BREAKTHROUGH TECHNOLOGIES, TOOLS, AND RESOURCES

Short title: Off-target analyses of a cytosine base editor

Corresponding author:

Yiping Qi,

Department of Plant Science and Landscape Architecture, University of Maryland, College Park, Maryland 20742, USA.

Email: yiping@umd.edu

Genome- and transcriptome-wide off-target analyses of an improved cytosine base editor

Linnell Bentley Randall[1, ǂ], Simon Sretenovic[2, ǂ], Yuechao Wu[3,4 ǂ], Desuo Yin[2], Tao Zhang[3,4*], Joyce Van Eck[1,5*], Yiping Qi[2,6*]

[1]The Boyce Thompson Institute, Ithaca, New York 14853, USA; [2]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, Maryland 20742, USA; [3]Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding/Key Laboratory of Plant Functional Genomics of the Ministry of Education, College of Agriculture, Yangzhou University, Yangzhou 225009, China; [4]Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou 225009, China; [5]Plant Breeding and Genetics Section, Cornell University, Ithaca, NY 14853; [6]Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, Maryland 20850.

ǂ These authors contributed equally to this work.

One-sentence Summary:

Whole-genome sequencing and transcriptome sequencing reveal high editing specificity of an improved cytosine base editor in tomato.

Author contributions:

YQ and JVE designed the experiments. SS and DY generated all base editing vectors. LBR conducted tomato protoplast and stable transformation. LBR and SS analyzed PCR amplicon NGS data. LBR analyzed mutagenesis of stable tomato lines. LBR prepared tomato samples for WGS and RNA-seq. YW and TZ analyzed the WGS and RNA-seq

data. LBR, YW, and SS made the figures. YQ, JVE, and TZ wrote the manuscript. All authors participated in discussion and revision of the manuscript.

Funding information:

This work was supported by the U.S. Department of Agriculture Biotechnology Risk Assessment Grant Program (award no. 2018-33522-28789) to YQ and JVE. It was also partly supported by the National Science Foundation Plant Genome Research Program (award no. IOS-1758745 and IOS-2029889) to YQ. SS is a Foundation for Food and Agriculture Research Fellow.

Competing interests

The authors declare that they have no competing financial interests.

Authors' emails:

Linnell Bentley Randall: linnell.randall@gmail.com

Simon Sretenovic: simonsre@umd.edu

Yuechao Wu: wuyuechao@zhangtaolab.org

Desuo Yin: yindesuo@163.com

Tao Zhang, Email: zhangtao@yzu.edu.cn

Joyce Van Eck, Email: jv27@cornell.edu

Yiping Qi, Email: Yiping@umd.edu

***Senior authors:** zhangtao@yzu.edu.cn; jv27@cornell.edu; Yiping@umd.edu

Abstract

Cytosine base editors (CBEs) are promising tools for precise genome editing in plants. It is important to investigate potential off-target effects of an efficient CBE at the genome and transcriptome levels in a major crop. Based on comparison of five cytidine deaminases and two different promoters for expressing sgRNAs, we tested a highly efficient A3A/Y130F-BE3 system for efficient C-to-T base editing in tomato (*Solanum lycopersicum*). We then conducted whole-genome sequencing (WGS) of four base-edited tomato plants, three GFP-expressing control plants, and two wild-type (WT) plants. The sequencing depths ranged from 25X to 49X with read mapping rates above 97%. No sgRNA-dependent off-target mutations were detected. Our data show an average of ~1000 single nucleotide variations (SNVs) and ~100 insertions and deletions (indels) per GFP control plant. Base-edited plants had on average elevated levels of SNVs (~1250) and indels (~300) per plant. On average, about 200 more C-to-T (G-to-A) mutations were found in a base-edited plant than a GFP control plant, suggesting some level of sgRNA-independent off-target effects, though the difference is not statistically significant. We also conducted RNA sequencing (RNA-seq) of the same four base-edited plants and three GFP control plants. An average of ~200 RNA SNVs was discovered per plant for either base-edited or GFP control plants. Furthermore, no specific enrichment of C-to-U mutations can be found in the base-edited plants. Hence, we cannot find any evidence for bona fide off-target mutations by A3A/Y130F-BE3 at the transcriptome level.

INTRODUCTION

Base editors such as cytosine base editors (CBEs) and adenine base editors (ABEs) are precise genome editing tools with wide applications in genetics, medicine and agriculture (Zhang et al., 2019; Anzalone et al., 2020; Gurel et al., 2020). The most widely used C-to-T base editing platform is BE3, which consists of a Cas9(D10A) nickase, a cytidine deaminase like rAPOBEC1, and a uracil glycosylase inhibitor (UGI) (Komor et al., 2016). The earlier CBEs utilized cytidine deaminases like rAPOBEC1 (Komor et al., 2016) and *Petromyzon marinus* cytidine deaminase 1(PmCDA1) (Nishida et al., 2016). In recent years, additional cytidine deaminases have been demonstrated in human cells, such as APOBEC3A (Gehrke et al., 2018; Wang et al., 2018) and some CDA1-like deaminases (Cheng et al., 2019). Multiple cytidine deaminases have been used for base editing in plants such as rAPOBEC1 (Li et al., 2017; Lu and Zhu, 2017; Zong et al., 2017), hAID (Ren et al., 2018; Kuang et al., 2020; Sretenovic et al., 2021), PmCDA1 (Shimatani et al., 2017; Tang et al., 2019; Zhong et al., 2019; Sretenovic et al., 2021), and A3A (Zong et al., 2018). Base editing efficiency, editing purity and activity windows are all key parameters of consideration when developing and implementing base editing systems (Molla and Yang, 2019; Anzalone et al., 2020; Gurel et al., 2020).

As with other genome editing tools, specificity of CBEs has been a focus of intensive investigation. It was reported that rAPOBEC1-based BE3 generated genome-wide off-target effects that were independent of single guide RNAs (sgRNAs) in mice (*Mus musculus*) (Zuo et al., 2019) and rice (*Oryza sativa*) (Jin et al., 2019). To reduce off-target C-to-T conversions by CBEs, engineered forms of cytidine deaminases were developed, such as eA3A in human cells (Gehrke et al., 2018), and A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3 in rice (Jin et al., 2020). While rAPOBEC1-BE3 was reported to generate off-target C-to-U mutations at the transcriptome level in humans (Grunewald et al., 2019; Zhou et al., 2019), two BE3 variants, BE3-R33A and BE3-R33A/K34A, showed substantially reduced RNA editing when highly expressed in HEK293T cells (Grunewald et al., 2019). It was also reported that CBEs based on PmCDA1, hAID and eA3A induced less off-target mutations at the transcriptome level when compared to those based on rAPOBEC1 and A3A in human cells (Grunewald et al., 2019). Furthermore, another CBE variant, YE1-BE3-FNLS, displayed high on-target activity with reduced off-target effects

at both DNA and RNA levels in mammalian cells (Zuo et al., 2020). Despite the progress in mammalian systems, the potential transcriptome off-target effects remain unclear for any established CBE systems in plants.

While it is important to further develop highly efficient CBEs for precise base editing in plants, it is also critical to assess the potential off-target effects for such CBEs at the genome and transcriptome levels. Tomato (*Solanum lycopersicum*) is a dicot model plant and crop, very suitable for developing and demonstrating genome editing technologies. For example, Brooks et al. reported CRISPR-Cas9-mediated genome editing in tomato in 2014 (Brooks et al., 2014) and this editing system has been applied to improve agronomic traits in tomato (Soyk et al., 2017). CBEs have also been demonstrated in tomato, notably with the PmCDA1 deaminase (Shimatani et al., 2017; Veillet et al., 2019). Here, we describe our efforts in confirming an improved CBE system, A3A/Y130F-BE3, for high-efficiency C-to-T base editing in tomato. We further assessed this promising CBE for editing specificity at both the genome and transcriptome levels by using whole-genome sequencing (WGS) and RNA sequencing (RNA-seq).

RESULTS

Comparison of different CBE systems in tomato protoplasts

To seek a robust CBE system in tomato, we compared five different BE3 systems (**Supplemental Figure. S1**): rAPOBEC1-BE3, hAID-BE3, PmCDA1-BE3, A3A-BE3, and an improved variant of A3A-BE3, A3A/Y130F-BE3, which allowed high-efficiency base editing in human cells (Wang et al., 2018). All these BE3 systems were expressed under the 2X35S promoter. We also compared two small RNA promoters, AtU6 (U6) and AtU3 (U3), for sgRNA expression. The T-DNA expression vectors were constructed according to our modular assembly system (Lowder et al., 2015). The AtU6 and AtU3 promoters were previously demonstrated in Arabidopsis (*Arabidopsis thaliana*) (Lowder et al., 2015) and carrot (*Daucus carota* subsp. *sativus*) (Klimek-Chodacka et al., 2018). Comparing both promoters in tomato would likely help us identify the most efficient one for base editing. Since the preferred start-nucleotide in transcription by AtU6 and AtU3 is G and A, respectively, when necessary, an additional G or A was added to the 5’ end of the

protospacer to meet these requirements (**Supplemental Table S1**). These five BE3 editors and two sgRNA expression promoters were compared for editing two independent target sites in tomato *AGO7* (*SolyA7*). This gene is involved in small RNA biogenesis (Husbands et al., 2009), whose knockout results in a needle-like or wiry leaf phenotype (Yifhar et al., 2012). C-to-T base editing efficiency was assessed in tomato protoplasts at 22ºC and quantified by next-generation sequencing (NGS) of PCR amplicons. The assay revealed different C-to-T conversion rates in tomato protoplasts when the AtU3 promoter was used for sgRNA expression (**Figure 1A and Figure 1B**). By contrast, all CBE systems showed very low editing activity when the AtU6 promoter was used (**Figure 1A and Figure 1B**), suggesting that the AtU3 promoter renders higher sgRNA expression in tomato. When the AtU3 promoter was used in combination with rAPOBEC1-BE3, hAID-BE3, and PmCDA1-BE3 there was poor C-to-T editing activity at both target sites, SolyA7-gRNA3 and SolyA7-gRNA4, while A3A-BE3 showed higher C-to-T editing rates than these three CBEs, consistent with the recent report of high editing activity for A3A-BE3 in rice, wheat (*Triticum aestivum*) and potato (*Solanum tuberosum*) (Zong et al., 2018). A3A/Y130F-BE3 showed the highest base editing efficiency among all CBEs tested, resulting in ~3% editing efficiency at both target sites (**Figure 1A and Figure 1B**). The slightly improved editing efficiency of A3A/Y130F-BE3 over A3A-BE3 is consistent with the previous report in human cells (Wang et al., 2018). Further analysis showed these five BE3 editors had different base editing windows at both target sites, with A3A-BE3 and A3A/Y130F-BE3 showing larger windows than those of other tested BE3 editors (**Figure 1C and Figure 1D**).

Given that Cas9 activity is temperature-sensitive in plants (LeBlanc et al., 2018), we tested hAID-BE3, PmCDA1-BE3 and A3A/Y130F-BE3 at three different temperatures: 22ºC, 28ºC and 32ºC. While there was no significant difference in base editing efficiency for all three CBEs at these temperatures, incubation of the transformed protoplasts at 32ºC seemed to result in more robust editing (**Figure 1E and Figure 1F**). This result is consistent with Cas9's temperature sensitivity (LeBlanc et al., 2018), but may also imply possible temperature sensitivity of cytidine deaminases.

High frequency base editing in stable lines by A3A/Y130F-BE3

To assess whether A3A/Y130F-BE3 can readily generate C-to-T base-edited plants, we conducted stable transformation of tomato using *Agrobacterium tumefaciens*-mediated T-DNA delivery. We included hAID-BE3 and PmCDA1-BE3 as controls since both systems have been widely used in plants (Shimatani et al., 2017; Ren et al., 2018; Tang et al., 2019; Zhong et al., 2019; Kuang et al., 2020; Sretenovic et al., 2021) and showed higher editing efficiency than rAPOBEC1-BE3 in rice (Tang et al., 2019). hAID-BE3 and PmCDA1-BE3 generated 20% and 25% C-to-T editing efficiency at the SolyA7-gRNA3 site (**Table 1**), respectively. At the SolyA7-gRNA4 site, neither of the two CBEs generated pure C-to-T base editing among the T0 lines examined (**Table 1**). By contrast, A3A/Y130F-BE3 resulted in 68.8% and 72.7% base editing efficiency at the SolyA7-gRNA3 and SolyA7-gRNA4 sites, respectively. At the SolyA7-gRNA3 site, 43.8% were pure C-to-T editing events and 25% included C-to-G or C-to-A editing. At the SolyA7-gRNA4 site, all base-edited T0 lines generated by A3A/Y130F-BE3 contained C-to-T mutations (**Table 1**). Our data suggest that A3A/Y130F-BE3 is a far more efficient CBE than hAID-BE3 and PmCDA1-BE3 in transgenic tomato plants, consistent with the data in tomato protoplasts (**Figure 1**). However, we also observed high frequencies of insertions and deletions (indels) in T0 lines expressing A3A/Y130F-BE3, 38% at the SolyA7-gRNA3 site and 27.3% at the SolyA7-gRNA4 site (**Table 1**). This is probably due to high frequency induction of DNA double strand breaks caused by high cytidine deaminase activity of A3A/Y130F.

The base editing windows identified in stable tomato lines were consistent with those identified in protoplasts. At the SolyA7-gRNA3 site, hAID-BE3 edited C4 to C8 in the target (**Figure 2A**). By contrast, PmCDA1-BE3's editing window shifted toward the 5' end of the protospacer (C1 to C7) (**Figure 2B**), while A3A/Y130F-BE3's editing window shifted toward the middle of the protospacer (C6 to C8) (**Figure 2C**). While hAID-BE3 and PmCDA1-BE3 only produced heterozygous base editing, A3A/Y130F-BE3 produced homozygous C-to-T editing at both target sites in multiple lines (**Figure 2C**). The high editing efficiency of A3A/Y130F-BE3 was very obvious as it generated simultaneous biallelic base conversions at multiple editable Cs at the target sites across several independent transgenic lines (**Figure 2C**). Biallelic base editing by A3A/Y130F-BE3 could generate loss-of-function of the *SolyA7* target gene. That was indeed the case at the

SolyA7-gRNA3 site as some lines containing biallelic editing showed wiry leaves (e.g., A3A/Y130F-BE3-gRNA3 #44 and #49) (**Figure 2D**). The data suggest that a P542F missense mutation caused by C-to-T base editing in A3A/Y130F-BE3-gRNA3 #44 and #49 lines is sufficient to largely if not completely abolish the AGO7 protein function in tomato (Yifhar et al., 2012; Brooks et al., 2014). Since the mutation is recessive, heterozygous lines (e.g., A3A/Y130F-BE3-gRNA3 #5B and #31) showed a wild-type phenotype as was observed for the GFP control plant (**Figure 2D**). Together, the results suggest A3A/Y130F-BE3 is a highly potent CBE that can reliably generate biallelic editing in the T0 generation of tomato.

To evaluate germline transmission of the base edits, we followed three T0 lines into the T1 generation and genotyped these lines by Sanger sequencing (**Supplemental Table S2**). Among 12 T1 lines from A3A/Y130F-BE3-gRNA3 #5B that were genotyped, six lines showed a heterozygous C7A,C8T,C13G genotype that resembles the T0 parental line and one line had a homozygous C7A,C8T,C13G genotype. In two heterozygous and one homozygous base-edited lines, the CBE transgene appeared to be segregated away. All the genotyped T1 lines from A3A/Y130F-Be3-gRNA4 #3 and #5B contained C6T,C8T homozygous base edits, the same genotype as the parental lines. Similarly, multiple base-edited T1 lines were transgene-free. Hence, in all three cases, the observed base edits in T0 lines were faithfully transmitted to the T1 generation and transgene-free base-edited T1 lines were readily recovered.

Confirmation of on-target base editing by WGS

Having identified A3A/Y130F-BE3 as an efficient CBE system in tomato, we investigated the potential genome-wide off-target effects of this promising base editor. We conducted WGS on four edited T0 lines (two independent lines each for the two target sites: SolyA7-gRNA3 and SolyA7-gRNA4), three GFP control T0 lines for tissue culture induced somaclonal variation, and two WT tomato plants. For all nine samples, the genome sequencing coverage ranged from 25x to 49x and sequencing reads were mapped to the genome at 97.48% or higher (**Supplemental Table S3**). To analyze the data, we adopted a similar analysis pipeline to what we previously used for assessing genome-wide off-target effects by Cas9 and Cas12a in rice (Tang et al., 2019). This rigorous pipeline used

three independent calling programs to identify single nucleotide variations (SNVs) and insertions and deletions (indels) (**Figure 3A**). Based on WGS, we re-confirmed the base editing events at the two target sites in the four selected lines, which were previously identified by Sanger sequencing (**Figure 3B and Figure 3C**), suggesting that our WGS pipeline and data analysis is highly reliable.

Investigation of sgRNA-dependent off-target effects of A3A/Y130F-BE3 by WGS

We next sought to investigate off-target effects of A3A/Y130F-BE3. By allowing five nucleotide mismatches to the protospacers, Cas-OFFinder (Bae et al., 2014) predicted 155 and 72 putative off-target sites for the SolyA7-gRNA3 and SolyA7-gRNA4 targets, respectively. However, WGS analysis did not reveal any mutations at these sites (**Figure 4A**). To further examine sgRNA-dependent off-target effects, we searched for shared mutations (SNVs or indels) between the two edited plants for each sgRNA. Only one shared mutation site was identified in the A3A/Y130F-BE3-gRNA4 samples (**Figure 4B**). However, this target site showed very low sequence similarity to the target protospacer sequence, and a C-to-A mutation was found at the first nucleotide of the protospacer at this putative off-target site (**Figure 4C**). Neither the base editing type nor the editing position matched the predicted base editing profile of A3A/Y130F-BE3. Hence, we were unable to identify any sgRNA-dependent off-target mutations in the genome of base-edited plants. This result is not surprising considering we designed two very specific sgRNAs for editing the *SolyA7* gene. The Cas-OFFinder predicted that off-target sites contained at least four nucleotide mismatches for SolyA7-gRNA3 and at least three for SolyA7-gRNA4 (**Figure 4A**). Based on our previous WGS study in rice, Cas9 generally could not induce off-target mutations at putative off-target sites with more than two nucleotide mismatches to the protospacers (Tang et al., 2019).

Investigation of sgRNA-independent off-target effects of A3A/Y130F-BE3 by WGS

To identify sgRNA-independent and deaminase-dependent off-target mutations, we first compared the total SNVs per plant for each sample group. About 200 SNVs were identified in each WT plant (**Figure 5A**), indicating a spontaneous SNV mutation rate. On average, ~1000 SNVs were found in each GFP control plant, which defines a level of

somaclonal variation for SNVs attributed to the tissue culture and Agrobacterium-mediated transformation process in our experimental conditions (**Figure 5A**). For the base-edited plants, an average of ~1250 SNVs were found for each line (**Figure 5A**). However, the numbers of SNVs between base-edited plants and GFP control plants are not statistically significant (P=0.193) (**Figure 5A**). For all sample types, more SNVs were found in the transposable elements (TEs) and repeats than in exons, introns or intergenic regions (**Figure 5B**). These SNVs were evenly distributed across the 12 tomato chromosomes (**Supplemental Figure. S2**). A further breakdown of the SNVs showed that A3A/Y130F-BE3-gRNA3 lines or A3A/Y130F-BE3-gRNA4 lines had around 600 C:G>T:A mutations, approximately 200 more than the GFP plants, although the difference is not statistically significant (**Figure 5C**). Nevertheless, the majority of additional mutations found in base-edited lines appeared to be C:G>T:A mutations, which are signatures of off-target mutations of a CBE system. Further analysis showed the fraction of C:G>T:A changes were slightly elevated in the A3A/Y130F-BE3-gRNA3 and gRNA4 lines when compared to the controls (e.g. GFP or WT) (**Supplemental Figure. S3**). Therefore, we concluded that the tomato tissue culture process generated ~1000 SNVs per regenerated plant as a result of tissue culture and Agrobacterium-mediated transformation, and A3A/Y130F-BE3 appeared to generate more C:G>TA mutations (~200) in tomato (**Figure 5C**), providing some evidence of genome-wide off-target effects.

Indels are common byproducts of CBEs which are likely due to base-excision repair-generated DNA double-strand breaks (Komor et al., 2016; Nishida et al., 2016). We also compared the number of indels among different sample groups. The WT plants each had ~50 indels, indicating a spontaneous indel mutation rate (**Figure 5D**). The GFP plants had an average of ~120 indels per plant, while the base-edited plants had ~300 indels per plant, albeit with large variations among the four samples, and consequently the numbers of indels between the GFP control plants and base-edited plants are not statistically significant (P=0.2217) (**Figure 5D**). The identified indels are evenly distributed across the 12 tomato chromosomes (**Supplemental Figure. S2B**). Interestingly, the indels were enriched in intergenic and intronic regions for the GFP and base-edited plants (**Figure 5E**). Compared to spontaneous mutations found in the WT plants, all seven Agrobacterium-transformed and tissue cultured plants (GFP and base-edited) had more

deletions than insertions (~80% vs ~20%) (**Figure 5F**). Furthermore, the vast majority of these indels were 1-bp deletions and insertions (**Figure 5G and Figure 5H**). Such 1-bp indels are not known as signature mutations for off-target effects of any CBEs. Since the GFP and base-edited samples showed very similar indel profiles, we concluded that very few indels, if any, resulted from the off-target effects of A3A/Y130F-BE3.

Transcriptome-wide off-target analysis of A3A/Y130F-BE3

Several CBEs were reported to induce off-target mutations at the transcriptome level in human cells (Grunewald et al., 2019; Zhou et al., 2019). To determine the potential off-target effects of A3A/Y130F-BE3 at the transcriptome level in tomato, we conducted RNA sequencing (RNA-seq) on the same four base-edited T0 lines and three GFP controls lines that were used for WGS. The total reads for each sample ranged from 59.06M to 113.58M (**Supplemental Table S4**). We established a robust pipeline for analyzing the RNA-seq data (**Figure 6A**). For all seven samples, most of the SNVs found at the transcriptome level were derived from genomic SNVs, with a Pearson correlation coefficient ≥0.50 among these two groups of SNVs in each plant (**Figure 6B**). The RNA-specific SNVs constituted 10% to 14.9% SNVs detected in the GFP samples, and of 11.4% to 15.5% SNVs detected in the base-edited samples (**Figure 6C**). There were ~200 RNA SNVs per sample for either GFP plants or base-edited plants (**Figure 6D**). Among these SNVs, C>U changes represented ~30% for both GFP and base-edited samples (**Figure 6E**). Considering unidirectional gene transcription, we further compared the GFP controls and base-edited lines for all 12 possible nucleotide changes, and again no difference was found among these samples (**Supplemental Figure. S4**). In addition, no specific motif around mutated cytosines (Cs) was enriched to show any preferred activity by A3A/Y130F-BE3, indicting no evidence of RNA editing (**Figure 6F**). Together, our analyses suggested that A3A/Y130F-BE3 did not elicit any detectable off-target C-to-U mutations at the transcriptome level.

DISCUSSION

The widely used rAPOBEC1-BE3 has been applied in numerous plant species such as rice (Li et al., 2017; Lu and Zhu, 2017), Arabidopsis (Chen et al., 2017; Xue et al., 2018;

Li et al., 2019), maize (*Zea mays*) (Lu and Zhu, 2017), wheat (Lu and Zhu, 2017; Zhang et al., 2019), and cotton (*Gossypium hirsutum*) (Qin et al., 2020). In recent years, there have been reports describing other BE3s based on different cytidine deaminases such as hAID (Ren et al., 2018; Tang et al., 2019; Sretenovic et al., 2021), PmCDA1 (Shimatani et al., 2017; Tang et al., 2019; Sretenovic et al., 2021) and A3A (Zong et al., 2018) with higher editing efficiency in plants. By comparing different BE3 systems in tomato, we found that A3A-BE3 outperformed rAPOBEC1-BE3, hAID-BE3 and PmCDA1-BE3 with much higher base editing efficiency (**Figure 1**). We want to note that the absolute values for base editing efficiency in tomato protoplasts were not high, which can be partly explained by the fact that these protoplast cells were largely non-dividing which subsequently limits base editing efficiency. This potential problem could be alleviated with the tissue culture system where there are actively dividing cells. We indeed observed very high base editing efficiency (~70%) with A3A/Y130F-BE3 in tomato stable lines at two independent target sites (**Table 1**). Our results are consistent with recent reports on high base editing activity of A3A-BE3 in rice, wheat and potato (Zong et al., 2018), oilseed rape (*Brassica napus*) (Cheng et al., 2021), as well as in human cells (Gehrke et al., 2018; Wang et al., 2018; Grunewald et al., 2019; Zhou et al., 2019; Doman et al., 2020; Tan et al., 2020). Previously, A3A/Y130F-BE3 was reported to have higher editing efficiency in human cells than A3A-BE3 (Wang et al., 2018). Similarly, we found that A3A-Y130F further promoted base editing efficiency in tomato. This mutation was previously implied to impact the deaminase-single strand DNA (ssDNA) interaction, especially at the target cytosine residues (Shi et al., 2017). In a recent parallel study, we also found that A3A/Y130F-BE3 resulted in highly efficient C-to-T conversions in a populus hybrid (*Populus tremula* × *P. alba* hybrid clone INRA 717-1B4) (Li et al., 2021). In light of these reports, A3A/Y130F-BE3 probably is the best CBE, or at least represents one of the best CBEs demonstrated so far, for precise C-to-T base editing in plants. However, we also observed high frequencies of indel byproducts induced by A3A/Y130F in tomato (**Table 1**). To reduce indel byproducts, it would most likely be helpful to have an additional copy of UGI by adoption of the BE4 configuration (Komor et al., 2017). Since A3A/W98Y/W104A rendered high editing efficiency when coupled with dCas12a in mammalian cells (Wang et al., 2020), it would be interesting to test whether this dual

mutation variant of A3A could further enhance C-to-T base editing in tomato and other plants. Our study also showed that the AtU3 promoter is superior to the AtU6 promoter for expression of the sgRNAs in tomato, which is consistent with our recent observation in poplar (Li et al., 2021). While both promoters were included in our CRISPR toolbox (Lowder et al., 2015), we would recommend the use of AtU3 for dicot applications.

WGS is a reliable method to assess off-target effects of genome editing reagents. Previously, we applied WGS for genome-wide analysis of off-target effects of Cas9 and Cas12a in rice (Tang et al., 2018). This study, along with reports for different plant species (Feng et al., 2014; Zhang et al., 2014; Nekrasov et al., 2017; Lee et al., 2019), showed that CRISPR-Cas systems were very specific genome editing tools as long as the guide RNAs were well-designed. However, in C-to-T base editing applications, sgRNA-independent off-target mutations could occur due to non-specific DNA binding and editing by the cytidine deaminase. This is more likely to happen if the cytidine deaminase has a high tendency for promiscuous DNA binding and high enzyme activity. In our study, we indeed found some evidence, albeit lack of statistical significance, of such genome-wide C-to-T off-target mutations caused by A3A/Y130F-BE3, averaging ~200 off-target mutations in each edited tomato plant. However, we also found that tissue culture-induced somaclonal variation amounted to ~1000 mutations per plant, which substantially dwarfed the number of off-target mutations caused by A3A/Y130F-BE3. Furthermore, only a fraction of the off-target mutations was located in exons, suggesting that the potential physiological and phenotypic off-target effects of this highly active A3A/Y130F-BE3 are marginal.

Based on our analysis, a low level of genome-wide off-target mutations by A3A/Y130F-BE3 is noticeable, despite lack of statistical significance. On the other hand, tissue culture and Agrobacterium-mediated transformation resulted in large and variable numbers of somaclonal variations in each plant, as we previously observed in rice (Tang et al., 2018). Although, it would still be desirable to further minimize the potential off-target effects of CBEs such as A3A/Y130F-BE3, and we reason that at least three different approaches can be used in plants. First, the extent of off-targeting may be controlled by the CBE expression level. To investigate this, we analyzed the RNA-seq data for the transcript levels of the CBE among four base-edited T0 lines (**Supplemental Figure. S5**).

Pearson correlation analysis showed no positive correlation between the CBE expression and the numbers of DNA SNVs, DNA Indels or RNA SNVs. To our surprise, the two A3A/Y130F-BE3-gRNA3 T0 lines (#31 and #5B) had very low levels of CBE expression. This result suggests that the base edits observed in these lines might be due to transient expression of the T-DNA during the early stage of Agrobacterium-mediated transformation. Genome editing from transient gene expression was previously reported in tobacco (*Nicotiana tabacum*) when a similar Agrobacterium-mediated transformation protocol was used (Chen et al., 2018). With this complication, we could not provide any direct evidence in this study for corelating high levels of CBE expression to high levels of off targeting. However, we cannot rule out this possibility. Second, the exposure time of the genome to the CBE reagents can be minimized by using ribonucleoprotein (RNP) delivery for transient expression, as shown in mammalian cells (Rees et al., 2017). Therefore, it is possible that RNP delivery of the CBE system would result in reduced off-targeting in plants. Hence, it will be interesting to compare RNP delivery and DNA-based delivery for base editing in plants. Finally, A3A/Y130F-BE3 or other highly efficient CBEs may be further engineered to mitigate off-target effects. Although, a big challenge is how to improve editing specificity without compromising on-target editing activity. For instance, an engineered A3A (eA3A) with reduced off-target effects (Gehrke et al., 2018) showed low on-target editing efficiency in human cells (Doman et al., 2020), rice (Jin et al., 2020), and yeast (*Saccharomyces cerevisiae*) (Tan et al., 2020). Recently, rationally designed A3B CBEs (A3Bctd-VHM-BE3 and A3Bctd-KKR-BE3) showed minimal genome-wide off-target editing in rice (Jin et al., 2020). These variants did not show higher on-target editing efficiency than A3A-BE3 (Jin et al., 2020) and hence are likely to be less efficient than A3A/Y130F-BE3. Nevertheless, it is promising to explore additional cytidine deaminases and variants for improvement of on-target base editing concomitant with negligible genome-wide off-target DNA editing in plants.

Recently, transcriptome-wide off-target mutations were reported for rAPOBEC1-BE3 in human cells (Grunewald et al., 2019; Zhou et al., 2019). To mitigate RNA off-targeting, one group used protein engineering to develop BE3-R33A and BE3-R33A/K34A variants that had similar on-target DNA editing activity to the rAPOBEC1-BE3 and substantially reduced RNA editing activity (Grunewald et al., 2019). Another group

took a similar approach and engineered multiple BE3 variants including A3A/Y130F-BE3 with undetectable RNA editing activity (Zhou et al., 2019). In our study, we did not find any off-target effects at the transcriptome level for A3A/Y130F-BE3 in tomato. Further supporting this work and earlier findings (Zhou et al., 2019), recent studies showed that A3A-BE3 resulted in off-target mutations in the transcriptome (Grunewald et al., 2019) and introduction of the Y130F mutation in A3A-BE3 completely abolished RNA editing activity in human cells (Zuo et al., 2020). However, A3A/Y130F-BE3 still possessed high DNA off-target effects which is correlated with its high DNA on-target editing activity (Wang et al., 2018; Zuo et al., 2020). Nevertheless, absence of transcriptome-wide off-target effects for A3A/Y130F-BE3 is very important as it can help avoid CBE self-editing at the RNA level and hence ensure high on-target editing efficiency and specificity in the genome (Grunewald et al., 2019).

Conclusions

We have demonstrated a highly efficient A3A/Y130F-BE3 system for C-to-T base editing in tomato. We provided some evidence based on WGS and RNA-seq of genome-wide, but not transcriptome-wide, off-target effects of this promising base editing system. Compared to the level of tissue culture-induced somaclonal variation, the putative genome-wide off-target mutations by A3A/Y130F-BE3 were neither substantial nor statistically significant. Hence, A3A/Y130F-BE3 represents a highly efficient and precise C-to-T base editor that can be utilized for many plant genome editing applications.

MATERIALS AND METHODS

Vector construction

Gateway compatible attL1-attR5 entry clones pYPQ265 (Addgene #164712) and pYPQ266 (Addgene #164713) were prepared using homologous recombination of two PCR amplicons in *Escherichia coli* DH5α of a maize codon optimized (z) Cas9(D10A) nickase fragment from pYPQ166-D10A vector with primers zCas9-F1-Rec and zCas9-R1-Rec in the case of pYPQ265 and with primers zCas9-F2-Rec and zCas9-R2-Rec in

the case of pYPQ266. The PCR amplified backbone from pYPQ255 (Addgene #124310) with primers UGI-F1-Rec and APB-R1-Rec was used to prepare pYPQ265 and the PCR amplified backbone from pYPQ256 (Addgene #124312) with primers CDA-F1-Rec and attI1-R1-Rec was used to prepare pYPQ266. pYPQ265C (Addgene #164715) was prepared by cloning the PCR amplified gBlock™ (IDT) hAID-XTEN with primers hAID-XTEN-F and hAID-XTEN-R after restriction digestion with BsrGI and NcoI into the BsrGI and NcoI digested pYPQ265. pYPQ265E1 (Addgene #164718) and pYPQ265E2 (Addgene #164719) were prepared by cloning the gBlocks™ A3A in the case of pYPQ265E1, A3A-Y130F in the case of pYPQ265E2 after restriction digestion with BsrGI and BsaI into the BsrGI and BsaI digested pYPQ265 backbone. All Gateway compatible attL1-attR5 entry clones were confirmed by Sanger sequencing with primers listed in **Supplemental Table S1**. All enzymes used for preparing Gateway compatible attL1-attR5 entry clones were purchased from NEB.

T-DNA vectors (**Supplemental Table S5**) for cytosine base editing were constructed using Golden Gate and Gateway LR assembly reactions based on the protocols described previously (Lowder et al., 2015). Briefly, forward and reverse primers (**Supplemental Table S1**) were phosphorylated with T4 polynucleotide kinase (NEB, catalog #M0201*), annealed, and ligated with T4 DNA ligase (NEB, catalog #M0202*) into BsmBI (ThermoFisher, catalog #ER045*) restriction digested pYPQ141A (Addgene #69290, with the AtU6 promoter) or pYPQ141B (Addgene #69291, with the AtU3 promoter) sgRNA entry clones in one-step Golden Gate reactions. Individual three-way Gateway LR reactions were conducted using attL5-attL2 sgRNA entry clone, attL1-attR5 base editor entry clone, and attR1-attR2 destination vector pCGS710 containing the 2x35S promoter for base editor expression (**Supplemental Figure. S6**). Both sgRNA and base editor entry clone recombination regions were confirmed by Sanger sequencing. Final T-DNA vectors were confirmed by restriction digestion with EcoRV-HF (NEB, catalog # R3195*).

Tomato protoplast isolation and transformation

Protoplasts were harvested from cotyledons of 6-9-day-old *in vitro*-grown seedlings of the M82-indeterminate tomato containing a functional SELF-PRUNING gene (M82 SP+). M82 SP+ seedlings were grown *in vitro* as previously described (Van Eck et al., 2019). Excised cotyledons were floated in enzyme solution (400 mM mannitol, 10 mM $CaCl_2$, 20 mM KCL, 10 mM MES, 0.3% (w/v) Cellulase Onozuka R-10 (Yakult Pharmaceutical), 0.15% (w/v) Macerozyme R-10 (Yakult Pharmaceutical), pH 5.7) for 16-20 hrs at 22°C in the dark with gentle agitation on a rotating shaker. The protoplast suspension was filtered through a 100 µm cell strainer and centrifuged for 10 min at 200 x g. The resulting protoplast pellet was resuspended in 0.55 M sucrose (pH 5.7), and slowly overlayed with W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES, pH 5.7) without mixing. After centrifugation for 30 min at 200 x g, protoplasts were extracted from the sucrose/W5 interface using a glass Pasteur pipette, washed with fresh W5 solution, and counted by hemocytometer. The final protoplast pellet was resuspended in MMG (500 mM mannitol, 15 mM $MgCl_2$, 4 mM MES, pH 5.7) to a density of 1 x 10^6 protoplasts/ml.

Protoplasts were transformed according to the method previously described (Zhang et al., 2013). Briefly, 200 µl of MMG protoplast suspension (2 x 10^5 protoplasts) were mixed with 10 µg of purified plasmid DNA (in 40 µl water), followed by 240 µl freshly prepared PEG solution (40% (w/v) PEG-3350, 200 mM mannitol, 100 mM $CaCl_2$), gently mixed, and incubated in the dark for 20 min. Then 800 µl W5 solution was added and protoplasts were collected by centrifugation for 5 min at 200 x g, washed with W5, centrifuged for 5 min at 200 x g, and resuspended in 2 ml W5 solution. The entire 2 ml protoplast preparation was transferred to a 6-well plate, and incubated in the dark for two days at the designated temperature (22°C, 28°C, or 32°C). Transformation efficiency for each experiment was estimated by counting the number of GFP-positive protoplasts from a sample transformed with JL33 (a binary vector containing the *neomycin phosphotransferase II* (*nptII)* selectable marker and *GFP* fluorescent reporter genes) (Floss et al., 2013) in at least three fields of view. Transformed protoplasts were collected by centrifugation at 10,000 x g for 10 min, and pellets were resuspended in 20 µl Phire Dilution Buffer (Thermo Fisher) and stored at -20°C.

Mutation analysis of transformed protoplasts

Target regions were PCR-amplified from protoplasts with barcoded primers (**Supplemental Table S1**) using Phire Plant Direct PCR Kit (Thermo Fisher) per the manufacturer's instructions. Amplicons were confirmed by gel electrophoresis, purified with QIAQuick PCR Purification Kit (QIAGEN), quantified by Nanodrop (Thermo Fisher), and combined in equal ratios into pools of 3-5 for deep sequencing. Amplicon-EZ sequencing was performed by Genewiz. Mutation analysis was performed on FASTQ sequence files using BE-Analyzer software (Hwang et al., 2018). Individual amplicon sequences, with 6-nt barcodes per end, were entered as Target Sequences, with a Base Editing Window from 1-20. Default parameters of R=10 and n=1 were used. C-to-T and indel frequencies for each construct are reported as the average of at least three independent biological replicates.

Tomato stable transformation

Agrobacterium tumefaciens-mediated transformation of cotyledons from the M82 SP+ was performed with the Agrobacterium AGL1 strain as previously described (Van Eck et al., 2019). All cultures were grown at 28°C.

Mutation analysis of stably transformed lines

DNA was extracted from leaf tissue of well-rooted T0 plants as previously described (Stewart and Via, 1993). Verification of the stable lines was confirmed by PCR amplification for the presence of Cas9 using GoTaq Green Mastermix (Promega) following the manufacturer's instructions. *SolyA7* target regions were PCR amplified using Q5 High-Fidelity DNA Polymerase (NEB) per the manufacturer's instructions (**Supplemental Table S1**). Amplicons were verified by gel electrophoresis, purified with QIAQuick PCR Purification Kit (QIAGEN), quantified by Nanodrop (Thermo Fisher), and Sanger sequenced at Genewiz or the Cornell Institute of Biotechnology. Sanger sequencing chromatograms were aligned to the tomato reference sequence (Solyc01g010970) using Geneious Prime software, and mutations were identified by visual inspection.

Whole genome sequencing

DNA was extracted from leaf tissue of well-rooted, *in vitro*-grown T0 plants and 3-week-old M82 SP+ wild-type seedlings in soil as previously described (Stewart and Via, 1993). Genomic DNA was quantified using the Qubit 2.0 Fluorometer (Life Technologies). DNA integrity was checked with ~1% (w/v) agarose gel with 50-100 ng sample loaded in each well. Samples were then chosen for library preparation based on the QC results. NEBNext® Ultra™ II DNA Library Prep Kit for Illumina, clustering, and sequencing reagents were used according to manufacturer recommendations (NEB). Briefly, the genomic DNA was fragmented by acoustic shearing with a Covaris S220 instrument. Fragmented DNA was purified, and end repaired. Adapters were ligated after adenylation of the 3' ends, followed by enrichment by limited cycle PCR. DNA libraries were validated using a DNA 1000 Chip on the Agilent 2100 Bioanalyzer (Agilent Technologies), and were quantified using Qubit 2.0 Fluorometer. The DNA libraries were also quantified by qPCR (Applied Biosystems). Illumina sequencing was performed by Genewiz on a HiSeq 4000 platform.

RNA sequencing

Leaf tissue (approximately 30 mg) was excised from each well-rooted, *in vitro*-grown T0 plant for RNA extraction and sequencing. RNA was extracted with the RNeasy Plant Mini kit per the manufacturer's instructions (QIAGEN). Extracted RNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies) and RNA integrity was checked using Agilent TapeStation 4200 (Agilent Technologies). RNA sequencing libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina following the manufacturer's instructions (NEB). Briefly, mRNAs were enriched with Oligo(dT) beads. Enriched mRNAs were fragmented for 15 min at 94°C. First and second strand cDNAs were subsequently synthesized. cDNA fragments were end-repaired, adenylated at 3'ends, and universal adapters were ligated to the cDNA fragments followed by index addition and library enrichment by limited-cycle PCR. The sequencing libraries were validated on the Agilent TapeStation (Agilent Technologies), and quantified using Qubit 2.0 Fluorometer (Invitrogen) as well as by quantitative PCR (KAPA Biosystems). RNA extraction, library preparation, and sequencing were done by Genewiz.

Whole genome sequencing data analysis

WGS analysis was done by following our previous method (Tang et al., 2018) with only minor modifications. Briefly, the adapters were trimmed by applying SKEWER (v. 0.2.2) (Jiang et al., 2014). All cleaned reads were mapped to tomato reference sequence M82 SP+ (ftp://ftp.solgenomics.net/genomes/tomato100/March_02_2020_sv_landscape/) with BWA mem (v. 0.7.17) software (Li and Durbin, 2010). Picard and Samtools (v. 1.9) (Li et al., 2009) were used to filter multiple mapping reads. GATK (v. 3.8) (McKenna et al., 2010) was used to realign the reads near indels. Then, whole genome SNVs and indels were detected by applying LoFreq (v. 2.1.2) (Wilm et al., 2012), Mutect2 (Cibulskis et al., 2013), VarScan2 (v. 2.4.3) (Koboldt et al., 2012), and Strelka2 (v2.9.10) (Kim et al., 2018). Bedtools (v. 2.27.1) (Li, 2011) were used for overlapping of SNVs and indels. Potential off-target sites were predicted by applying Cas-OFFinder software (v. 2.4) (Bae et al., 2014) with up to 5-nt mismatches. Data processing and analysis were done using Python and R.

Whole transcriptome sequencing data analysis

Data preprocessing has been described in the WGS analysis section. All cleaned reads were mapped to tomato reference sequence M82 SP+ (ftp://ftp.solgenomics.net/genomes/tomato100/March_02_2020_sv_landscape/) with Hisat2 (v. 2.2.0) software(Kim et al., 2019). Picard tools were then applied to sort and mark duplicates of the mapped BAM files. The modified BAM files were subjected to spanned splice junctions, local realignment and variant calling with SplitNCigarReads, IndelRealigner, and HaplotypeCaller tools from GATK (v. 3.8) (McKenna et al., 2010), respectively. From all called RNA variants, downstream analyses focused solely on SNVs on canonical (Chr1–Chr12) chromosomes. To identify transcriptome-wide SNVs with high confidence, VariantFiltration tool was used to filter RNA SNVs. Sequence logos is made by WebLogo3 tool (http://weblogo.threeplusone.com/) (Crooks et al., 2004). Data processing and analysis were done using Python and R.

Statistical Analyses

Two-way ANOVA and Tukey's test were used for statistical analyses in this study.

Accession Numbers

Addgene numbers for five BE3 entry clones: rAPOBEC1-BE3 (164712), PmCDA1-BE3 (164713), hAID-BE3 (164715), A3A-BE3 (164718), and A3A/Y130F-BE3 (164719). The WGS raw data and RNA-seq data reported in this article have been deposited to the Sequence Read Archive in National Center for Biotechnology Information (NCBI) under the accession numbers PRJNA672142 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA672142) and PRJNA670713 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA670713), respectively.

Supplemental Data

Supplemental Figure S1. Five BE3 systems tested in tomato.

Supplemental Figure S2. The distributions of DNA SNVs and INDELs on tomato chromosomes.

Supplemental Figure S3. Nucleotide substitution types among different plants.

Supplemental Figure S4. SNVs in transcripts and motif analysis of altered cytosine (C).

Supplemental Figure S5. Pearson correlation between A3A/Y130F-zCas9D10A-UGI mRNA and the numbers of DNA or RNA level mutations in edited T0 plants.

Supplemental Figure S6. Map of the Gateway destination vector pCGS710.

Supplemental Table S1. Oligos and gBlocks™ used in this study.

Supplemental Table S2. Genotypes of T1 lines.

Supplemental Table S3. WGS coverage of each sample.

Supplemental Table S4. RNA-seq reads for each sample.

Supplemental Table S5. T-DNA vectors used in this study.

Acknowledgements

We thank Colby Starker and Daniel Voytas at University of Minnesota for providing the Gateway Destination vector pCGS710. We also thank Maria Harrison at Boyce Thompson Institute for providing the GFP-expressing JL33 plasmid.

Table 1. Base editing in stable tomato T0 lines

Base editor	Targeted site	Tested T0 lines	Base edited (C-to-T) T0 lines (number; ratio)	Base edited (C-to-R) T0 lines (number; ratio)	Indels in T0 lines (number; ratio)
hAID-BE3	SolyA7-gRNA3	10	2; 20%	0; 0%	1; 10%
hAID-BE3	SolyA7-gRNA4	15	0; 0%	0; 0%	0; 0%
PmCDA1-BE3	SolyA7-gRNA3	12	3; 25%	0; 0%	0; 0%
PmCDA1-BE3	SolyA7-gRNA4	14	0; 0%	1; 7.1%	0; 0%
A3A-Y130F-BE3	SolyA7-gRNA3	16	7; 43.8%	4; 25%	6; 38%
A3A-Y130F-BE3	SolyA7-gRNA4	11	8; 72.7%	0; 0%	3; 27.3%

Figure legends

Figure 1. Testing multiple CBEs in tomato protoplasts. (**A, B**) Assessment of five BE3 base editors in tomato protoplasts at two independent target sites. (**C, D**) C-to-T editing windows of different CBEs at the two target sites. (**E, F**) Assessment of three BE3 base editors in tomato protoplasts at three temperatures at two independent target sites. n=3-5 (except for the A3A/Y130F-AtU3-gRNA3 sample for which only two replicates were used). Error bars represent standard error of the mean (SEM). Significance was calculated using two-way ANOVA analysis of variance test, $^{**}p < 0.01$, $^{*}p < 0.05$; Letters denote statistical differences with Tukey's test ($p<0.05$).

Figure 2. Comparison of three CBEs in stable transgenic tomato lines. (**A**) Base edited T0 lines at the SolyA7-gRNA3 site by hAID-BE3. (**B**) Base edited T0 lines at the SolyA7-gRNA3 and SolyA7-gRNA4 sites by PmCDA-BE3. (**C**) Base edited T0 lines at the SolyA7-gRNA3 and SolyA7-gRNA4 sites by A3A/Y130F-BE3. Chromatograms of Sanger sequencing shown with base changes indicated by asterisks. (**D**) Phenotype of a GFP control plant, two T0 lines containing heterozygous base editing events, and two T0 lines containing homozygous base-edited alleles.

Figure 3. Whole genome sequencing of base-edited plants and control plants. (**A**) Workflow of whole-genome detection of SNV and indel mutations. SNV analysis involves using three computer programs: LoFreq, VarScan2, and MuTect2. Indel analysis also involves using three programs: VarScan2, MuTect2, and Strelka2. (**B**) Identification of targeted base editing by WGS in lines A3A/Y130F-BE3-gRNA3 #31 and A3A/Y130F-BE3-gRNA3 #5B. (**C**) Identification of targeted base editing by WGS in lines A3A/Y130F-BE3-gRNA4 #3 and A3A/Y130F-BE3-gRNA4 #5B.

Figure 4. Guide RNA-specific off-targeting effects of A3A/Y130F-BE3 in tomato. (**A**) Number of off-target sites identified in replicate edited plants vs the number of all potential off-target sites that are predicted by Cas-OFFinder with allowing up to 5-nt mismatch for both gRNA3 and gRNA4. (**B**) Identification of shared DNA SNVs and INDELs between

two replicated T_0 edited plants. (**C**) Top panel shows potential off-target sites identified in A3A/Y130F-BE3-gRNA4 samples based on shared mutations within two T_0 plants; bottom panel shows sequence analysis of the shared mutations in A3A/Y130F-BE3-gRNA4 samples.

Figure 5. Genome-wide off-target analysis of A3A/Y130F-BE3 in tomato. (**A**) Total number of DNA SNVs identified in the WT, GFP, and A3A/Y130F-BE3 plants. The average numbers of SNVs in WT, GFP, and A3A/Y130F-BE3 were 203.5, 940, and 1304, respectively. Each triangle represents the number of SNVs from an individual plant. Error bars represent SEM. (**B**) Annotation of genome-wide distribution of SNVs mutations found in WT, GFP, A3A/Y130F-BE3-gRNA3 and A3A/Y130F-BE3-gRNA4 plants. TE: transposable element. Error bars represent SEM. (**C**) Comparison of total C>T(G>A) SNVs in GFP, A3A/Y130F-BE3-gRNA3 and A3A/Y130F-BE3-gRNA4 plants. The average numbers of SNVs were 380.3, 597.5 and 573, respectively. Error bars represent SEM. (**D**) Total number of DNA indels identified in the WT, GFP, and A3A/Y130F-BE3 plants. Each triangle represents the number of indels from an individual plant. The average numbers of indels in WT, GFP, and A3A/Y130F-BE3 were 61.5, 116.3, and 314.75, respectively. Error bars represent SEM. (**E**) Annotation of genome-wide distribution of indels found in WT, GFP, A3A/Y130F-BE3-gRNA3 and A3A/Y130F-BE3-gRNA4 plants. Error bars represent SEM. (**F**) The fractions of indels in each plant. (**G**) Fractions of deletions of different sizes in each plant. (**H**) Fractions of insertions of different sizes in each plant. . *p* values were calculated by the two-sided unpaired t-test; * $p < 0.05$, ** $p < 0.01$.

Figure 6. Transcriptome-wide off-target analysis of A3A/Y130F-BE3 in tomato. (**A**) Workflow of detection of RNA-level SNV mutations. RNA SNVs are identified by HaplotypeCaller. (**B**) Scatter plot correlating RNA mutation rates of RNA SNVs as identified by HaplotypeCaller with DNA mutation rates as determined by WGS. The x-axis depicts fractions of RNA SNVs due to RNA-level mutations; The y-axis depicts fractions of RNA SNVs due to DNA-level mutations. Each dot represents an RNA SNV mutation. The Pearson's correlation was calculated between DNA mutation rates and RNA mutation rates. (**C**) Comparison of RNA-level SNVs in each plant. All identified RNA SNVs are

divided into DNA-level SNVs (>5%) and RNA-level SNVs (⩽5%) according to their DNA mutation rates. (**D**) Comparison of RNA-level SNV counts in GFP, A3A/Y130F-BE3-gRNA3 and A3A/Y130F-BE3-gRNA4 plants, respectively. *p* values were calculated by the two-sided unpaired t-test; $p < 0.05$ was considered significant. Error bars represent SEM. (**E**) The fraction of RNA-level SNV types in each plant. (**F**) Sequence logos derived from specific strand RNA level SNVs of each plant. Every T depicted should be considered a U in RNA. X-axis 0 indicated mutant cytosines position.

References

Anzalone AV, Koblan LW, Liu DR (2020) Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol **38:** 824-844

Bae S, Park J, Kim JS (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics **30:** 1473-1475

Brooks C, Nekrasov V, Lippman ZB, Van Eck J (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. Plant Physiol **166:** 1292-1297

Chen L, Li W, Katin-Grazzini L, Ding J, Gu X, Li Y, Gu T, Wang R, Lin X, Deng Z, McAvoy RJ, Gmitter FG, Jr., Deng Z, Zhao Y, Li Y (2018) A method for the production and expedient screening of CRISPR/Cas9-mediated non-transgenic mutant plants. Hortic Res **5:** 13

Chen Y, Wang Z, Ni H, Xu Y, Chen Q, Jiang L (2017) CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in Arabidopsis. Sci China Life Sci **60:** 520-523

Cheng H, Hao M, Ding B, Mei D, Wang W, Wang H, Zhou R, Liu J, Li C, Hu Q (2021) Base editing with high efficiency in allotetraploid oilseed rape by A3A-PBE system. Plant Biotechnol J **19:** 87-97

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

Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology **31:** 213-219

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Research **14:** 1188-1190

Doman JL, Raguram A, Newby GA, Liu DR (2020) Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. Nat Biotechnol **38:** 620-628

Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang DL, Wang Z, Zhang Z, Zheng R, Yang L, Zeng L, Liu X, Zhu JK (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proc Natl Acad Sci U S A **111:** 4632-4637

Floss DS, Levy JG, Levesque-Tremblay V, Pumplin N, Harrison MJ (2013) DELLA proteins regulate arbuscule formation in arbuscular mycorrhizal symbiosis. Proc Natl Acad Sci U S A **110:** E5025-5034

Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, Pinello L, Joung JK (2018) An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nat Biotechnol **36:** 977-982

Grunewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, Joung JK (2019) Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature **569:** 433-437

Grunewald J, Zhou R, Iyer S, Lareau CA, Garcia SP, Aryee MJ, Joung JK (2019) CRISPR DNA base editors with reduced RNA off-target and self-editing activities. Nat Biotechnol **37:** 1041-1048

Gurel F, Zhang Y, Sretenovic S, Qi Y (2020) CRISPR-Cas nucleases and base editors for plant genome editing. aBiotech**:** 1-14

Husbands AY, Chitwood DH, Plavskin Y, Timmermans MC (2009) Signals and prepatterns: new insights into organ polarity in plants. Genes Dev **23:** 1986-1997

Hwang GH, Park J, Lim K, Kim S, Yu J, Yu E, Kim ST, Eils R, Kim JS, Bae S (2018) Web-based design and analysis tools for CRISPR base editing. BMC Bioinformatics **19:** 542

Jiang HS, Lei R, Ding SW, Zhu SF (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. Bmc Bioinformatics **15**

Jin S, Fei H, Zhu Z, Luo Y, Liu J, Gao S, Zhang F, Chen YH, Wang Y, Gao C (2020) Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity. Mol Cell **79:** 728-740 e726

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

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology **37:** 907-+

Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, Chen XY, Kim Y, Beyter D, Krusche P, Saunders CT (2018) Strelka2: fast and accurate calling of germline and somatic variants. Nature Methods **15:** 591-+

Klimek-Chodacka M, Oleszkiewicz T, Lowder LG, Qi Y, Baranski R (2018) Efficient CRISPR/Cas9-based genome editing in carrot cells. Plant Cell Rep **37:** 575-586

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res **22:** 568-576

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature **533:** 420-424

Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH, Liu DR (2017) Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. Sci Adv **3:** eaao4774

Kuang Y, Li S, Ren B, Yan F, Spetz C, Li X, Zhou X, Zhou H (2020) Base-Editing-Mediated Artificial Evolution of OsALS1 In Planta to Develop Novel Herbicide-Tolerant Rice Germplasms. Mol Plant **13:** 565-572

LeBlanc C, Zhang F, Mendez J, Lozano Y, Chatpar K, Irish VF, Jacob Y (2018) Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. Plant J **93:** 377-386

Lee K, Zhang Y, Kleinstiver BP, Guo JA, Aryee MJ, Miller J, Malzahn A, Zarecor S, Lawrence-Dill CJ, Joung JK, Qi Y, Wang K (2019) Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. Plant Biotechnol J **17:** 362-372

Li G, Sretenovic S, Eisenstein E, Coleman G, Qi Y (2021) Highly efficient C‐to‐T and A‐to‐G base editing in a Populus hybrid. Plant Biotechnol J

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27:** 2987-2993

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078-2079

Li J, Sun Y, Du J, Zhao Y, Xia L (2017) Generation of Targeted Point Mutations in Rice by a Modified CRISPR/Cas9 System. Mol Plant **10:** 526-529

Li Z, Xiong X, Wang F, Liang J, Li JF (2019) Gene disruption through base editing-induced messenger RNA missplicing in plants. New Phytol **222:** 1139-1148

Lowder LG, Zhang D, Baltes NJ, Paul JW, Tang X, Zheng X, Voytas DF, Hsieh TF, Zhang Y, Qi Y (2015) A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. Plant Physiol **169:** 971-985

Lu Y, Zhu JK (2017) Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. Mol Plant **10:** 523-525

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research **20:** 1297-1303

Molla KA, Yang Y (2019) CRISPR/Cas-Mediated Base Editing: Technical Considerations and Practical Applications. Trends Biotechnol **37:** 1121-1142

Nekrasov V, Wang C, Win J, Lanz C, Weigel D, Kamoun S (2017) Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. Sci Rep **7:** 482

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

Qin L, Li J, Wang Q, Xu Z, Sun L, Alariqi M, Manghwar H, Wang G, Li B, Ding X, Rui H, Huang H, Lu T, Lindsey K, Daniell H, Zhang X, Jin S (2020) High-efficient and precise base editing of C*G to T*A in the allotetraploid cotton (Gossypium hirsutum) genome using a modified CRISPR/Cas9 system. Plant Biotechnol J **18:** 45-56

Rees HA, Komor AC, Yeh WH, Caetano-Lopes J, Warman M, Edge ASB, Liu DR (2017) Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun **8:** 15790

Ren B, Yan F, Kuang Y, Li N, Zhang D, Zhou X, Lin H, Zhou H (2018) Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. Mol Plant **11:** 623-626

Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, McCann JL, Starrett GJ, Duffy JV, Demir O, Amaro RE, Harki DA, Harris RS, Aihara H (2017) Structural basis for targeted

DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat Struct Mol Biol **24:** 131-139

Shimatani Z, Kashojiya S, Takayama M, Terada R, Arazoe T, Ishii H, Teramura H, Yamamoto T, Komatsu H, Miura K, Ezura H, Nishida K, Ariizumi T, Kondo A (2017) Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat Biotechnol **35:** 441-443

Soyk S, Muller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, Zhang L, Van Eck J, Jimenez-Gomez JM, Lippman ZB (2017) Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. Nat Genet **49:** 162-168

Sretenovic S, Yin D, Levav A, Selengut JD, Mount SM, Qi Y (2021) Expanding plant genome-editing scope by an engineered iSpyMacCas9 system that targets A-rich PAM sequences. Plant Commun **2:** 100101

Stewart CN, Jr., Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques **14:** 748-750

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

Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, Xin X, Zhong Z, Liu B, Zheng X, Zhang D, Malzahn A, Gong Z, Qi Y, Zhang T, Zhang Y (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

Tang X, Ren Q, Yang L, Bao Y, Zhong Z, He Y, Liu S, Qi C, Liu B, Wang Y, Sretenovic S, Zhang Y, Zheng X, Zhang T, Qi Y, Zhang Y (2019) Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. Plant Biotechnol J **17:** 1431-1445

Van Eck J, Keen P, Tjahjadi M (2019) Agrobacterium tumefaciens-Mediated Transformation of Tomato. Methods Mol Biol **1864:** 225-234

Veillet F, Perrot L, Chauvin L, Kermarrec MP, Guyon-Debast A, Chauvin JE, Nogue F, Mazier M (2019) Transgene-Free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor. Int J Mol Sci **20**

Wang X, Ding C, Yu W, Wang Y, He S, Yang B, Xiong YC, Wei J, Li J, Liang J, Lu Z, Zhu W, Wu J, Zhou Z, Huang X, Liu Z, Yang L, Chen J (2020) Cas12a Base Editors Induce Efficient and Specific Editing with Low DNA Damage Response. Cell Rep **31:** 107723

Wang X, Li J, Wang Y, Yang B, Wei J, Wu J, Wang R, Huang X, Chen J, Yang L (2018) Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. Nat Biotechnol **36:** 946-949

Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N (2012) LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Research **40:** 11189-11201

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

Yifhar T, Pekker I, Peled D, Friedlander G, Pistunov A, Sabban M, Wachsman G, Alvarez JP, Amsellem Z, Eshed Y (2012) Failure of the tomato trans-acting short interfering RNA program to regulate AUXIN RESPONSE FACTOR3 and ARF4 underlies the wiry leaf syndrome. Plant Cell **24:** 3575-3589

Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Zhang H, Xu N, Zhu JK (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J **12:** 797-807

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

Zhang Y, Malzahn AA, Sretenovic S, Qi Y (2019) The emerging and uncultivated potential of CRISPR technology in plant science. Nat Plants **5:** 778-794

Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, Bogdanove AJ, Voytas DF (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol **161:** 20-27

Zhong Z, Sretenovic S, Ren Q, Yang L, Bao Y, Qi C, Yuan M, He Y, Liu S, Liu X, Wang J, Huang L, Wang Y, Baby D, Wang D, Zhang T, Qi Y, Zhang Y (2019) Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. Mol Plant

Zhou C, Sun Y, Yan R, Liu Y, Zuo E, Gu C, Han L, Wei Y, Hu X, Zeng R, Li Y, Zhou H, Guo F, Yang H (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 JL, 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 JL, Wang D, 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 LM, Li Y, Yang H (2019) Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science **364:** 289-292

Zuo E, Sun Y, Yuan T, He B, Zhou C, Ying W, Liu J, Wei W, Zeng R, Li Y, Yang H (2020) A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. Nat Methods **17:** 600-604

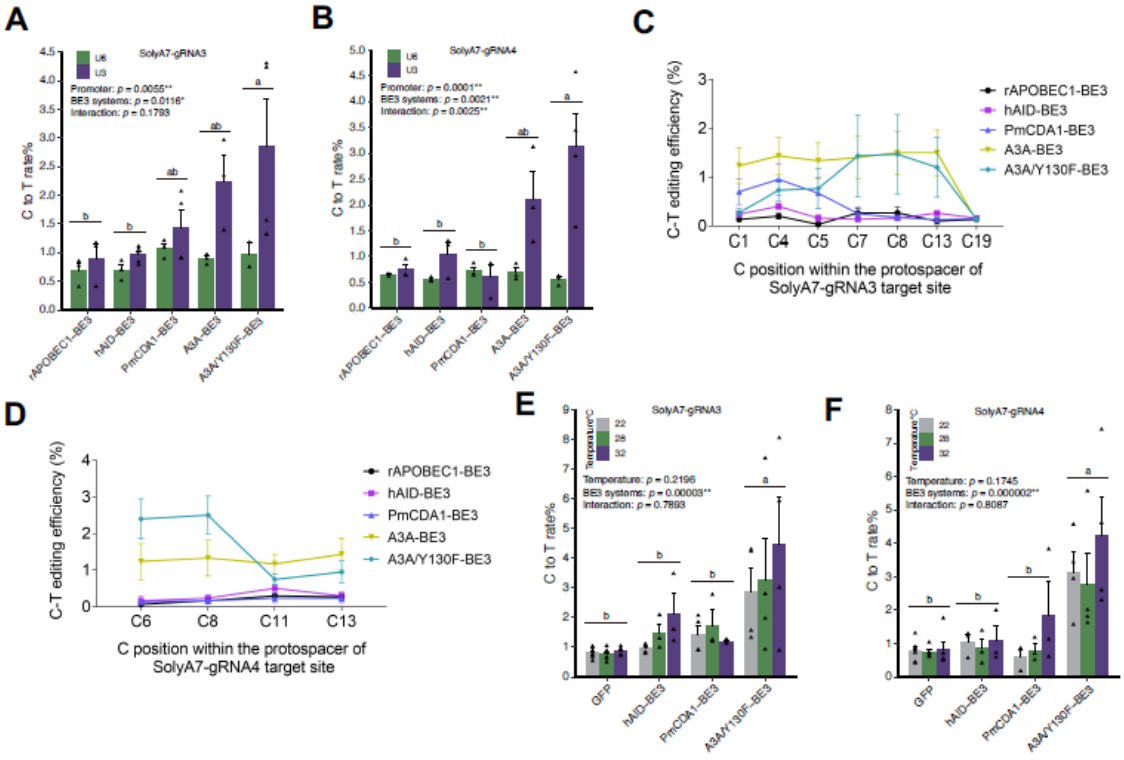

Figure 1. Testing multiple CBEs in tomato protoplasts.
(**A, B**) Assessment of five BE3 base editors in tomato protoplasts at two independent target sites. (**C, D**) C-to-T editing windows of different CBEs at the two target sites. (**E, F**) Assessment of three BE3 base editors in tomato protoplasts at three temperatures at two independent target sites. n=3-5 (except for the A3A/Y130F-AtU3-gRNA3 sample for which only two replicates were used). Error bars represent standard error of the mean (SEM). Significance was calculated using two-way ANOVA analysis of variance test, $^{**}p < 0.01$, $^{*}p < 0.05$; Letters denote statistical differences with Tukey's test ($p < 0.05$).

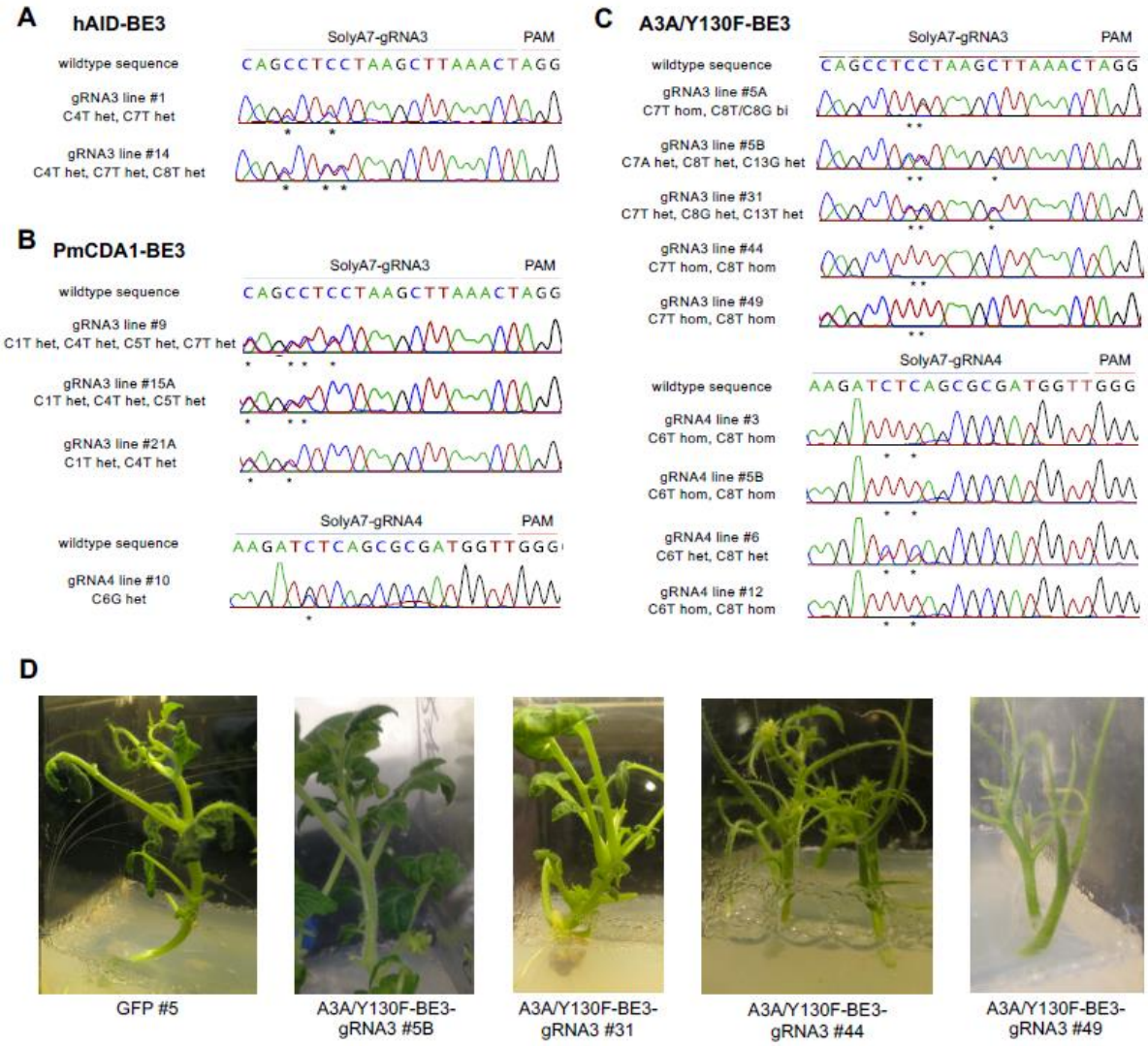

Figure 2. Comparison of three CBEs in stable transgenic tomato lines.
(**A**) Base edited T0 lines at the SolyA7-gRNA3 site by hAID-BE3. (**B**) Base edited T0 lines at the SolyA7-gRNA3 and SolyA7-gRNA4 sites by PmCDA-BE3. (**C**) Base edited T0 lines at the SolyA7-gRNA3 and SolyA7-gRNA4 sites by A3A/Y130F-BE3. Chromatograms of Sanger sequencing shown with base changes indicated by asterisks. (**D**) Phenotype of a GFP control plant, two T0 lines containing heterozygous base editing events, and two T0 lines containing homozygous base-edited alleles.

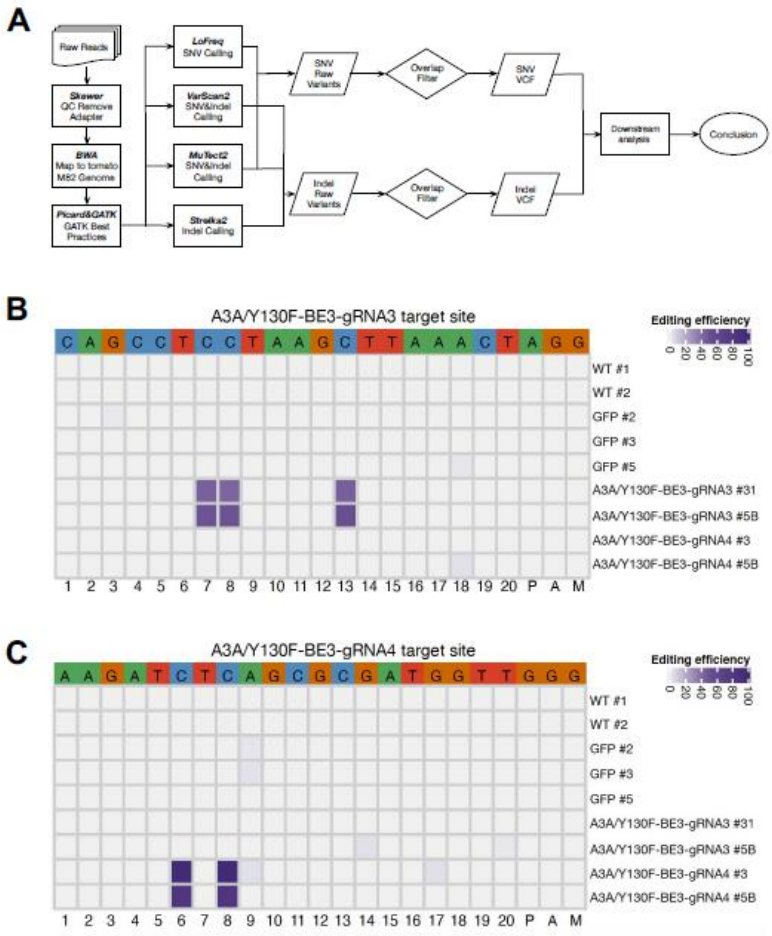

Figure 3. Whole genome sequencing of base-edited plants and control plants.
(**A**) Workflow of whole-genome detection of SNV and indel mutations. SNV analysis involves using three computer programs: LoFreq, VarScan2, and MuTect2. Indel analysis also involves using three programs: VarScan2, MuTect2, and Strelka2. (**B**) Identification of targeted base editing by WGS in lines A3A/Y130F-BE3-gRNA3 #31 and A3A/Y130F-BE3-gRNA3 #5B. (**C**) Identification of targeted base editing by WGS in lines A3A/Y130F-BE3-gRNA4 #3 and A3A/Y130F-BE3-gRNA4 #5B.

A

Nucleotide mismatch	sgRNA	A3A/Y130F-BE3-gRNA3	A3A/Y130F-BE3-gRNA4
≤1	T0 plant 1	0/0	0/0
	T0 plant 2	0/0	0/0
≤2	T0 plant 1	0/0	0/0
	T0 plant 2	0/0	0/0
≤3	T0 plant 1	0/0	0/1
	T0 plant 2	0/0	0/1
≤4	T0 plant 1	0/13	0/10
	T0 plant 2	0/13	0/10
≤5	T0 plant 1	0/155	0/72
	T0 plant 2	0/155	0/72

B

Shared mutations in T0 replication and NOT in GFP plants

sample	A3A/Y130F-BE3-gRNA3	A3A/Y130F-BE3-gRNA4
SNVs	0	1
INDELs	0	0

C

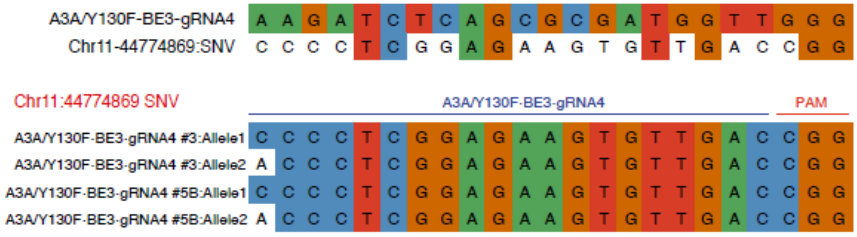

Figure 4. Guide RNA-specific off-targeting effects of A3A/Y130F-BE3 in tomato.
(**A**) Number of off-target sites identified in replicate edited plants vs the number of all potential off-target sites that are predicted by Cas-OFFinder with allowing up to 5-nt mismatch for both gRNA3 and gRNA4. (**B**) Identification of shared DNA SNVs and INDELs between two replicated T_0 edited plants. (**C**) Top panel shows potential off-target sites identified in A3A/Y130F-BE3-gRNA4 samples based on shared mutations within two T_0 plants; bottom panel shows sequence analysis of the shared mutations in A3A/Y130F-BE3-gRNA4 samples.

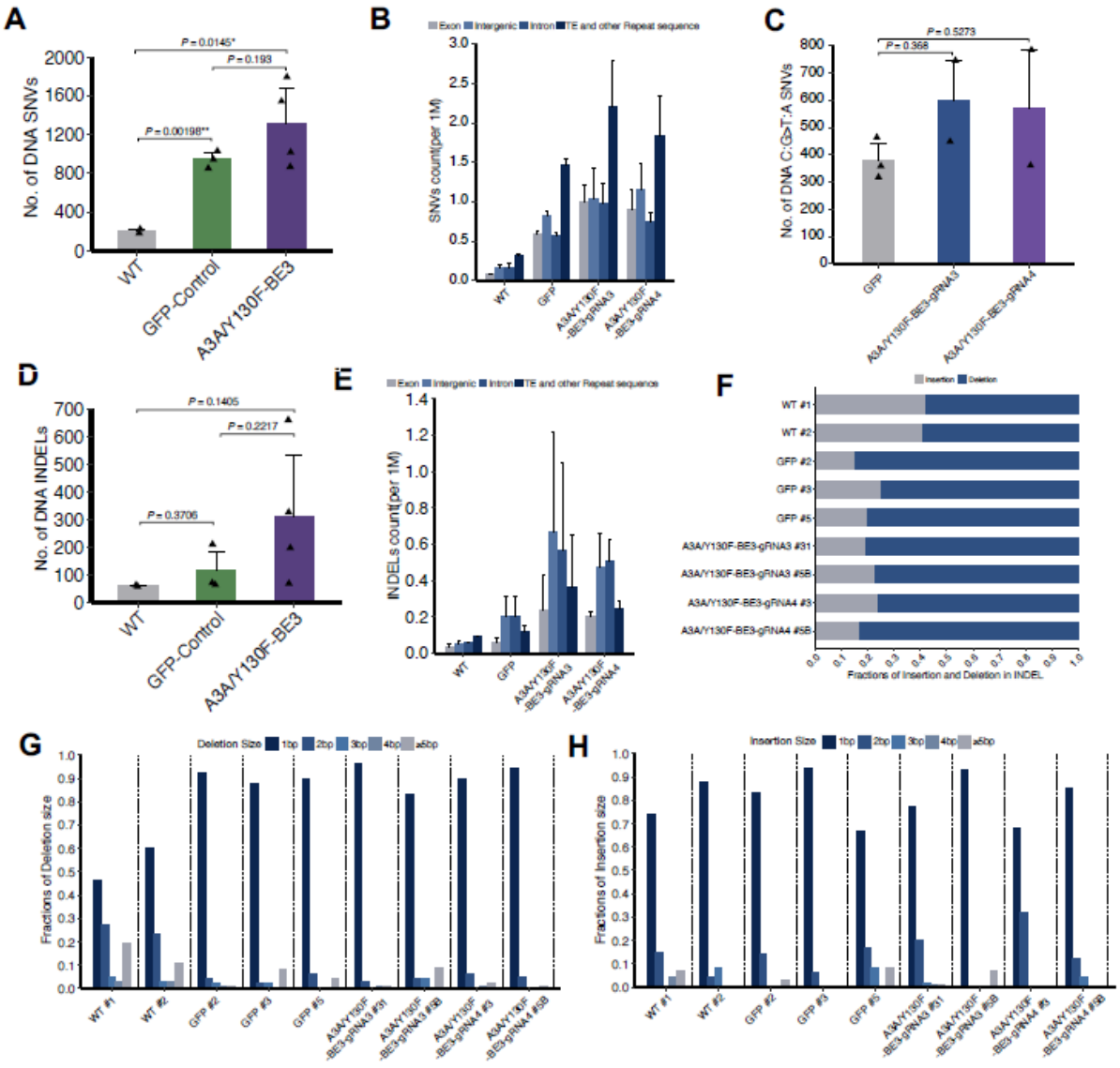

Figure 5. Genome-wide off-target analysis of A3A/Y130F-BE3 in tomato.
(**A**) Total number of DNA SNVs identified in the WT, GFP, and A3A/Y130F-BE3 plants. The average numbers of SNVs in WT, GFP, and A3A/Y130F-BE3 were 203.5, 940, and 1304, respectively. Each triangle represents the number of SNVs from an individual plant. *p* values were calculated by the two-sided unpaired t-test; * $p < 0.05$, ** $p < 0.01$. Error bars represent SEM. (**B**) Annotation of genome-wide distribution of SNVs mutations found in WT, GFP, A3A/Y130F-BE3-gRNA3 and A3A/Y130F-BE3-gRNA4 plants. TE: transposable element. Error bars represent SEM. (**C**) Comparison of total C>T(G>A) SNVs in GFP, A3A/Y130F-BE3-gRNA3 and A3A/Y130F-BE3-gRNA4 plants. The average numbers of SNVs were 380.3, 597.5 and 573, respectively. Error bars represent SEM. (**D**) Total number of DNA indels identified in the WT, GFP, and A3A/Y130F-BE3 plants. Each triangle represents the number of indels from an individual plant. The average numbers of indels in WT, GFP, and A3A/Y130F-BE3 were 61.5, 116.3, and 314.75, respectively. Error bars represent SEM. (**E**) Annotation of genome-wide distribution of indels found in WT, GFP, A3A/Y130F-BE3-gRNA3 and A3A/Y130F-BE3-gRNA4 plants. Error bars represent SEM. (**F**) The fractions of indels in each plant. (**G**) Fractions of deletions of different sizes in each plant. (**H**) Fractions of insertions of different sizes in each plant.

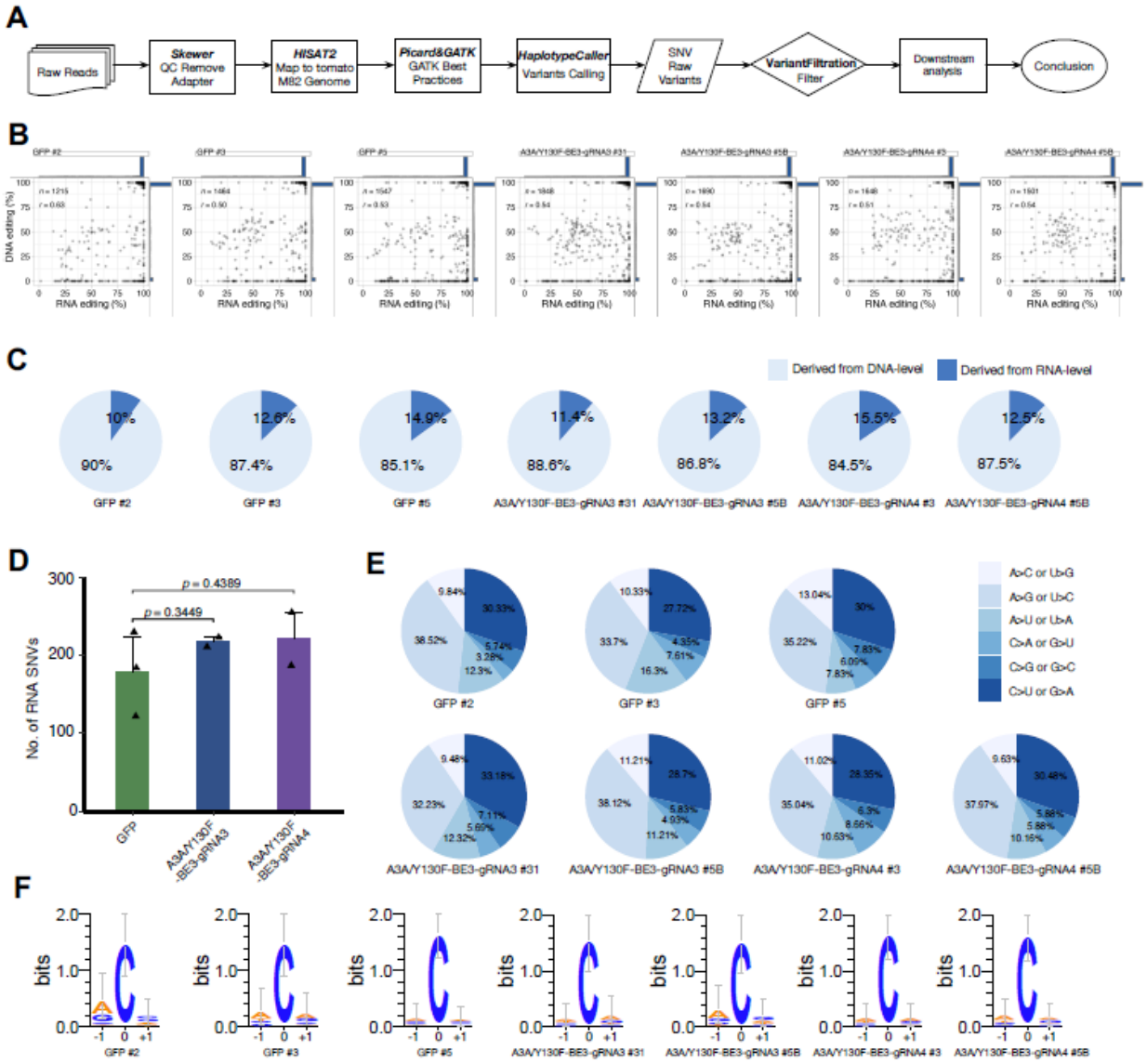

Figure 6. Transcriptome-wide off-target analysis of A3A/Y130F-BE3 in tomato.
(**A**) Workflow of detection of RNA-level SNV mutations. RNA SNVs are identified by HaplotypeCaller. (**B**) Scatter plot correlating RNA mutation rates of RNA SNVs as identified by HaplotypeCaller with DNA mutation rates as determined by WGS. Top histograms depict fractions of RNA SNVs due to RNA-level mutations; right histograms depict fractions of RNA SNVs due to DNA-level mutations. Each dot represents an RNA SNV mutation. The Pearson's correlation was calculated between DNA mutation rates and RNA mutation rates. (**C**) Comparison of RNA-level SNVs in each plant. All identified RNA SNVs are divided into DNA-level SNVs (>5%) and RNA-level SNVs (≤5%) according to their DNA mutation rates. (**D**) Comparison of RNA-level SNV counts in GFP, A3A/Y130F-BE3-gRNA3 and A3A/Y130F-BE3-gRNA4 plants, respectively. *p* values were calculated by the two-sided unpaired t-test; $p < 0.05$ was considered significant. Error bars represent SEM. (**E**) The fraction of RNA-level SNV types in each plant. (**F**) Sequence logos derived from specific strand RNA level SNVs of each plant. Every T depicted should be considered a U in RNA. X-axis 0 indicated mutant cytosines position.

Parsed Citations

Anzalone AV, Koblan LW, Liu DR (2020) Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol 38: 824-844
Google Scholar: Author Only Title Only Author and Title

Bae S, Park J, Kim JS (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30: 1473-1475
Google Scholar: Author Only Title Only Author and Title

Brooks C, Nekrasov V, Lippman ZB, Van Eck J (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. Plant Physiol 166: 1292-1297
Google Scholar: Author Only Title Only Author and Title

Chen L, Li W, Katin-Grazzini L, Ding J, Gu X, Li Y, Gu T, Wang R, Lin X, Deng Z, McAvoy RJ, Gmitter FG, Jr., Deng Z, Zhao Y, Li Y (2018) A method for the production and expedient screening of CRISPR/Cas9-mediated non-transgenic mutant plants. Hortic Res 5: 13
Google Scholar: Author Only Title Only Author and Title

Chen Y, Wang Z, Ni H, Xu Y, Chen Q, Jiang L (2017) CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in Arabidopsis. Sci China Life Sci 60: 520-523
Google Scholar: Author Only Title Only Author and Title

Cheng H, Hao M, Ding B, Mei D, Wang W, Wang H, Zhou R, Liu J, Li C, Hu Q (2021) Base editing with high efficiency in allotetraploid oilseed rape by A3A-PBE system. Plant Biotechnol J 19: 87-97
Google Scholar: Author Only Title Only Author and Title

Cheng TL, Li S, Yuan B, Wang X, Zhou W, Qiu Z (2019) Expanding C-T base editing toolkit with diversified cytidine deaminases. Nat Commun 10: 3612
Google Scholar: Author Only Title Only Author and Title

Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology 31: 213-219
Google Scholar: Author Only Title Only Author and Title

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Research 14: 1188-1190
Google Scholar: Author Only Title Only Author and Title

Doman JL, Raguram A, Newby GA, Liu DR (2020) Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. Nat Biotechnol 38: 620-628
Google Scholar: Author Only Title Only Author and Title

Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang DL, Wang Z, Zhang Z, Zheng R, Yang L, Zeng L, Liu X, Zhu JK (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proc Natl Acad Sci U S A 111: 4632-4637
Google Scholar: Author Only Title Only Author and Title

Floss DS, Levy JG, Levesque-Tremblay V, Pumplin N, Harrison MJ (2013) DELLA proteins regulate arbuscule formation in arbuscular mycorrhizal symbiosis. Proc Natl Acad Sci U S A 110: E5025-5034
Google Scholar: Author Only Title Only Author and Title

Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, Pinello L, Joung JK (2018) An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nat Biotechnol 36: 977-982
Google Scholar: Author Only Title Only Author and Title

Grunewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, Joung JK (2019) Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature 569: 433-437
Google Scholar: Author Only Title Only Author and Title

Grunewald J, Zhou R, Iyer S, Lareau CA, Garcia SP, Aryee MJ, Joung JK (2019) CRISPR DNA base editors with reduced RNA off-target and self-editing activities. Nat Biotechnol 37: 1041-1048
Google Scholar: Author Only Title Only Author and Title

Gurel F, Zhang Y, Sretenovic S, Qi Y (2020) CRISPR-Cas nucleases and base editors for plant genome editing. aBiotech: 1-14
Google Scholar: Author Only Title Only Author and Title

Husbands AY, Chitwood DH, Plavskin Y, Timmermans MC (2009) Signals and prepatterns: new insights into organ polarity in plants. Genes Dev 23: 1986-1997
Google Scholar: Author Only Title Only Author and Title

Hwang GH, Park J, Lim K, Kim S, Yu J, Yu E, Kim ST, Eils R, Kim JS, Bae S (2018) Web-based design and analysis tools for CRISPR base editing. BMC Bioinformatics 19: 542
Google Scholar: Author Only Title Only Author and Title

Jiang HS, Lei R, Ding SW, Zhu SF (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. Bmc Bioinformatics 15
Google Scholar: Author Only Title Only Author and Title

Jin S, Fei H, Zhu Z, Luo Y, Liu J, Gao S, Zhang F, Chen YH, Wang Y, Gao C (2020) Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity. Mol Cell 79: 728-740 e726
Google Scholar: Author Only Title Only Author and Title

Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, Liang C, Wang D, Qiu JL, Zhang F, Gao C (2019) Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science 364: 292-295
Google Scholar: Author Only Title Only Author and Title

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology 37: 907-+
Google Scholar: Author Only Title Only Author and Title

Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, Chen XY, Kim Y, Beyter D, Krusche P, Saunders CT (2018) Strelka2: fast and accurate calling of germline and somatic variants. Nature Methods 15: 591-+
Google Scholar: Author Only Title Only Author and Title

Klimek-Chodacka M, Oleszkiewicz T, Lowder LG, Qi Y, Baranski R (2018) Efficient CRISPR/Cas9-based genome editing in carrot cells. Plant Cell Rep 37: 575-586
Google Scholar: Author Only Title Only Author and Title

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22: 568-576
Google Scholar: Author Only Title Only Author and Title

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533: 420-424

Google Scholar: Author Only Title Only Author and Title

Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH, Liu DR (2017) Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. Sci Adv 3: eaao4774
Google Scholar: Author Only Title Only Author and Title

Kuang Y, Li S, Ren B, Yan F, Spetz C, Li X, Zhou X, Zhou H (2020) Base-Editing-Mediated Artificial Evolution of OsALS1 In Planta to Develop Novel Herbicide-Tolerant Rice Germplasms. Mol Plant 13: 565-572
Google Scholar: Author Only Title Only Author and Title

LeBlanc C, Zhang F, Mendez J, Lozano Y, Chatpar K, Irish VF, Jacob Y (2018) Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. Plant J 93: 377-386
Google Scholar: Author Only Title Only Author and Title

Lee K, Zhang Y, Kleinstiver BP, Guo JA, Aryee MJ, Miller J, Malzahn A, Zarecor S, Lawrence-Dill CJ, Joung JK, Qi Y, Wang K (2019) Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. Plant Biotechnol J 17: 362-372
Google Scholar: Author Only Title Only Author and Title

Li G, Sretenovic S, Eisenstein E, Coleman G, Qi Y (2021) Highly efficient C-to-T and A-to-G base editing in a Populus hybrid. Plant Biotechnol J
Google Scholar: Author Only Title Only Author and Title

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987-2993
Google Scholar: Author Only Title Only Author and Title

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595
Google Scholar: Author Only Title Only Author and Title

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079
Google Scholar: Author Only Title Only Author and Title

Li J, Sun Y, Du J, Zhao Y, Xia L (2017) Generation of Targeted Point Mutations in Rice by a Modified CRISPR/Cas9 System. Mol Plant 10: 526-529
Google Scholar: Author Only Title Only Author and Title

Li Z, Xiong X, Wang F, Liang J, Li JF (2019) Gene disruption through base editing-induced messenger RNA missplicing in plants. New Phytol 222: 1139-1148
Google Scholar: Author Only Title Only Author and Title

Lowder LG, Zhang D, Baltes NJ, Paul JW, Tang X, Zheng X, Voytas DF, Hsieh TF, Zhang Y, Qi Y (2015) A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. Plant Physiol 169: 971-985
Google Scholar: Author Only Title Only Author and Title

Lu Y, Zhu JK (2017) Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. Mol Plant 10: 523-525
Google Scholar: Author Only Title Only Author and Title

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20: 1297-1303
Google Scholar: Author Only Title Only Author and Title

Molla KA, Yang Y (2019) CRISPR/Cas-Mediated Base Editing: Technical Considerations and Practical Applications. Trends Biotechnol 37: 1121-1142
Google Scholar: Author Only Title Only Author and Title

Nekrasov V, Wang C, Win J, Lanz C, Weigel D, Kamoun S (2017) Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. Sci Rep 7: 482
Google Scholar: Author Only Title Only Author and Title

Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, Shimatani Z, Kondo A (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353
Google Scholar: Author Only Title Only Author and Title

Qin L, Li J, Wang Q, Xu Z, Sun L, Alariqi M, Manghwar H, Wang G, Li B, Ding X, Rui H, Huang H, Lu T, Lindsey K, Daniell H, Zhang X, Jin S (2020) High-efficient and precise base editing of C*G to T*A in the allotetraploid cotton (Gossypium hirsutum) genome using a modified CRISPR/Cas9 system. Plant Biotechnol J 18: 45-56
Google Scholar: Author Only Title Only Author and Title

Rees HA, Komor AC, Yeh WH, Caetano-Lopes J, Warman M, Edge ASB, Liu DR (2017) Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun 8: 15790
Google Scholar: Author Only Title Only Author and Title

Ren B, Yan F, Kuang Y, Li N, Zhang D, Zhou X, Lin H, Zhou H (2018) Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. Mol Plant 11: 623-626
Google Scholar: Author Only Title Only Author and Title

Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, McCann JL, Starrett GJ, Duffy JV, Demir O, Amaro RE, Harki DA, Harris RS, Aihara H (2017) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat Struct Mol Biol 24: 131-139
Google Scholar: Author Only Title Only Author and Title

Shimatani Z, Kashojiya S, Takayama M, Terada R, Arazoe T, Ishii H, Teramura H, Yamamoto T, Komatsu H, Miura K, Ezura H, Nishida K, Ariizumi T, Kondo A (2017) Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat Biotechnol 35: 441-443
Google Scholar: Author Only Title Only Author and Title

Soyk S, Muller NA, Park SJ, Schmalenbach I, Jiang K, Hayama R, Zhang L, Van Eck J, Jimenez-Gomez JM, Lippman ZB (2017) Variation in the flowering gene SELF PRUNING 5G promotes day-neutrality and early yield in tomato. Nat Genet 49: 162-168
Google Scholar: Author Only Title Only Author and Title

Sretenovic S, Yin D, Levav A, Selengut JD, Mount SM, Qi Y (2021) Expanding plant genome-editing scope by an engineered iSpyMacCas9 system that targets A-rich PAM sequences. Plant Commun 2: 100101
Google Scholar: Author Only Title Only Author and Title

Stewart CN, Jr., Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques 14: 748-750
Google Scholar: Author Only Title Only Author and Title

Tan J, Zhang F, Karcher D, Bock R (2020) Expanding the genome-targeting scope and the site selectivity of high-precision base editors. Nat Commun 11: 629
Google Scholar: Author Only Title Only Author and Title

Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, Xin X, Zhong Z, Liu B, Zheng X, Zhang D, Malzahn A, Gong Z, Qi Y, Zhang T, Zhang Y (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
Google Scholar: Author Only Title Only Author and Title

Tang X, Ren Q, Yang L, Bao Y, Zhong Z, He Y, Liu S, Qi C, Liu B, Wang Y, Sretenovic S, Zhang Y, Zheng X, Zhang T, Qi Y, Zhang Y (2019) Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. Plant Biotechnol J 17: 1431-1445
Google Scholar: Author Only Title Only Author and Title

Van Eck J, Keen P, Tjahjadi M (2019) Agrobacterium tumefaciens-Mediated Transformation of Tomato. Methods Mol Biol 1864: 225-234
Google Scholar: Author Only Title Only Author and Title

Veillet F, Perrot L, Chauvin L, Kermarrec MP, Guyon-Debast A, Chauvin JE, Nogue F, Mazier M (2019) Transgene-Free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor. Int J Mol Sci 20
Google Scholar: Author Only Title Only Author and Title

Wang X, Ding C, Yu W, Wang Y, He S, Yang B, Xiong YC, Wei J, Li J, Liang J, Lu Z, Zhu W, Wu J, Zhou Z, Huang X, Liu Z, Yang L, Chen J (2020) Cas12a Base Editors Induce Efficient and Specific Editing with Low DNA Damage Response. Cell Rep 31: 107723
Google Scholar: Author Only Title Only Author and Title

Wang X, Li J, Wang Y, Yang B, Wei J, Wu J, Wang R, Huang X, Chen J, Yang L (2018) Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. Nat Biotechnol 36: 946-949
Google Scholar: Author Only Title Only Author and Title

Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N (2012) LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Research 40: 11189-11201
Google Scholar: Author Only Title Only Author and Title

Xue C, Zhang H, Lin Q, Fan R, Gao C (2018) Manipulating mRNA splicing by base editing in plants. Sci China Life Sci 61: 1293-1300
Google Scholar: Author Only Title Only Author and Title

Yifhar T, Pekker I, Peled D, Friedlander G, Pistunov A, Sabban M, Wachsman G, Alvarez JP, Amsellem Z, Eshed Y (2012) Failure of the tomato trans-acting short interfering RNA program to regulate AUXIN RESPONSE FACTOR3 and ARF4 underlies the wiry leaf syndrome. Plant Cell 24: 3575-3589
Google Scholar: Author Only Title Only Author and Title

Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Zhang H, Xu N, Zhu JK (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J 12: 797-807
Google Scholar: Author Only Title Only Author and Title

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

Zhang Y, Malzahn AA, Sretenovic S, Qi Y (2019) The emerging and uncultivated potential of CRISPR technology in plant science. Nat Plants 5: 778-794
Google Scholar: Author Only Title Only Author and Title

Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, Bogdanove AJ, Voytas DF (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol 161: 20-27
Google Scholar: Author Only Title Only Author and Title

Zhong Z, Sretenovic S, Ren Q, Yang L, Bao Y, Qi C, Yuan M, He Y, Liu S, Liu X, Wang J, Huang L, Wang Y, Baby D, Wang D, Zhang T, Qi Y, Zhang Y (2019) Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. Mol Plant
Google Scholar: Author Only Title Only Author and Title

Zhou C, Sun Y, Yan R, Liu Y, Zuo E, Gu C, Han L, Wei Y, Hu X, Zeng R, Li Y, Zhou H, Guo F, Yang H (2019) Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. Nature 571: 275-278
Google Scholar: Author Only Title Only Author and Title

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

Zong Y, Wang Y, Li C, Zhang R, Chen K, Ran Y, Qiu JL, Wang D, Gao C (2017) Precise base editing in rice, wheat and maize with a Cas9- cytidine deaminase fusion. Nat Biotechnol 35: 438-440
Google Scholar: Author Only Title Only Author and Title

Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, Yuan L, Steinmetz LM, Li Y, Yang H (2019) Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science 364: 289-292
Google Scholar: Author Only Title Only Author and Title

Zuo E, Sun Y, Yuan T, He B, Zhou C, Ying W, Liu J, Wei W, Zeng R, Li Y, Yang H (2020) A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. Nat Methods 17: 600-604
Google Scholar: Author Only Title Only Author and Title