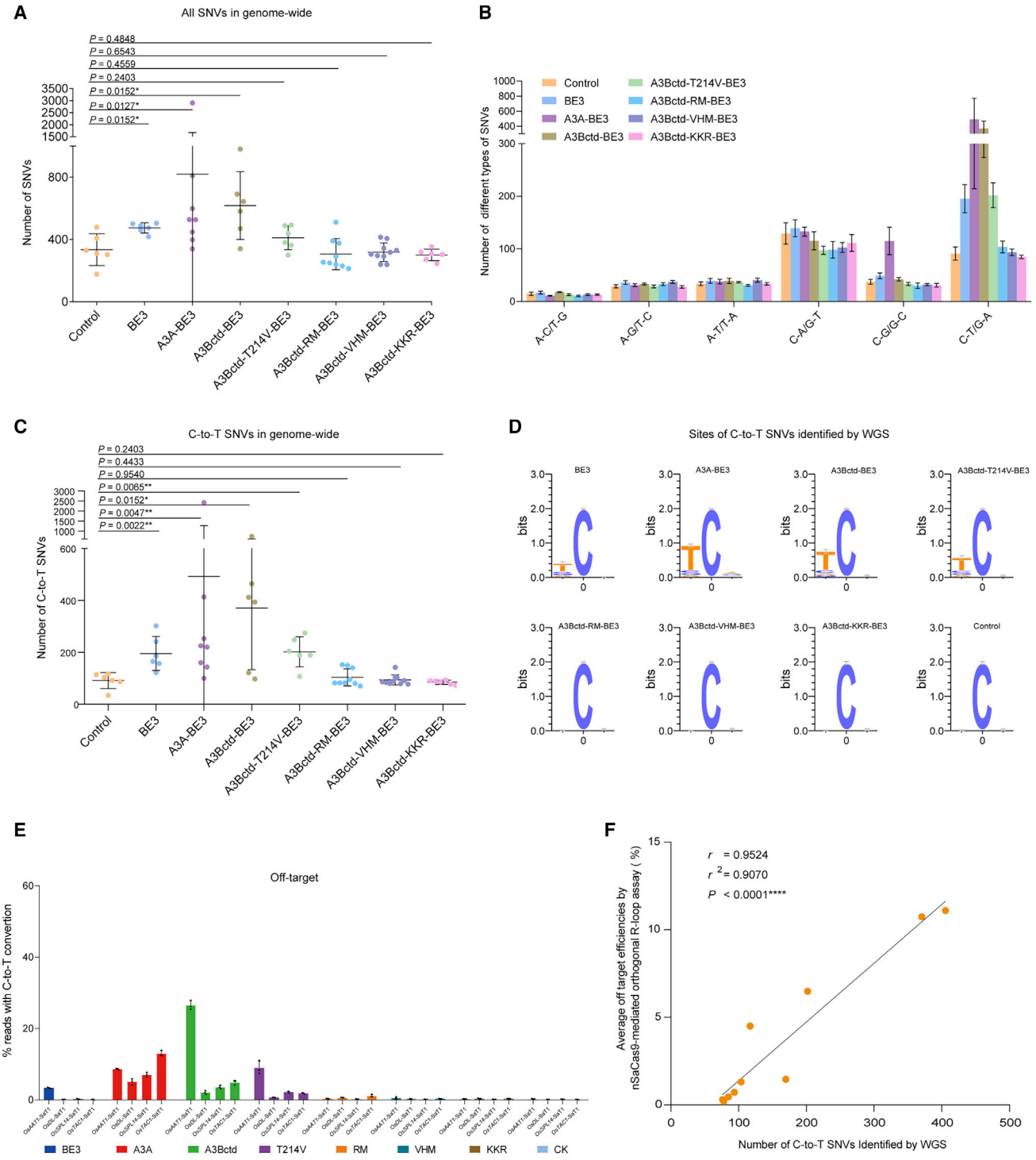


Figure 6. Further Evaluation of the DNA Off-Target Activities of the Selected A3Bctd-Based New Base Editors and Validation of the nSaCas9-Mediated Orthogonal R-Loop Assay in Plants by WGS

(A)–(C) Numbers of all SNVs (A), different types of SNVs (B), and C-to-T SNVs (C) identified in control, and BE3, A3A-BE3, A3Bctd-BE3, A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3 edited plants measured by WGS. The average numbers of total SNVs in control, BE3, A3A-BE3, A3Bctd-BE3, A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3 plants were 335, 475, 820, 618, 411, 306, 319, and 301, respectively, and the average numbers of C-to-T SNVs were 91, 195, 492, 371, 202, 204, 94, and 85, respectively. The statistical significance of differences was calculated using the 2-sided Mann-Whitney *U* test.

(legend continued on next page)

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

in tested base editors (Jin et al., 2019; Figure 6B). Further analyses indicated that the numbers of C-to-T transitional mutations generated by BE3, A3A-BE3, and A3Bctd-BE3 groups (averages of 195, 492, and 371, respectively) were significantly higher than in control plants (average of 91) (Figure 6C). Similarly, the single-mutation variant A3Bctd-T214V-BE3 also produced higher numbers of C-to-T SNVs than the control group (Figure 6C). In the double mutant variant A3Bctd-RM-BE3 transformation group, although the total number of SNVs was similar, a slight increase in C-to-T SNVs was observed between the variant and control groups (averages of 104 and 91, respectively; Figure 6C). Conversely, A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 produced comparable or even fewer SNVs (averages of 94 and 85, respectively) than the control group (average of 91) (Figure 6C). These results thus confirm that rationally designed A3Bctd-BE3 variants A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 were highly specific at the whole-genome level, as indicated by the results from the R-loop assay.

Because previous studies have shown that C-to-T edits induced by APOBEC-based CBEs are most likely to occur at TC motifs in plants (Zong et al., 2017, 2018), we extended our analyses to the sequences that are 5 bp upstream and downstream of the identified off-target sites with C-to-T SNVs identified by WGS (Figure 6D). The results showed that the A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3 variants displayed no sequence preference, as did the control group (Figure 6D). In contrast, consistent with the results from the previous study (Doman et al., 2020), the off-target C-to-T SNVs induced by BE3, A3A-BE3, A3Bctd-BE3, and A3Bctd-T214V-BE3 showed a sequence preference for TC dinucleotides (Figure 6D), although the level of sequence preference in BE3 appeared to be lower than those of A3A-BE3, A3Bctd-BE3, and A3Bctd-T214V-BE3 (Figure 6D), probably due to its lower overall base editing efficiency (Figure 5C).

To further validate the correlation between the R-loop assay and the WGS results, we selected an additional 4 nSaCas9 target sites (Os*DL*-SaT1, Os*AAT1*-SaT1, Os*SPL14*-SaT1, and Os*TAC1*-SaT1) to test the new CBEs variants A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3 (Figure 6E; Table S1). Each CBE was co-transformed with nSaCas9 and its Sa-sgRNA into rice protoplasts. Deep sequencing showed that A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 yielded almost undetectable levels of sgRNA-independent off-target edits compared with the control group across 4 nSaCas9 sites (Figure 6E). A correlational analysis based on the average off-target activities of 10 base editors tested by WGS (Figure 6C) and by R-loop off-target sites (Figures S6D and S6E) used in this study showed that the results of the nSaCas9-mediated orthogonal R-loop assay and of WGS results were strongly correlated ($r^2 = 0.9070$, ****$p < 0.0001$) (Figure 6F). This again indicated that the nSaCas9-mediated orthogonal R-loop assay can effectively evaluate the off-target activities of base editors.

DISCUSSION

In this study, we tested and optimized the nSaCas9-mediated orthogonal R-loop assay in plant cells and validated it with WGS data. Using this assay, we screened a series of rationally designed A3Bctd-BE3 variants and identified 2 variants, A3Bctd-KKR-BE3 and A3Bctd-VHM-BE3, that retained efficient on-target activities but with dramatically reduced off-target activities (Figures S6D and S6E; Tables S4 and S5). These 2 novel A3Bctd-BE3 variants not only extend the toolbox for precise genome editing but also provide new insights into the factors affecting the specificity and precision of base editing systems.

WGS analysis has been the gold standard for direct, thorough, and unbiased assessment of unexpected off-target edits by Cas proteins and base editors (Iyer et al., 2018; Tang et al., 2018; Willi et al., 2018; Jin et al., 2019; Li et al., 2019; Zuo et al., 2019; Lee et al., 2020). However, its major drawbacks, namely low-throughput, expense, and length have prevented this approach from being used widely (Iyer et al., 2018; Li et al., 2019; Willi et al., 2018). The orthogonal R-loop assay avoids these difficulties by reducing off-target assessment to a limited number of designated loci, and thus greatly reduces the cost and improves the throughput capacity (Doman et al., 2020). It also allows comparative and quantitative evaluation of CBEs with more statistical reliability, thereby overcoming the limitations of the WGS assessment (Doman et al., 2020). Thus far, the orthogonal R-loop assay had only been tested in *E. coli* and human cells. Moreover, its accuracy had not been fully validated in plants and by large-scale individual WGS data.

In this study, we extended the use of the orthogonal R-loop assay to plants and tested 3 Cas9-orthogonal proteins (dLbCpf1, dSaCas9, or nSaCas9). We found that nSaCas9 but not dLbCas12a or dSaCas9 was effective in creating ssDNA regions for evaluating off-target base editing in plant cells, whereas Doman et al. (2020) found that dSaCas9 had this ability in human cells.

One explanation could be that rice protoplasts may possess a robust repair system but not DNA replication. Plant protoplasts, isolated from leaves, do not continue dividing during the transformation process, and there is no DNA replication. Our previous

(D) The sequence context of C · G-to-T · A SNVs in edited and control plants identified by WGS. Sequence conservation at positions from −1 to +1 is shown, with the mutated C at position 0. The sequence logos were produced on the online website http://weblogo.threeplusone.com/ (Crooks et al., 2004).

(E) Frequencies (percentage) of C-to-T conversions induced by 4 key A3Bctd-BE3 variants and other editors across 4 new off-target sites (*OsDL*-SaT1, *OsAAT1*-SaT1, *OsSPL14*-SaT1, and *OsTAC1*-SaT1) in rice protoplasts measured by deep sequencing. The values and error bars indicate means ± SEMs of 3 independent experiments.

(F) Analysis of the correlation between C-to-T off-target activities in the nSaCas9-mediated orthogonal R-loop assay and average numbers of C-to-T SNVs identified by WGS for various editors. Off-target activities were calculated from the average C-to-T efficiency at 4 nSaCas9 created ssDNA regions induced by RK-BE3, YEE-BE3, and eA3A-BE3; off-target activities were calculated from the average C-to-T efficiencies of 8 nSaCas9-created ssDNA regions induced by BE3, A3A-BE3, A3Bctd-BE3, A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3. The statistical significance of differences was calculated using the 2-sided Mann-Whitney *U* test. Pearson's correlation coefficient (r) and p values (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$) are shown.
See also Figure S6.

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

study showed that when we used BE2 (APOBEC1-dCas9-UGI) to edit targeted sites, we found almost no C-to-T edits (Zong et al., 2017). We reasoned that this was largely due to the absence of DNA replication in the protoplasts, in contrast to dividing human cells. In the R-loop assays in the present work, we used dSaCas9 to create ssDNA regions, which did not produce nicks when the tested cytidine deaminases induced C-to-U changes, so we could not detect C-to-T mutations in protoplasts. When we used nSaCas9 to produce ssDNA and nicks in the ssDNA in the opposite strand, exactly as with BE3, then, if the tested cytidine deaminases converted C to U in the nSaCas9 ssDNA region, the DNA nick in the opposite strand made by nSaCas9 could stimulate MMR or long-patch BER so as to preferentially change the U:G mismatch into the desired U:A and T:A products. Thus, nCas9-CBEs induces C-to-T edits in dividing cells (DNA replication) and non-dividing cells (no-DNA replication), whereas dCas9-CBEs do not induce efficient C-to-T editing in non-replicating cells. It is suggested that only nickase-based BEs possess high on-target efficiencies in plants; however, the nicking requirement could limit the application of some genome-editing reagents without non-target nickase variants in BE and R-loop systems, such as Cas12a, Cas12b, and TALEN.

The dSaCas9- and nSaCas9-based R-loop assays create extended ssDNA regions at target sites that theoretically do not exist in the normal genomic environment, and prolonged exposure of the ssDNA region may to some extent amplify the off-target deamination activities. This suggests that if the d/nSaCas9-based R-loop assay fails to detect any off-target editing activity of a given base editor, then that base editor would not produce any off-target edits in the normal genome. Our present data also support this point: we found that A3Bctd-KKR-BE3 and A3Bctd-VHM-BE3 variants retained very low off-target activities by R-loop assay (Figures 5C and 6E), whereas in WGS experiments, they failed to induce any off-target SNVs at all (Figure 6C). This result suggests that the R-loop assay is a high-throughput assay that can be used to detect sgRNA-independent off-target activities of base editors over a wide genomic range.

In the R-loop assay, our data also showed that the same set of base editors had different levels of off-target activities at different target sites. This could reflect the sequence preferences of various deaminases or the composition of targeted sequences. For example, A3A-BE3 and A3Bctd-BE3 had very high off-target activities at all 8 off-target sites, while A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 had almost no off-target activities at these off-target sites. However, the original BE3 produced changed off-target edits across 8 off-target sites, perhaps due to the deamination preference of the APOBEC1 deaminase. The Cs within tested sites that we selected covered nearly all 20 nucleotides in the protospacer and all 4 types of dinucleotide pairs (TC, AC, CC, GC), both at on-target sites (total numbers of AC, CC, GC, and TC across 8 on-target sites 13, 20, 14, and 9) and off-target sites (total numbers of AC, CC, GC, and TC across 8 off-target sites are 8, 12, 9, and 15) (Figure S6F). In view of the different deamination characteristics of the different deaminases, we suggest that when using the R-loop assay to detect sgRNA-independent DNA off-target edits, at least 4–6 sites should be tested to compensate for the biases imposed by the target sites.

The objective of this study was to develop base editors with higher on-target efficiencies and lower off-target activities than the original BE3. Protein engineering by the structure-based rational design of deaminases has been proven to be an efficient method for improving the efficiency, specificity, and precision of designer nucleases. We hypothesized that modifying the ssDNA-binding domain of a highly efficient deaminase may reduce its ssDNA-binding affinity and thus lower its sgRNA-independent off-target activity while retaining high on-target activity. In fact, the highly specific BE3 variants developed previously, YEE-BE3, RK-BE3, and eA3A-BE3 (Kim et al., 2017), had amino acid mutations in their ssDNA- or RNA-binding domains. In the present work, we focused on A3Bctd-BE3, which has higher on-target activity than the original BE3 and its derivatives mentioned above. Of 16 A3Bctd-BE3 variants with single amino acid substitutions in the ssDNA-binding domain, 8 had reduced sgRNA-independent off-target activities while retaining high base editing efficiencies. Moreover, 7 of these variants had mutations in loops 1 and 7, the key ssDNA-binding domains in A3Bctd (Byeon et al., 2016; Hou et al., 2019; Shi et al., 2015; Siriwardena et al., 2015). When the amino acid substitutions in these loop regions were further combined to form the variants with triple amino acid substitutions, VHM and KKR, both had even lower off-target and higher on-target activities than BE3.

Nearly all of these variants with mutations in the ssDNA-binding domains also had altered on-target editing characteristics, with somewhat decreased on-target editing efficiencies but enhanced editing precision (Figures S6D and S6E). The double mutant variants RM, KR, VR, and KK produced markedly reduced proportions of multiplex C edits, while the triple mutants mainly generated single and double C edits (Figures 4E, 5D, and S6E). The increased precision of editing may be due to decreased on-target editing activity, stronger dinucleotide preferences, or the sequences surrounding target sites. Defining the editing characteristics of large numbers of base editors based on 4 or 8 endogenous sites is not very accurate. Research aimed at defining the reason for the precise editing behavior of these new base editors would be valuable and is needed, but for large numbers of base editors, amplicon deep sequencing assays would be too labor intensive to permit the testing of enough endogenous target sites, and high-throughput assays would be needed to analyze editing window or dinucleotide preferences.

In summary, we have extended the use of the nSaCas9-mediated orthogonal R-loop assay for assessing the sgRNA-independent off-target editing activity of CBEs in plants and validated it by WGS. The assay allowed us to rapidly screen a series of rationally designed A3Bctd-BE3 variants for reduced sgRNA-independent off-target activity, and yielded the VHM and KKR variants of A3Bctd-BE3, which exhibited efficient C-to-T base editing with almost no sgRNA-independent off-target activity. The framework presented in this study could be widely used to assess the off-target activities of newly developed base editors and for screening work to develop base editors with improved specificity and precision. We anticipate that the improved base editors, combined with a variety of different Cas proteins, will

permit rapid expansion of the precision and targeting range of genome editing.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Plasmid construction
  - Rice protoplasts transfection
  - Flow cytometry analysis
  - Analysis of randomly selected variants
  - Amplicon deep sequencing and data analysis
  - *Agrobacterium*-mediated transformation of rice
  - *De novo* assembly of the Zhonghua 11 genome
  - Whole-genome sequencing and data analysis
  - Analysis of BE3 and BE3 variants induced genome-wide C·G-to-T·A SNVs in plants
  - hA3Bctd structural analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.07.005.

ACKNOWLEDGMENTS

We thank Dr. Fei Lu (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences [IGDB, CAS]) for advice in the statistics analysis. This work was supported by grants from the National Transgenic Science and Technology Program (2019ZX08010-003 and 2019ZX08010-001), the Strategic Priority Research Program of the CAS (Precision Seed Design and Breeding, XDA24020102), the National Natural Science Foundation of China (31788103), the Youth Innovation Promotion Association of the Chinese Academy of Sciences (2020000003), and by a grant from the United Soybean Board (1920-152-0131-B) to F.Z.

AUTHOR CONTRIBUTIONS

Y.W. and C.G. designed the experiments. S.J., H.F., and Z.Z. performed most of the experiments. Y.L., S.J., and S.G. performed the WGS analysis. J.L. performed the rice transformation experiments. Y.-H.C. analyzed the structure of the A3Bctd protein. Y.W. and C.G. supervised the project. S.J., H.F., Z.Z., Y.W., F.Z., and C.G. wrote the manuscript.

DECLARATION OF INTERESTS

The authors have submitted two patent applications based on the results reported in this article.

Received: February 12, 2020
Revised: May 8, 2020
Accepted: July 2, 2020
Published: July 27, 2020

REFERENCES

Byeon, I.J., Byeon, C.H., Wu, T., Mitra, M., Singer, D., Levin, J.G., and Gronenborn, A.M. (2016). Nuclear magnetic resonance structure of the APOBEC3B catalytic domain: structural basis for substrate binding and DNA deaminase activity. Biochemistry *55*, 2944–2959.

Chadwick, A.C., Wang, X., and Musunuru, K. (2017). *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.

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. Annu. Rev. Plant Biol. *70*, 667–697.

Cheng, T.-L., Li, S., Yuan, B., Wang, X., Zhou, W., and Qiu, Z. (2019). Expanding C-T base editing toolkit with diversified cytidine deaminases. Nat. Commun. *10*, 3612.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190.

Doman, J.L., Raguram, A., Newby, G.A., and Liu, D.R. (2020). Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base. Nat. Biotechnol. *38*, 620–628.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

Gehrke, J.M., Cervantes, O., Clement, M.K., Wu, Y., Zeng, J., Bauer, D.E., Pinello, L., and Joung, J.K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nat. Biotechnol. *36*, 977–982.

Grünewald, J., Zhou, R., Garcia, S.P., Iyer, S., Lareau, C.A., Aryee, M.J., and Joung, J.K. (2019a). Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature *569*, 433–437.

Grünewald, J., Zhou, R., Iyer, S., Lareau, C.A., Garcia, S.P., Aryee, M.J., and Joung, J.K. (2019b). CRISPR DNA base editors with reduced RNA off-target and self-editing activities. Nat. Biotechnol. *37*, 1041–1048.

Hou, S., Silvas, T.V., Leidner, F., Nalivaika, E.A., Matsuo, H., Kurt Yilmaz, N., and Schiffer, C.A. (2019). Structural analysis of the active site and dna binding of human cytidine deaminase APOBEC3B. J. Chem. Theory Comput. *15*, 637–647.

Iyer, V., Boroviak, K., Thomas, M., Doe, B., Riva, L., Ryder, E., and Adams, D.J. (2018). No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice. PLOS Genet. *14*, e1007503.

Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.L., Zhang, F., and Gao, C. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science *364*, 292–295.

Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. *35*, 371–376.

Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell *168*, 20–36.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. *27*, 722–736.

Lee, H.K., Willi, M., Miller, S.M., Kim, S., Liu, C., Liu, D.R., and Hennighausen, L. (2018). Targeting fidelity of adenine and cytosine base editors in mouse embryos. Nat. Commun. *9*, 4804.

Lee, H.K., Smith, H.E., Liu, C., Willi, M., and Hennighausen, L. (2020). Cytosine base editor 4 but not adenine base editor generates off-target mutations in mouse embryos. Commun. Biol. *3*, 19.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, J., Manghwar, H., Sun, L., Wang, P., Wang, G., Sheng, H., Zhang, J., Liu, H., Qin, L., Rui, H., et al. (2019). Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. Plant Biotechnol. J. *17*, 858–868.

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.

Liang, P., Ding, C., Sun, H., Xie, X., Xu, Y., Zhang, X., Sun, Y., Xiong, Y., Ma, W., Liu, Y., et al. (2017). Correction of β-thalassemia mutant by base editor in human embryos. Protein Cell *8*, 811–822.

McGrath, E., Shin, H., Zhang, L., Phue, J.N., Wu, W.W., Shen, R.F., Jang, Y.Y., Revollo, J., and Ye, Z. (2019). Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. Nat. Commun. *10*, 5353.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, 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.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*, 6305.

Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. *19*, 770–788.

Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics *19*, 460.

Ryu, S.M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.T., Kim, H.S., Kim, D.E., Lee, H., Chung, E., and Kim, J.S. (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nat. Biotechnol. *36*, 536–539.

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., Zhang, K., Liu, J., Voytas, D.F., Zheng, X., et al. (2013). Rapid and efficient gene modification in rice and Brachypodium using TALENs. Mol. Plant *6*, 1365–1368.

Shi, K., Carpenter, M.A., Kurahashi, K., Harris, R.S., and Aihara, H. (2015). Crystal structure of the DNA deaminase APOBEC3B catalytic domain. J. Biol. Chem. *290*, 28120–28130.

Siriwardena, S.U., Guruge, T.A., and Bhagwat, A.S. (2015). Characterization of the catalytic domain of human APOBEC3B and the critical structural role for a conserved methionine. J. Mol. Biol. *427*, 3042–3055.

St Martin, A., Salamango, D., Serebrenik, A., Shaban, N., Brown, W.L., Donati, F., Munagala, U., Conticello, S.G., and Harris, R.S. (2018). A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC-Cas9 or cleavage by Cas9 in living cells. Nucleic Acids Res. *46*, e84.

Tan, J., Zhang, F., Karcher, D., and Bock, R. (2020). Expanding the genome-targeting scope and the site selectivity of high-precision base. Nat. Commun. *11*, 629.

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J.C., Schnable, P.S., Lyons, E., and Lu, J. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. *16*, 3.

Tang, X., Liu, G., Zhou, J., Ren, Q., You, Q., Tian, L., Xin, X., Zhong, Z., Liu, B., Zheng, X., et al. (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. Genome Biol. *19*, 84.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS ONE *9*, e112963.

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.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO Applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. *35*, 543–548.

Willi, M., Smith, H.E., Wang, C., Liu, C., and Hennighausen, L. (2018). Mutation frequency is not increased in CRISPR-Cas9-edited mice. Nat. Methods *15*, 756–758.

Wilm, A., Aw, P.P.K., Bertrand, D., Yeo, G.H.T., Ong, S.H., Wong, C.H., Khor, C.C., Petric, R., Hibberd, M.L., and Nagarajan, N. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. *40*, 11189–11201.

Xue, C., Zhang, H., Lin, Q., Fan, R., and Gao, C. (2018). Manipulating mRNA splicing by base editing in plants. Sci. China Life Sci. *61*, 1293–1300.

Zhang, Y., Li, J., and Gao, C. (2016). Generation of stable transgenic rice (Oryza sativa L.) by *Agrobacterium*-mediated transformation. Curr. Protoc. Plant Biol. *1*, 235–246.

Zhang, R., Liu, J., Chai, Z., Chen, S., Bai, Y., Zong, Y., Chen, K., Li, J., Jiang, L., and Gao, C. (2019). Generation of herbicide tolerance traits and a new selectable marker in wheat using base editing. Nat. Plants *5*, 480–485.

Zhang, T., Luo, Y., Liu, K., Pan, L., Zhang, B., Yu, J., and Hu, S. (2011). BIGpre: a quality assessment package for next-generation sequencing data. Genom. Proteom. Bioinf. *9*, 238–244.

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.

Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., Qiu, J., and Gao, C. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nat. Biotechnol. *36*, 950–953.

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.L., Wang, D., and Gao, C. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat. Biotechnol. *35*, 438–440.

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L.M., Li, Y., and Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science *364*, 289–292.

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
DH5a *E.coli* competent cells	2nd labs	CAT#DL1001M
AGL1 *Agrobacterium* competent cells	2nd labs	CAT#AE1020M
Chemicals, Peptides, and Recombinant Proteins		
Ampicillin	Transgen	CAT#GG101-01
Kanamycin	Transgen	CAT#GG201-01
Rifampicin	ThermoFisher	CAT#T0681-200
Hygromycin	Roche	CAT#10843555001
Carbenicillin	Yeasen Biotech Co	CAT#60202ES60
Ethanol	Sigma-Aldrich	CAT#E7023
Tryptone	OXOID	CAT#LP0042
NaCl	IGDB	CAT#YC-SJ02027
Agarose	Invitrogen	CAT#16500500
Yeast Extract	OXOID	CAT#LP0021
2-(N-morpholino) ethanesulfonic acid(MES)	Sigma-Aldrich	CAT#M8250
Mannitol	Sigma-Aldrich	CAT#M4125
CaCl2	Sigma-Aldrich	CAT#C7902
KCl	Sigma-Aldrich	CAT#P3911
MgCl2	Sigma-Aldrich	CAT#M9272
BSA	Sigma-Aldrich	CAT#A6793
Cellulase R10	Yakult Pharmaceutical Industry	CAT#L0012-10
Macerozyme R10	Yakult Pharmaceutical Industry	CAT#L0021-5
PEG4000	Sigma-Aldrich	CAT#95904
Murashige and Skoog (MS) salt, including vitamins	Phytotech	CAT#M519
Phytagel	Sigma-Aldrich	CAT#P8169-5Kg
Sucrose	Sigma-Aldrich	CAT#V900116
2,4-Dichlorophenoxyacetic acid	Sigma-Aldrich	CAT#D7299
CuSO4·5H2O	Sigma-Aldrich	CAT#C3036
KOH	Sigma-Aldrich	CAT#484016
Naphthaleneacetic acid (NAA)	Sigma-Aldrich	CAT#N0640
Casein Hydrolysate	Duchefa	CAT#C1301.0500
Inositol	Sangon	CAT#A600536-0500
Proline	Sangon	CAT#A600923-0500
L–glutamine	Sigma	CAT#G3126-100 g
Thiamine hydrochloride	Sigma	CAT#T1270-25G
Pyridoxine hydrochloride	Sigma	CAT#P6280-25G
Nicotinic acid BioReagent	Sigma	CAT#N0761-100G
Mannitol	Sigma	CAT#M1902-5KG
N6-Benzyladenine	Cayman	CAT#21711
Kinetin	Sigma	CAT#K3378-5G
Naphthalene acetic acid	Sigma	CAT#N0640
PCR enzyme KOD FX	TOYOBO	CAT#KFX-101
Ethyl alcohol	Aladdin	CAT#E111993-4 × 4L
FastPfu DNA polymerase	TransGen Biotech	CAT#AP221-03
10 × Fastdigest Green Buffer	Fermentas/Thermo Scientific	CAT#B72

(*Continued on next page*)

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AxyPrep DNA gel extraction kit	Axygen	CAT#AP-GX-250
Trans2K PlusII DNA marker	TransGen Biotech	CAT#BM121-02
10 × TAE buffer	Cellgro	CAT#46-010-CM
T4 DNA ligase	Fermentas/Thermo Scientific	CAT#EL0011
Critical Commercial Assays		
Clone Express®II One Step Cloning Kit	Vazyme Biotech co	CAT#C112-02
Wizard Plus midipreps	Promega	CAT#A7640
2 × Rapid Taq Master Mix	Vazyme Biotech co	CAT#P222-03
DNA Quick Plant System	Tiangen Biotech	CAT#DP321-03
TransStart FastPfu Fly DNA Polymerase	TransGen Biotech	CAT#AP231-13
VAHTSTM Universal DNA Library Prep Kit for Illumina® V3	Vazyme Biotech co	CAT#ND607-01
VAHTSTM Multiplex Oligos Set 4/5 for Illumina®	Vazyme Biotech co	CAT#N321
Deposited Data		
All deep sequencing and whole-genome sequencing data for this study	This study	PRJNA602608 (NCBI Bioproject)
Experimental Models: Organisms/Strains		
Japonica rice (Oryza sativa) variety Zhonghua11	This study	N/A
Oligonucleotides		
See Tables S1, S3, S4, S5 for sequences of oligonucleotides used in this study	N/A	N/A
Recombinant DNA		
pJIT163-Ubi-dLbCpf1	This study	N/A
pJIT163-Ubi-dSaCas9	This study	N/A
pJIT163-Ubi-nSaCas9	This study	N/A
A3A-PBE	Zong et al., 2018	Addgene # 19768
pnCas9-PBE	Zong et al., 2017	Addgene #98164
pYEE (W90Y+R126E+R132E)-BE3	This study	N/A
pRK (R33A+K34A)-BE3	This study	N/A
peA3A (N57G)-BE3	This study	N/A
pA3B-BE3	This study	N/A
pA3Bctd-BE3	This study	N/A
pA3C-BE3	This study	N/A
pA3D-BE3	This study	N/A
pA3F-BE3	This study	N/A
pA3G-BE3	This study	N/A
pA3H-BE3	This study	N/A
pA3A-R128A-BE3	This study	N/A
pA3Bctd-Y130F-BE3	This study	N/A
pA3Bctd-R210A-BE3	This study	N/A
pA3Bctd-R210K-BE3	This study	N/A
pA3Bctd-T214G-BE3	This study	N/A
pA3Bctd-T214V-BE3	This study	N/A
pA3Bctd-T214S-BE3	This study	N/A
pA3Bctd-T214C-BE3	This study	N/A
pA3Bctd-L230K-BE3	This study	N/A
pA3Bctd-N240A-BE3	This study	N/A
pA3Bctd-W281H-BE3	This study	N/A

(*Continued on next page*)

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pA3Bctd-F308K-BE3	This study	N/A
pA3Bctd-R311K-BE3	This study	N/A
pA3Bctd-Y313F-BE3	This study	N/A
pA3Bctd-D314R-BE3	This study	N/A
pA3Bctd-D314H-BE3	This study	N/A
pA3Bctd-Y315M-BE3	This study	N/A
pA3Bctd-KK(R211K+R311K)-BE3	This study	N/A
pA3Bctd-KF1(R211K+Y313F)-BE3	This study	N/A
pA3Bctd-KR(R211K+D314R)-BE3	This study	N/A
pA3Bctd-VR(T214V+D314R)-BE3	This study	N/A
pA3Bctd-RM(D314R+Y315M)-BE3	This study	N/A
pA3Bctd-KKR(R211K+R311K+D314R)-BE3	This study	N/A
pA3Bctd-KVF(R211K+T214V+Y313F)-BE3	This study	N/A
pA3Bctd-VHM(T214V+D314H+Y315M)-BE3	This study	N/A
pU3-gRNA	Shan et al., 2013	Addgene # 53063
pJIT163-Ubi-LbcrRNA	This study	N/A
pJIT163-Ubi-Sa-sgRNA	This study	N/A
pH-RK-BE3	This study	N/A
pH-YEE-BE3	This study	N/A
pH-nCas9-PBE	Zong et al., 2017	Addgene # 98163
pH-A3A-PBE	Zong et al., 2018	Addgene # 119774
pH-eA3A-BE3	This study	N/A
Software and Algorithms		
BIGpre-2.0.2	N/A	https://sourceforge.net/projects/bigpre/
Canu v1.8	Koren et al., 2017	https://canu.readthedocs.io/en/latest/index.html
Pilon v1.23	Walker et al., 2014	https://github.com/broadinstitute/pilon
ALLMAPS v1.0	Tang et al., 2015	https://github.com/tanghaibao/jcvi/wiki/ALLMAPS
PurgeHaplotigs	Roach et al., 2018	https://bitbucket.org/mroachawri/purge_haplotigs/src/master/
BUSCO v3.0	N/A	https://busco-archive.ezlab.org/v3/
BWA-MEM v0.7.17	Li and Durbin, 2010	https://github.com/lh3/bwa
Picard tools v1.119	N/A	https://github.com/broadinstitute/picard
GATK v3.5	McKenna et al., 2010	https://github.com/broadgsa/gatk/releases/tag/3.5
LoFreq v2.1.2	Wilm et al., 2012	https://sourceforge.net/projects/lofreq/
Strelka2 v2.9.10	Kim et al., 2018	https://github.com/Illumina/strelka
PyMOL	PyMOL Team	https://pymol.org/2/
WebLogo 3	Crooks et al., 2004	http://weblogo.threeplusone.com/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact: Caixia Gao (cxgao@genetics.ac.cn).

Materials Availability

All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

Data and Code Availability

Original data of this study have been deposited to NCBI Sequence Read Archive database: PRJNA602608.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

DH5a *E.coil* component cells (2nd labs) were used for plasmid DNAs amplifying for rice protoplast transfection experiments. And the transfected *E.coil* cells were grown at 37°C in Lysogeny Broth (LB) medium supplemented with 100 mg/mL ampicillin overnight. AGL1 *Agrobacterium* competent cells (2nd labs) cells were used for *Agrobacterium*-mediated transformation of rice callus cells. And the transfected *Agrobacterium* cells were grown at 28°Cin Lysogeny Broth (LB) medium supplemented with 50 mg/mL kanamycin and 50 mg/mL rifampicin.

METHOD DETAILS

Plasmid construction

To construct the pJIT163-Ubi-dSaCas9, pJIT163-Ubi-nSaCas9, pJIT163-Ubi-dLbCpf1, pJIT163-OsU3-Sa-sgRNA and pJIT163-Ubi-LbcrRNA vectors, the SaCas9 and LbCpf1 sequences were codon-optimized for cereal plants and synthesized commercially (Sangon Biotech (Shanghai) co., Ltd). The OsU3-Sa-sgRNA and HNH-LbcrRNA-HDV sequence were directly synthesized commercially without codon optimization. Mismatch PCR experiments were performed to induce the D10A, D10A+N580A, and D832A point mutations in commercially synthesized SaCas9 and LbCpf1 fragments, respectively, yielding nSaCas9, dSaCas9 and dLbCpf1 fragments. These mutated fragments were then cloned into HindIII- and EcoRI-linearized pJIT163-Ubi-GFP vectors using the plasmid recombination kit, Clone Express (Vazyme Biotech co., ltd, C112-02), to yield pJIT163-Ubi-nSaCas9. pJIT163-Ubi-dSaCas9, pJIT163-Ubi-dLbCpf1, pJIT163-OsU3-Sa-sgRNA and pJIT163-Ubi-LbcrRNA expression vectors.

To construct APOBEC3-BE3 variant expression vectors, human A3B, A3Bctd, A3C, A3D, A3F, A3G, A3H deaminase were codon-optimized for cereal plants and synthesized commercially (Genewiz, Suzhou, China) (Table S6). The synthesized DNA fragments were PCR amplified and cloned into HindIII- and AflII-linearized A3A-PBE vector. The RK, YEE, eA3A and all the A3Bctd-BE3 variants were produced by mismatch PCR (Table S6). For example, two primer sets, (A3B-GB-F/A3B-R210A-R) and (A3B-R210A-F/A3B-GB-R) (Table S4), were used to create the R210A-mutated fragment, A3Bctd-R210A-BE3. Then this fragment was cloned into HindIII- and AflII-linearized A3Bctd-BE3 expression vector to yield pA3Bctd-R210A-BE3 expression vector. All the other A3Bctd-BE3 variants were constructed in the same way. The plant transformation vectors, pH-RK-BE3, pH-YEE-BE3, pH-eA3A-BE3, pH-A3Bctd-BE3, pH-A3Bctd-T214V-BE3, pH-A3Bctd-RM-BE3, pH-A3Bctd-VHM-BE3 and pH-A3Bctd-KKR-BE3 were also made by the same way on the published pH-CBE backbone (Jin et al., 2019). All the sgRNA cloning experiments were performed as previously described (Shan et al., 2013), all primer sets used were synthesized by Beijing Genomics Institute (BGI) and listed in Table S4.

Rice protoplasts transfection

All the plant protoplasts used in this study were prepared from Japonica rice (*Oryza sativa*) variety Zhonghua11. Rice protoplasts isolation and transformation were performed as previously described (Shan et al., 2013). Plasmids (10 μg per construct) were introduced by PEG-mediated transfection. For the experiments using nSaCas9-mediated orthogonal R-loop assay to measure the off-target activities of CBEs, 10 μg SpCas9 sgRNA plasmid, 10 μg base editor plasmid, 10 μg SaCas9 sgRNA plasmid and 10μg nSaCas9 plasmid were co-transfected into rice protoplasts by PEG-mediated transfection. The transfected protoplasts were incubated at 23°C for 48 h. After incubation, genomic DNA was extracted with a Plant Genomic DNA Kit (TIANGEN) used for library construction and deep amplicon sequencing.

Flow cytometry analysis

Flow cytometry analysis was carried out using a FACS Aria III (BD Biosciences). Each of eight human APOBEC3-BE3s expression plasmids (A3A, A3B, A3Bctd, A3C, A3D, A3F, A3G and A3H-BE3s) or the BE3 expression plasmid, together with the guide RNA expression plasmids and fluorophore expression plasmids were transfected into rice protoplasts (Zong et al., 2017). All samples were sorted for GFP-positive cells, and they were counted.

Analysis of randomly selected variants

A DNA Quick Plant System (Tiangen Biotech, Beijing, China) was used to extract rice genomic DNA. Specific primers were used to amplify the targeted rice genomic sequences; the amplicons products were purified with an EasyPure PCR Purification Kit (TransGen Biotech, Beijing, China) and the purified DNAs were quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The purified amplicons were Sanger sequenced to verify randomly selected mutations caused by base editors. The primers used for amplifying various genomic sites are listed in Table S3.

Amplicon deep sequencing and data analysis

Rice genomic DNA was extracted from protoplast samples 48 h after transfection and used as template. In the first-round PCR amplification, site-specific primers were used to amplify the target regions using TransStart FastPfu Fly DNA Polymerase (TransGen

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

Biotech, Beijing, China) (Table S4). In the second-round, both forward and reverse barcodes were added to the ends of the first-round PCR amplification products for library construction (Table S4). Equal numbers of PCR amplification products were pooled and sequenced commercially using the NovaSeq platform (Novagene, Beijing, China). After deep sequencing, the substitutions and indels within sgRNA target sites in the sequenced reads were examined and counted. Analyses of base-editing processivity were performed as previously described (Komor et al., 2016).

Agrobacterium-mediated transformation of rice

The plant transformation binary vectors, pH-RK-BE3, pH-YEE-BE3, pH-eA3A-BE3, pH-A3A-BE3, pH-nCas9-PBE, pH-A3Bctd-BE3, pH-A3Bctd-T214V-BE3, pH-A3Bctd-RM-BE3, pH-A3Bctd-VHM-BE3 and pH-A3Bctd-KKR-BE3 were transformed into *Agrobacterium tumefaciens* train AGL1 by electroporation. Then they were transformed into callus cells of Zhonghua11 via *Agrobacterium*-mediated transformation as reported (Zhang et al., 2016). Hygromycin (50 μg/ml) was added to the medium to select transgenic rice plants.

De novo assembly of the Zhonghua 11 genome

An individual rice plant (Zhonghua 11) was used to isolate high-quality genomic DNA. A PacBio Sequel library was constructed and generated 28.6 Gb long reads, while a PCR-free Illumina pair-end library yielded 26.4 Gb NGS short reads. The assembly was first generated by Canu (v1.8) (Koren et al., 2017) using the PacBio long reads, and then improved by two rounds of PacBio long read correction and three rounds of NGS short read correction using Pilon (version 1.23) (Walker et al., 2014). The assembly was further anchored into chromosome-level pseudo-molecules based on the synteny of the protein-coding genes of the rice reference genome (*Oryza sativa* ssp. japonica cv. Nipponbare, https://rapdb.dna.affrc.go.jp/) using ALLMAPS (v1.0) (Tang et al., 2015). After removing redundant contigs using PurgeHaplotigs (Roach et al., 2018), the final assembly consisted of 72 scaffolds, including 12 pseudo-chromosomes and 60 unplaced scaffolds. The final Zhonghua 11 assembly has the same percentage of complete genes as that of the reference rice genome (98.3% versus 98.3%) by BUSCO analysis (BUSCO.v3) (Waterhouse et al., 2018). The proportion of properly-mapped NGS reads for the majority of both wild and callus-regenerated rice plants was greater than 98%. This indicates that this study generated a high-quality Zhonghua 11 genome suitable for calling of SNVs and small indels.

Whole-genome sequencing and data analysis

Raw NGS reads were processed using BIGpre (v2.0.2) (Zhang et al., 2011) and the qualified reads for each rice plant were first mapped to the corresponding vector sequence to confirm that the expected vector had been inserted into the genome. After this, the qualified NGS reads for each rice plant were mapped to the Zhonghua 11 assembly generated in this study with BWA-MEM (v0.7.17) (Li and Durbin, 2010). The alignment for each plant was processed by marking duplicated reads using Picard tools (v1.119) (https://github.com/broadinstitute/picard). The reads in insertion/deletion regions were realigned using RealignerTargetCreator and IndelRealigner modules in the Genome Analysis Toolkit (GATK v3.5) (McKenna et al., 2010). Finally, the refined alignments for each plant were used for variant calling by three softwares: GATK (v3.5) (McKenna et al., 2010), LoFreq (v2.1.2) (Wilm et al., 2012) and Strelka2 (v2.9.10) (Kim et al., 2018). After removing the background variations between wild-type plants and the reference assembly, common SNVs identified by the above three softwares (GATK/LoFreq/Strelka2) were considered as the high-confidence SNVs, while common indels identified by two softwares (GATK/Strelka2) were considered as the high-confidence indels.

Analysis of BE3 and BE3 variants induced genome-wide C·G-to-T·A SNVs in plants

Using the genomic locations of all C·G-to-T·A SNVs detected in Figure 2 and Figure 6, the flanking sequences (5 bp on either side) were extracted from the Zhonghua 11 genome (mentioned in De Novo Assembly of the Zhonghua 11 Genome). These flanking sequences were aligned, fixing the mutant Cs in each case at position 0, and the resulting alignments were used to produce sequence logos on the online website http://weblogo.threeplusone.com/ (Crooks et al., 2004).

hA3Bctd structural analysis

The structural representation in Figure 4A was generated by PyMOL (https://pymol.org/2/). The structure information for hA3Bctd was from PDB accession number (PDB: 2NBQ) (Siriwardena et al., 2015). The side chains of the mutated residues are labeled with the single-letter amino-acid abbreviation.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed with R package 3.4.3 (http://www.Rproject.org/). All experiments for deep sequencing were performed with 3-6 biological replicates. For WGS experiments, *P value*s were calculated using the Mann-Whitney test, and $p < 0.05$ was considered as being statistically significant. For the correlation analysis of the nSaCas9-mediated orthogonal R-loop assay, correlation coefficients (R^2) and *P value*s were generated by an R program, and $p < 0.05$ was considered statistically significant.

Please cite this article in press as: Jin et al., Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity, Molecular Cell (2020), https://doi.org/10.1016/j.molcel.2020.07.005

DATA AND CODE AVAILABILITY

All data supporting the findings of this study are available in the article and its supplementary figures and tables or are available from the corresponding author on request. All the sequencing data were deposited in NCBI BioProject under accession code PRJNA602608. For sequence data, rice LOC_Os identifiers (http://rice.plantbiology.msu.edu/) are: LOC_Os01 g55540 (*OsAAT*-T1,*OsAAT*-SaT1), LOC_Os02 g30630 (*OsALS*-T1), LOC_Os03 g05730 (*OsCDC48*-T1, *OsCDC48*-T2, *OsCDC48*-SaT1), LOC_Os09 g26999 (*OsDEP1*-T1, *OsDEP1*-SaT2), LOC_Os10 g40600 (*OsNRT1.1B*-T1, *OsNRT1.1B*-SaT1), LOC_Os02 g11010 (*OsEV*-T1), LOC_Os11 g32080 (*OsACTG*-T1), LOC_Os03 g08570 (*OsPDS*-T1, *OsPDS*-Cpf1T1), LOC_Os06 g04280.1 (*OsEPSPS*-Cpf1T1), LOC_Os12 g36030 (*OsLDMAR*-T1), LOC_Os08 g39890 (*OsSPL14*-SaT1), LOC_Os09 g35980 (*OsTAC1*-SaT1), LOC_Os10 g27380.1 (*OsDL*-SaT1).

Molecular Cell, Volume *79*

Supplemental Information

Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity

Shuai Jin, Hongyuan Fei, Zixu Zhu, Yingfeng Luo, Jinxing Liu, Shenghan Gao, Feng Zhang, Yu-Hang Chen, Yanpeng Wang, and Caixia Gao

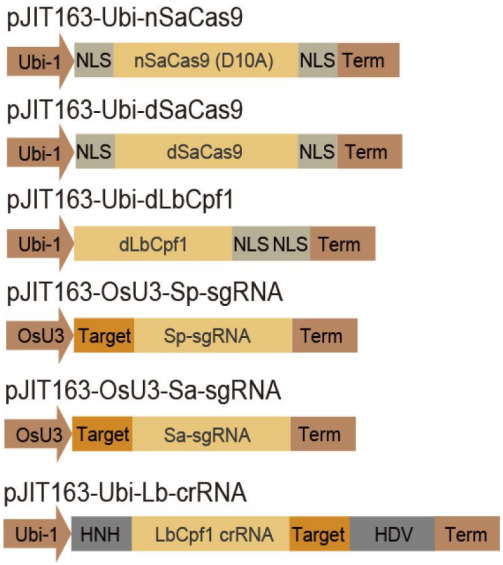


Figure S1. Schematic representation of the various CRISPR systems used in optimizing the orthogonal R-loop assay. Related to Figure 1.

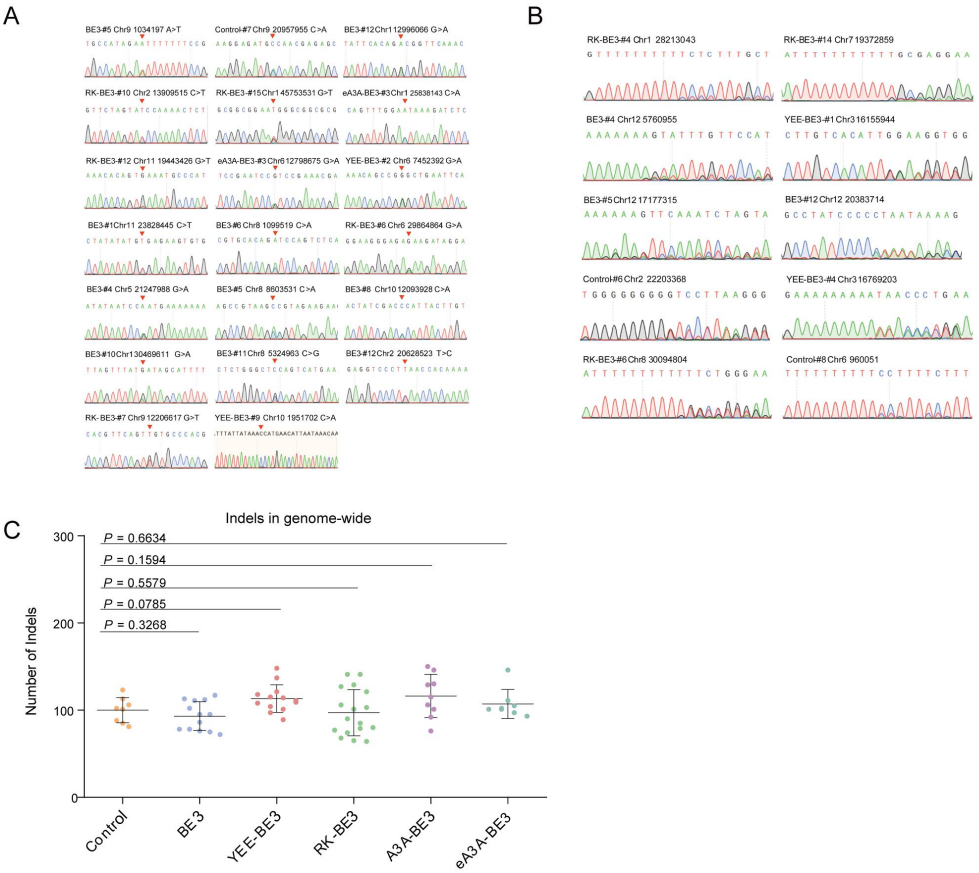


Figure S2. Validating the nSaCas9-mediated orthogonal R-loop assay in plants by whole-genome sequencing, related to Figure 2.

(A) Sanger sequencing chromatograms of representative SNVs detected by WGS. SNVs from the indicated samples were PCR amplified and verified by Sanger sequencing. Red arrows represent mutated nucleotides. Primers are listed in Table S3.

(B) Sanger sequencing chromatograms of representative indels detected by WGS. Indels from the indicated samples were PCR amplified and verified by Sanger sequencing. Primers are listed in Table S3.

(C) Number of genome-wide indels identified in five base editors. Number of indels identified in control, BE3, YEE-BE3, RK-BE3, A3A-BE3 and eA3A-BE3 plants by WGS. Average numbers of indels in control, BE3, YEE-BE3, RK-BE3, A3A-BE3 and eA3A-BE3 plants were 100, 93, 97, 113, 116 and 107. Statistical significance of differences was calculated using the two-sided Mann–Whitney *U*-test.

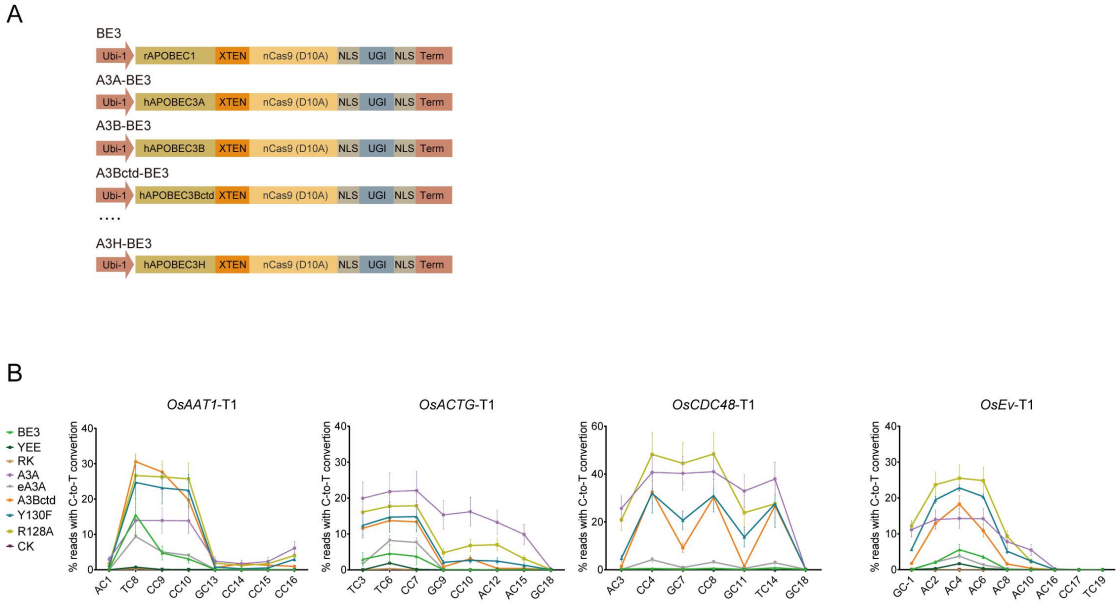


Figure S3. Screening human APOBEC3-based base editors for the rationally-designed platform, related to Figure 3.

(A) Schematic representation of BE3 and hAPOBEC3-BE3s.

(B) Frequencies of single C-to-T conversions by A3Bctd-BE3 and seven reported base editors at four targeted sites. Single C-to-T editing efficiencies of BE3, YEE-BE3, RK-BE3, A3A-BE3, eA3A-BE3, A3Bctd-BE3, A3A-Y130F-BE3, A3A-R128A-BE3 at four targets in rice protoplasts (*OsAAT1*-T1, *OsACTG*-T1, *OsCDC48*-T1 and *OsEv*-T1, counting the PAM as positions 21-23). Frequencies (mean ± s.e.m.) were calculated using the data in Figure 3C, n=3-6 independent biological replicates.

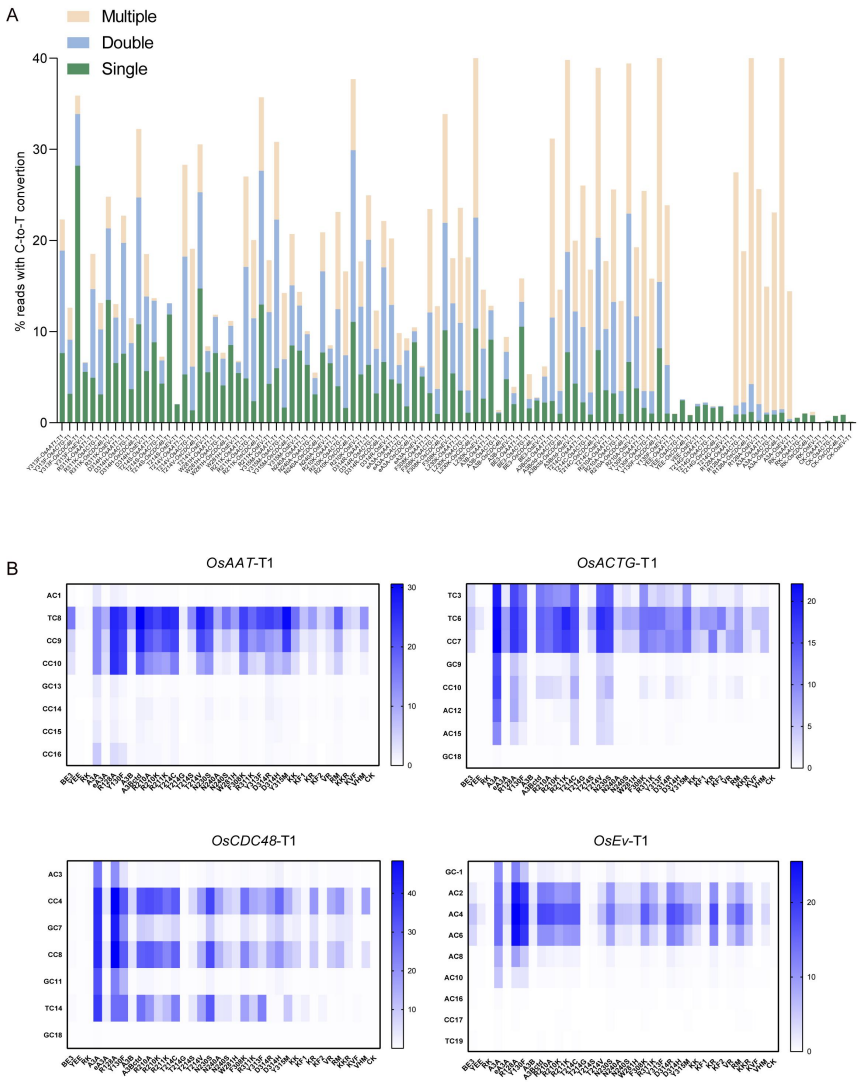


Figure S4. Evaluating the frequencies of on-target and sgRNA-independent off-target edits induced by the rationally-designed A3Bctd-BE3 variants, related to Figure 4.

(A) Mutation types analyses of editing products on different target sites by the A3Bctd-BE3 variants. Types of edits created by the A3Bctd-BE3 variants bearing single amino acid substitutions at four on-target sites (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1 and *OsEv*-T1). Single C-to-T changes are shown in green, double in blue and multiple in orange. The average efficiencies were calculated by the mean of 3-6 biologically independent experiments.

(B) **Single C-to-T conversions by A3Bctd-BE3 variants at four targeted sites.** Heat maps showing single C-to-T editing efficiencies of cytosine base editors used in this study at four endogenous rice gene target sites in rice protoplasts (*OsAAT1*-T1,

OsACTG-T1, *OsCDC48*-T1 and *OsEv*-T1, counting the PAM as positions 21-23, window from -1 to 20).

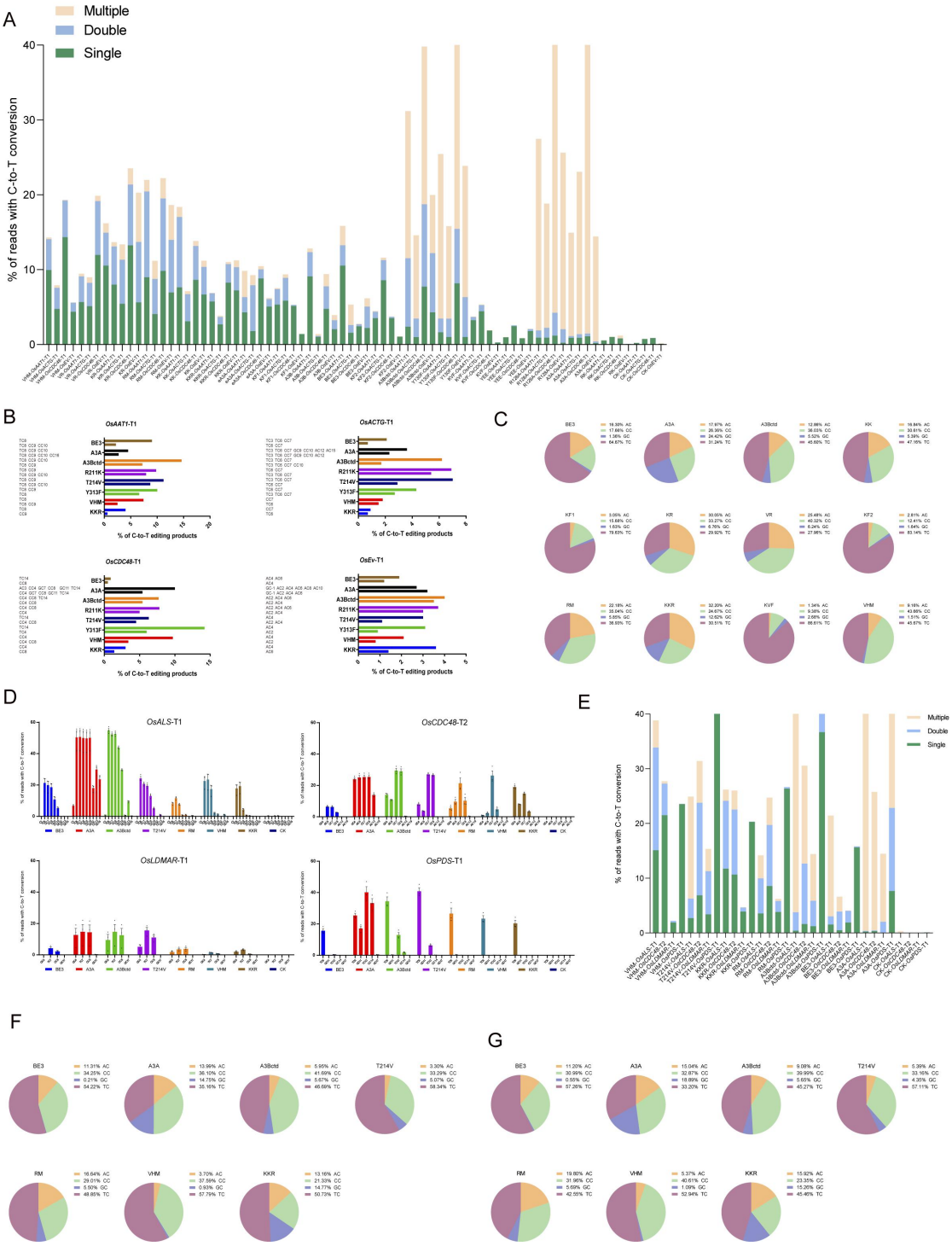


Figure S5. Some of the newly engineered base editors exhibit high specificity and increased precision, related to Figure 5.

(A) Mutation types analyses of editing products on different target sites induced by A3Bctd-BE3 variants. Types of edits created by A3Bctd-BE3 variants bearing double or triple mutations across the four on-target sites (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1 and *OsEv*-T1). Single C-to-T changes are shown in green, double in

blue and multiple in orange. The average efficiencies were calculated by the mean of 3-6 biologically independent experiments.

(B) Analysis of C-to-T editing products induced by BE3, A3A-BE3, A3Bctd-BE3, R211K-BE3, Y313F-BE3, VHM and KKR at four target sites. The frequencies of the editing products in the amplicon deep sequencing reads at four rice genes targeted by the various BE3 editors were further analyzed and the two most frequent types of editing product are shown. Edited Cs are shown on the left and represent the mean of 3-6 biologically independent experiments.

(C) Analysis of editing sequence preferences induced by various BE3 base editors at four target sites. The frequencies of the editing products in the amplicon deep sequencing reads across four on-target sites (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1 and *OsEv*-T1) targeted by nine combinatory A3Bctd-BE3 variants, BE3, A3A-BE3 and A3Bctd-BE3 were further analyzed and the frequencies of C-to-T conversion of AC, CC, GC, and TC motifs versus the frequencies of total C-to-T conversion are shown in orange, green, blue and magenta, respectively. The proportion of edited Cs were calculated by the mean of 3-6 biologically independent experiments.

(D) Frequencies (percentages) of single C-to-T conversions induced by various CBEs at four new on-target sites. C-to-T conversion induced by four key A3Bctd-BE3 variants (A3Bctd-VHM-BE3, A3Bctd-T214V-BE3, A3Bctd-KKR-BE3 and A3Bctd-RM-BE3), BE3, A3A-BE3 and A3Bctd-BE3 across four new on-target sites (*OsALS*-T1, *OsCDC48*-T2, *OsLDMAR*-T1 and *OsPDS*-T1) in rice protoplasts, counting the PAM as positions 21-23, window from 1 to 20. The average efficiencies were calculated by the mean ± s.e.m of 3 biologically independent experiments.

(E) Mutation types analyses of editing products on four new on-target sites induced by selected A3Bctd-BE3 variants. Types of edits induced by four A3Bctd-BE3 variants (A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3, and A3Bctd-KKR-BE3) and BE3, A3A-BE3 in four new on-target sites (*OsALS*-T1, *OsCDC48*-T2, *OsLDMAR*-T1 and *OsPDS*-T1). Single C-to-T changes are shown in green, double in blue and multiple in orange. The average efficiencies were calculated by the mean of 3 biologically independent experiments.

(F-G) Analysis of editing sequence preferences induced by various selected A3Bctd-BE3 variants. The frequencies of the editing products in the amplicon deep sequencing reads across four on-target sites (*OsALS*-T1, *OsCDC48*-T2, *OsLDMAR*-T1 and *OsPDS*-T1) (F) and eight on-target sites (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1,*OsEv*-T1, *OsALS*-T1, *OsCDC48*-T2, *OsLDMAR*-T1 and *OsPDS*-T1) (G) targeted by four key A3Bctd-BE3 variants (A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3), BE3, A3A-BE3 and A3Bctd-BE3 were further analyzed and the frequencies of C-to-T conversion of AC, CC, GC, and TC motif versus the frequencies of total C-to-T conversion are shown in orange, green, blue and magenta, respectively. The proportion of edited Cs were calculated by the mean of 3-6 biologically independent experiments.

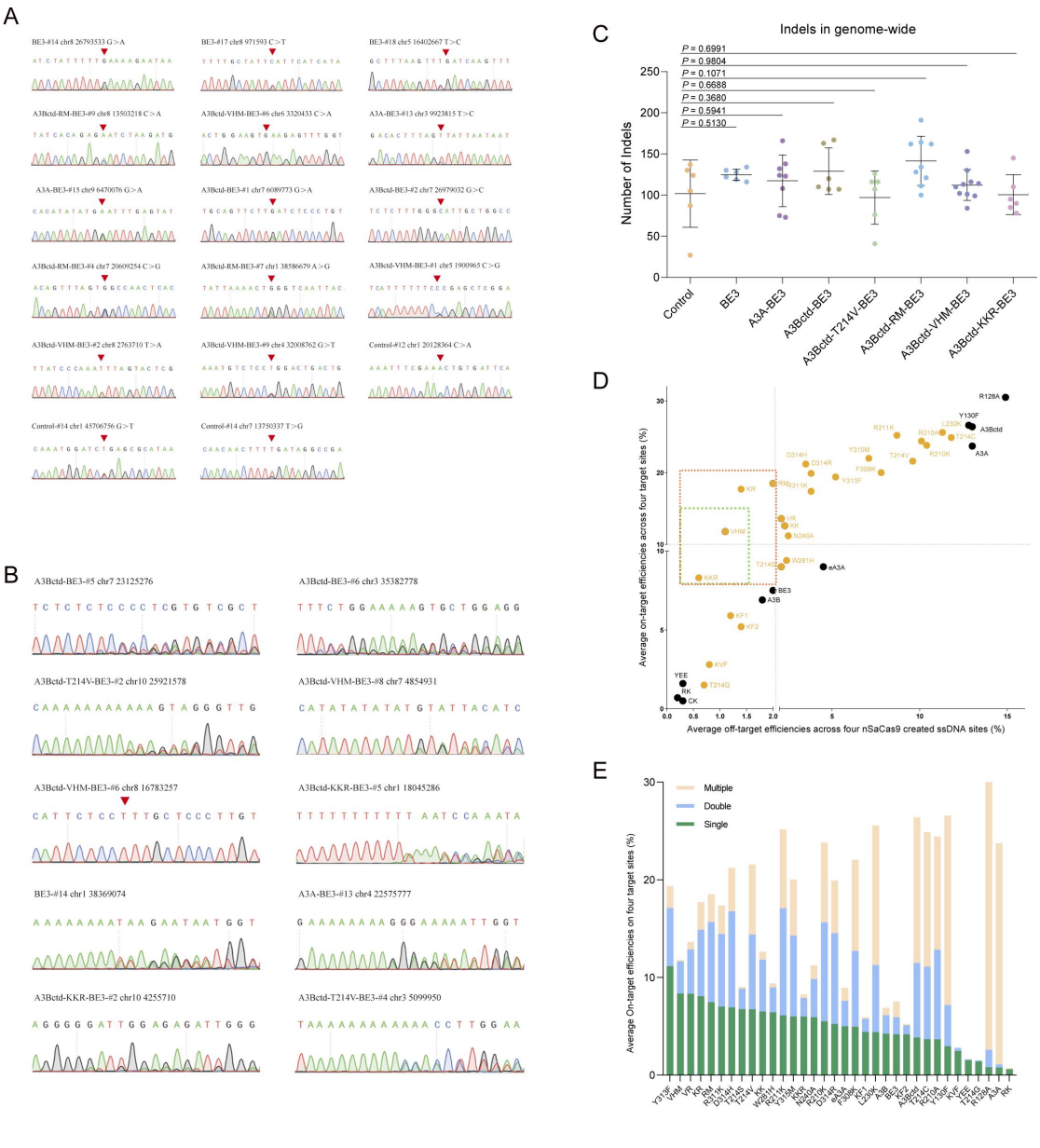


Figure S6. Further evaluation of the DNA off-target activities of the selected A3Bctd-based new base editors, and validation of the nSaCas9-mediated orthogonal R-loop assay in plants by whole-genome sequencing, related to Figure 6.

(A) Sanger sequencing chromatograms of representative SNVs detected by WGS. SNVs from the indicated samples were PCR amplified and verified by Sanger sequencing. Red arrows represent mutated nucleotides. Primers are listed in Table S3.

(B) Sanger sequencing chromatograms of representative indels detected by WGS. Indels from the indicated samples were PCR amplified and verified by Sanger sequencing. Primers are listed in Table S3.

(C) Number of genome-wide indels identified in seven base editors. Number of indels identified in control, BE3, A3A-BE3, A3Bctd-BE3, A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 plants by WGS. Average numbers of indels in control, BE3, A3A-BE3, A3Bctd-BE3, A3Bctd-T214V-BE3, A3Bctd-RM-BE3, A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 plants were 102,125, 117, 129, 97, 141, 112 and 101, respectively. Statistical significance of differences was calculated using the two-sided Mann–Whitney *U*-test.

(D) Evaluating the average on-target edits verses sgRNA-independent off-target edits induced by the rationally-designed A3Bctd-BE3 variants. Average on-target C-to-T conversion frequencies across four on-target sites(*OsAAT1*-T1, *OsACTG*-T1, *OsCDC48*-T1 and *OsEv*-T1) versus average sgRNA-independent off-target C-to-T conversion frequencies across four ssDNA regions (*OsCDC48*-SaT1, *OsDEP1*-SaT2, *OsNRT1.1B*-SaT1 and *OsDEP1*-SaT1) for 34 base editors used in this study .Base editors created in this study are shown in orange dots and reported base editors are shown in black dots. The orange box shows A3Bctd-BE3 variants possessed increased or comparable on-target activities and decreased or comparable off-target activities than those of BE3. The green box showed A3Bctd-BE3 variants with increased on-target activities and further lower off-target activities than BE3.

(E) Mutation types analyses of editing products induced by A3Bctd-BE3 variants. Average efficiency of reads with single, double or multiple C-to-T conversion induced by 34 CBEs used in this study across four on-target sites (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1 and *OsEv*-T1). Single C-to-T changes are shown in green, double in blue and multiple in orange. The average frequencies are mean values of four sites, calculated by the mean of 3-6 biologically independent experiments.

(F) Protospacer, PAM and flanking sequences of all on- and off-target sites. Sequence information of eight on-target (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1,*OsEv*-T1, *OsALS*-T1, *OsCDC48*-T2, *OsLDMAR*-T1 and *OsPDS*-T1) and eight off-target sites (*OsAAT*-SaT1, *OsCDC48*-SaT1, *OsDEP1*-SaT1,*OsDEP1*-SaT1, *OsDL*-SaT1, *OsNRT1.1B*-SaT1, *OsSPL14*-SaT1 and *OsTAC1*-SaT1) used to test all CBEs in this study. Protospacer (counting base distant from PAM as position 1) are shown in yellow and all C bases within protospacer are shown in blue, PAM sequences in red, flanking sequences (10 bp upstream of protospacer position 1 and 10 bp downstream of PAM sequence) in gray.

Supplementary Table 1. List of on-target and off-target loci used in this study with the corresponding primers used for cloning sgRNA. Related to Figures 1-6.

sgRNA	Target sequence	Oligo-F	Oligo-R
OsAAT1-T1	CAAGGATCCCAGCCCCGTGA**AGG**	GGCGCAAGGATCCCAGCCCCGTGA	AAACTCACGGGGCTGGGATCCTTG
OsACTG-T1	ATCATCCGCCACGACGGCGG**CGG**	GGCGATCATCCGCCACGACGGCGG	AAACCCGCCGTCGTGGCGGATGAT
OsALS-T1	CAGGTCCCCCGCCGCATGAT**CGG**	GGCGCAGGTCCCCCGCCGCATGAT	AAACATCATGCGGCGGGGGACCTG
OsCDC48-T1	GACCAGCCAGCGTCTGGCGC**CGG**	GGCGGACCAGCCAGCGTCTGGCGC	AAACGCGCCAGACGCTGGCTGGTC
OsCDC48-T2	TAGCACCCATGACAATGACA**TGG**	GGCGTAGCACCCATGACAATGACA	AAACTGTCATTGTCATGGGTGCTA
OsEV-T1	ACACACACACTAGTACCTCT**GGG**	GGCGACACACACACTAGTACCTCT	AAACAGAGGTACTAGTGTGTGTGT
OsLDMAR-T1	GTGCATCCTGGGAAAGCTTG**TGG**	GGCGGTGCATCCTGGGAAAGCTTG	AAACCAAGCTTTCCCAGGATGCAC
OsPDS-T1	GTTGGTCTTTGCTCCTGCAG**AGG**	GGCGTTGGTCTTTGCTCCTGCAG	AAACCTGCAGGAGCAAAGACCAA
OsAAT1-SaT1	AGGTGACGGTCGCGTACAAC**AAGGAT**	GGCGAGGTGACGGTCGCGTACAAC	AAACGTTGTACGCGACCGTCACCT
OsCDC48-SaT1	CTCGTTCCCATGTCATTGTC**ATGGGT**	GGCGCTCGTTCCCATGTCATTGTC	AAACGACAATGACATGGGAACGAG
OsDEP1-SaT1	TGGTCACTCAGCCTGCAGTA**CTGAAT**	GGCGTGGTCACTCAGCCTGCAGTA	AAACTACTGCAGGCTGAGTGACCA
OsDEP1-SaT2	GTCGTGCCCTGAATGTTCCT**GTGGGT**	GGCGGTCGTGCCCTGAATGTTCCT	AAACAGGAACATTCAGGGCACGAC
OsDL-SaT1	ACTGACCTGAAACGGGCCCA**AGGGGT**	GGCGACTGACCTGAAACGGGCCCA	AAACTGGGCCCGTTTCAGGTCAGT
OsNRT1.1B-SaT1	GATCATCGACAGGTCGGCGG**CGGAGT**	GGCGGATCATCGACAGGTCGGCGG	AAACCCGCCGACCTGTCGATGATC

OsSPL14-SaT1	TGGTCACTCTAGTGCAGTGG**CGGGAT**	GGCGTGGTCACTCTAGTGCAGTGG	AAACCCACTGCACTAGAGTGACCA
OsTAC1-SaT1	AAATCCCCGCAAAAGGTGAAA**GGGAAT**	GGCGAAATCCCGCAAAAGGTGAAA	AAACTTTCACCTTTTGCGGGATTT
OsEPSPS-Cpf1T1	**TTTA**TGAAAATATGTATGGAATTCATG	AGATTGAAAATATGTATGGAATTCATG	GGCCCATGAATTCCATACATATTTTCA
OsPDS-Cfp1T1	**TTTG**GAGTGAAATCTCTTGTCTTAAGG	AGATGAGTGAAATCTCTTGTCTTAAGG	GGCCCCTTAAGACAAGAGATTTCACTC

The C bases in the deamination window are highlighted in red. The PAM motif in each target sequence is showed in bold. Oligo-F and Oligo-R were used to clone sgRNA.

Supplementary Table 5. Ratio of on-target to off-target editing of all CBEs in this paper. Related to Figures 4-5.

CBEs	On-target efficiencies(%)	Off-target efficiencies(%)	On/Off ratio
BE3	7.5	2	3.75
YEE	1.6	0.3	5.33
RK	0.7	0.2	3.50
A3A	23.7	13	1.82
eA3A	9	4.5	2.00
R128A	30.5	14.9	2.05
Y130F	26.6	12.8	2.08
A3B	6.9	1.8	3.83
A3Bctd	26.4	13	2.03
R210A	24.4	10.1	2.42
R210K	23.8	10.4	2.29
R211K	25.2	8.7	2.90
T214V	21.6	9.6	2.25
T214S	9	2.1	4.29

T214C	24.9	11.8	2.11
L230K	25.6	11.3	2.27
N240A	11.2	2.5	4.48
W281H	9.4	2.4	3.92
F308K	22	7.1	3.10
R311K	17.4	3.8	4.58
Y313F	19.4	5.2	3.73
D314R	19.9	3.8	5.24
D314H	21.2	3.5	6.06
Y315M	20	7.8	2.56
T214G	1.5	0.7	2.14
KVF	2.8	0.8	3.50
KF2	5.2	1.4	3.71
VHM	11.8	1.1	10.73
RM	18.5	2	9.25
KK	12.6	2.3	5.48

KR	17.7	1.4	12.64
KKR	8.3	0.6	13.83
KF1	5.9	1.2	4.92
VR	13.6	2.1	6.48
CK	0.5	0.3	1.67

Table shows the ratio of on-target to off-target editing Frequencies of on-target C-to-T conversion at four rice genes loci (*OsAAT*-T1, *OsACTG*-T1, *OsCDC48*-T1, and *OsEv*-T1) and sgRNA-independent off-target C-to-T edits at four ssDNA regions loci (*OsCDC48*-SaT1, *OsDEP1*-SaT2, *OsNRT1.1B*-SaT1 and *OsDEP1*-SaT1) created by nSaCas9 in rice protoplasts were measured by deep sequencing.