



# An unbiased method for evaluating the genome-wide specificity of base editors in rice

Shuai Jin<sup>1,2</sup>, Qiang Gao<sup>3</sup> and Caixia Gao<sup>1,2</sup>

**Base editors can achieve targeted genomic base conversion. However, the off-target issue is one of the major concerns in their application. Whole-genome sequencing (WGS) at the individual level can provide direct information on genome-wide specificity, but it is difficult to distinguish true off-target single-nucleotide variants (SNVs) induced by base editors from background variation. Here we describe an unbiased WGS method for evaluating the specificity of base editors in rice. In this protocol, we describe the experimental design and provide details of vector construction, rice transformation and tissue culture, as well as a comprehensive WGS data analysis pipeline for overcoming two related core problems in various plant species: high background mutation rates and the heterogeneity of examined populations. Using this protocol, researchers can straightforwardly and accurately assess the genome-wide specificity of base editors and other genome editing tools in 12–15 weeks.**

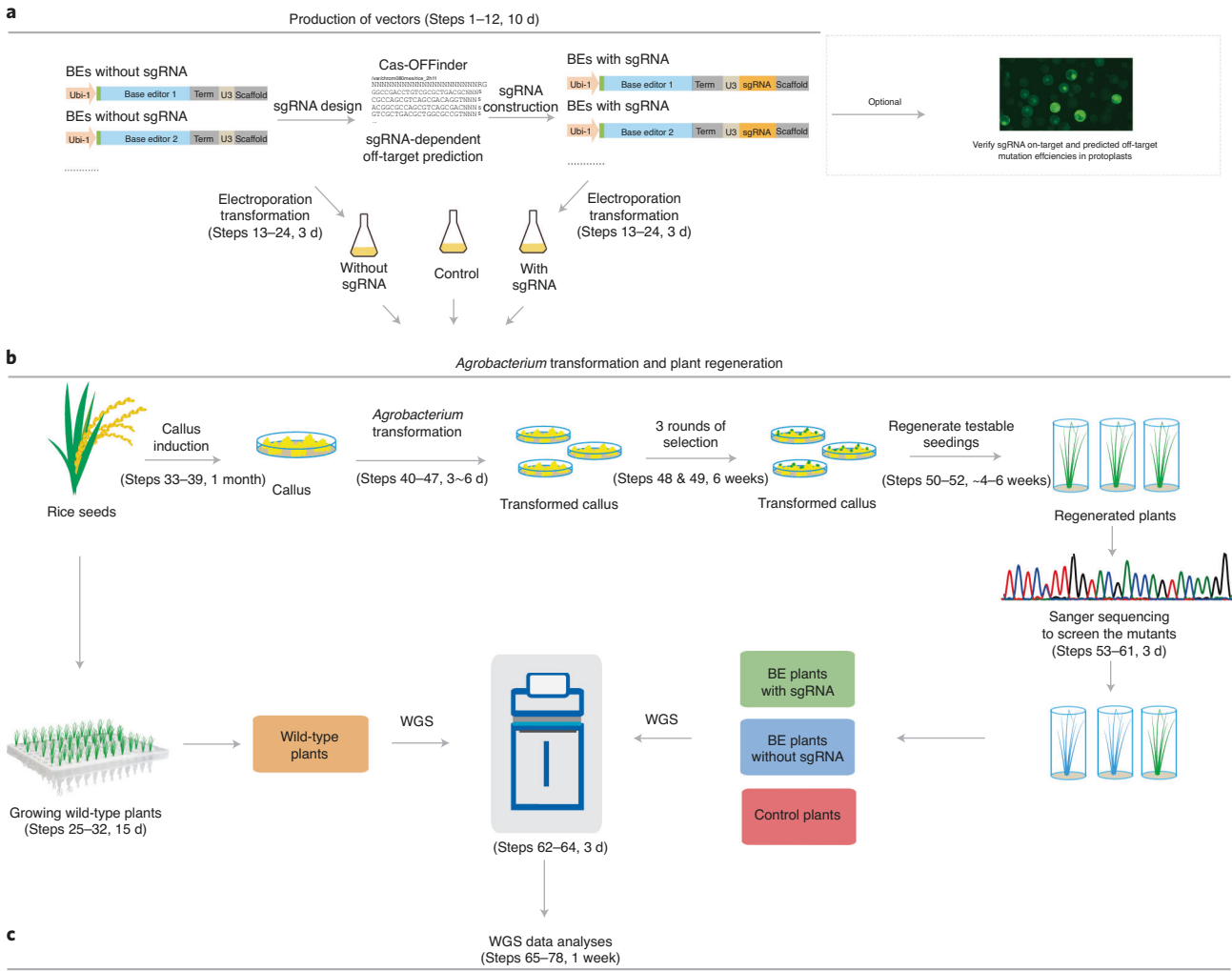
## Introduction

Cytosine base editors (CBEs) and adenine base editors (ABEs) are parts of the clustered regularly interspaced short palindromic repeats (CRISPR)-based genome editing toolbox<sup>1–3</sup>. They catalyze C-to-T and A-to-G base conversions, respectively, at target sites, without requiring double-strand breaks or donor DNA templates, and are widely used in a variety of systems, such as human cells<sup>1–3</sup>, mice<sup>4,5</sup>, rabbits<sup>6</sup>, zebrafish<sup>7,8</sup>, rice<sup>9–11</sup> and wheat<sup>9–11</sup>. Base editors can correct mutational errors and introduce desired point mutations and thus have enormous potential in therapeutics and agricultural applications<sup>12–15</sup>. However, off-target effects are the greatest technical barrier to the clinical application of CRISPR-based genome editing technologies<sup>16–18</sup>. These off-target effects can be divided into two types: (i) sgRNA-dependent off-target mutations (the CRISPR-based genome editing tools induce these off-target mutations because of similarity between the sequence targeted by the relevant sgRNA and the off-target sequences<sup>16,17</sup>); and (ii) sgRNA-independent off-target mutations; several groups have reported that the cytosine base editors (BE3 and BE4) induce genome-wide sgRNA-independent DNA off-target mutations in rice<sup>19</sup>, mice<sup>20,21</sup> and human cells<sup>22</sup> because of the high binding affinity of the cytosine deaminase for single-stranded DNA (ssDNA)<sup>19–22</sup>. In this protocol, both types of off-target effects are detected.

In silico, in vitro and in vivo methods have been developed for assessing the specificity of CRISPR systems in eukaryotic cells<sup>16,17</sup>. The in silico approach is used to predict potential off-target sites on the basis of the similarity between on-target and off-target sequences<sup>16,17</sup>. The in vitro and in vivo approaches, such as IDLV<sup>23,24</sup>, HTGTS<sup>25</sup>, Digenome-seq<sup>26,27</sup>, SITE-seq<sup>28</sup>, CIRCLE-seq<sup>29</sup>, BLESS<sup>30</sup>, GUIDE-seq<sup>31</sup> and DISCOVER-seq<sup>32</sup>, are widely used for high-throughput detection of genome-wide off-target events. However, all these approaches depend on capturing in vitro or in vivo DSBs (double-strand breaks) and do not detect SNVs, which are the actual products of base editing.

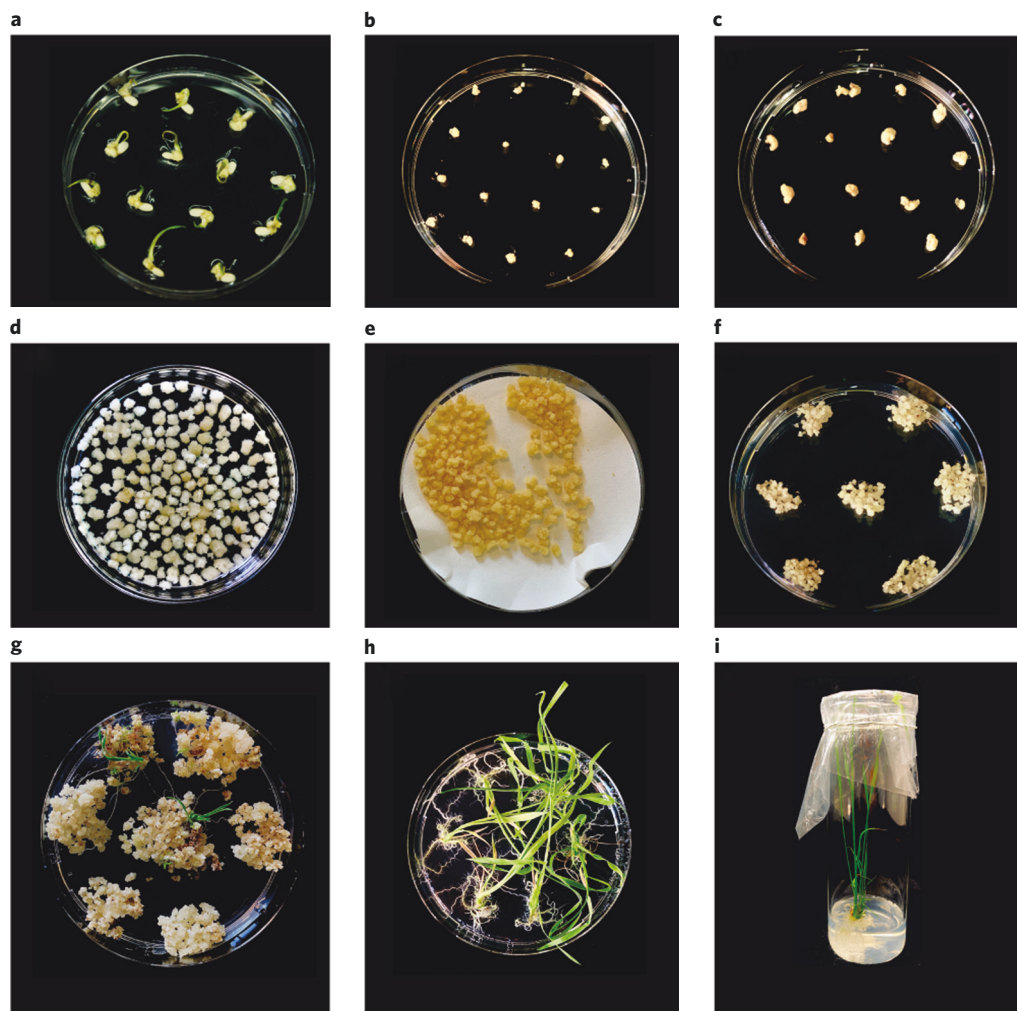
With the increased attention to base editing in therapeutics, plant breeding and research, a simple and accurate method for assessing the genome-wide specificity of base editors in vivo is urgently needed. WGS at the individual level can provide information on genome-wide variants; this approach has been used for assessing the specificity of CRISPR-associated (Cas)9/Cas12a-induced knockouts<sup>33–36</sup>. On the other hand, off-target SNVs are difficult to identify because they are hard to distinguish from the high mutation background in the samples analyzed<sup>37</sup>. Here we describe an unbiased WGS method for evaluating the specificity of base editing in rice (Fig. 1). We use japonica

<sup>1</sup>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. <sup>2</sup>College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China. <sup>3</sup>BGI Genomics, BGI-Shenzhen, Shenzhen, China. ✉e-mail: [cxgao@genetics.ac.cn](mailto:cxgao@genetics.ac.cn)



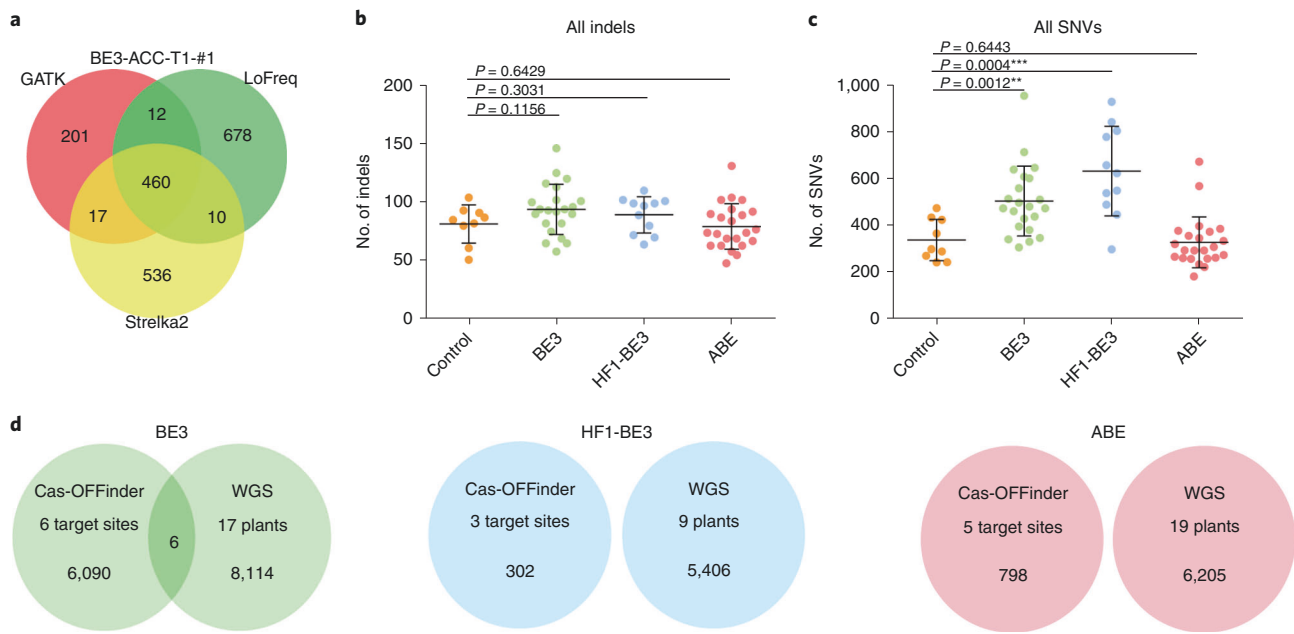
rice (*Oryza sativa*) as the experimental organism because it has a relatively small genome (380+ megabases (Mb)), and a high-quality reference sequence is available<sup>38</sup>; furthermore, individual transgenic plants can be regenerated from calluses derived from single cells to ensure they have the

◀ **Fig. 1 | Overview of WGS for evaluating the genome-wide specificity of base editors.** Vector construction, *Agrobacterium*-mediated rice callus transformation, water culture of wild-type rice plants, callus tissue culture and WGS data analysis workflow are illustrated. **a**, First the base editor vectors containing the deaminase sequence are constructed, and then the sgRNA sequence is added (Steps 1–11). Optionally, sgRNA-dependent and sgRNA-independent edits can be examined in a transient rice protoplast assay (Step 12). Then the vectors are transformed into *Agrobacterium* (Steps 13–24). **b**, The rice seeds are harvested and wild-type plants are cultured in water (Steps 25–32). Calluses are induced from mature seeds after 2 weeks (Steps 33–39), and the induced calluses are transformed with *Agrobacterium* (Steps 40–47). Three rounds of hygromycin selection of successfully transformed calluses and tissue culture of edited calluses are conducted over 6 weeks (Steps 48 & 49), after which regenerated edited seedlings (Steps 50–52) can be examined by Sanger sequencing or WGS. Sanger sequencing is used to screen mutants (Steps 53–61), and selected plants are analyzed by WGS (Steps 62–64). **c**, Clean data are processed with BWA, Samtools, Picard and GATK (Steps 65–69), using the packages GATK, Strelka2 and LoFreq to call SNVs, and GATK and Strelka2 to call indels (Steps 70–78). **a–c**, adapted from Jin et al.<sup>19</sup>, American Association for the Advancement of Science.



**Fig. 2 | *Agrobacterium tumefaciens*-mediated transformation of rice calluses and plant regeneration.** **a**, Seven days after cultivating the seeds, developed calluses are visible on mature seeds. **b**, Ten days later, yellow, compact calluses are sub-cultured in fresh medium. **c**, Calluses after a 2-week culture cycle. **d**, The calluses are incubated in an *A. tumefaciens* suspension for 30 min on a clean bench. **e**, *Agrobacterium*-infected rice calluses spread on a sterile filter paper. **f**, Healthy, yellowish calluses in fresh solid M2 medium in the second round of selection. **g**, Seedlings from resistant calluses are visible. **h**, Seedlings 2 weeks after the transfer to new plates in solid M4 medium. **i**, The seedlings after transfer to containers in solid M4 medium. **a, b**, adapted from Jin et al.<sup>19</sup>, American Association for the Advancement of Science.

same genetic background (Fig. 2). In this way, the problem of population heterogeneity should be easier to overcome. Below we describe the detailed experimental design and WGS data analysis pipeline for assessing the specificity of base editors and other genome editing tools in rice.



**Fig. 3 | Typical results.** **a**, A Venn diagram of SNVs detected by the three software packages. **b,c**, Numbers of indels (**b**) and total SNVs (**c**) identified in BE3-, HF1-BE3- and ABE-derived plants. Each dot represents the number of indels or SNVs in an individual plant. All error bars represent mean  $\pm$  SD. The horizontal lines represent medians. Nine control plants, 22 BE3-derived plants, 11 HF1-BE3-derived plants and 23 ABE-derived plants were analyzed. The average numbers of indels in BE3, HF1-BE3 ABE, and control plants were 94, 89, 79 and 82, respectively, whereas the average numbers of total SNVs were 504, 632, 327 and 338, respectively. **d**, Venn diagrams of the SNVs identified in each base editor group by WGS, together with the off-target sites predicted by Cas-OFFinder. Only six SNVs were identified by both WGS and Cas-OFFinder in the BE3 group, and no SNVs were identified by both WGS and Cas-OFFinder in the HF1-BE3 and ABE groups, indicating that the SNVs identified by WGS were sgRNA-independent off-target mutations. *P* values were calculated with the Mann-Whitney test; *P* < 0.05 was considered significant. All values represent means  $\pm$  SD. \*\**P* < 0.01; \*\*\**P* < 0.001. Adapted from Jin et al.<sup>19</sup>, American Association for the Advancement of Science.

### Applications of the method

This method can be easily and accurately used to evaluate the specificity of base editors in rice and other plants. When the method is applied in other organisms, the only change necessary is the replacement of the reference genome. Theoretically, it could also be used to evaluate the specificity of other genome editing tools, such as ZFNs<sup>39,40</sup>, TALENs<sup>41–43</sup> and CRISPR-based technologies<sup>15,44,45</sup>.

### Development of the protocol

Off-target indels (insertions and deletions) induced by CRISPR/Cas9 have been detected by WGS in edited human cells<sup>46</sup>, mice<sup>34,35</sup>, *Arabidopsis*<sup>47</sup>, rice<sup>36</sup>, maize<sup>48</sup>, tomato<sup>49</sup> and cotton<sup>33</sup>. However, the SNV variants<sup>37</sup> induced by base editors have not been detected by WGS.

Here we describe a simple, accurate and systematic method for evaluating the specificity of base editing in rice. There are several factors that are key to the method (Fig. 1): (i) Use of a high-quality reference genome (here the genome of the rice variety Zhonghua 11) to decrease the genetic differences between the sequenced variety and the reference sequence. (ii) Induction of calluses from rice seeds, all of which are derived from one plant, so as to reduce the heterogeneity of the plant population (Fig. 2). (iii) Sequencing of wild-type plants to further ‘train’ the reference genome, shown as the orange box in Fig. 1b. (iv) Use of *Agrobacterium*-transformed plants as the optimum control group, shown as the red box in Fig. 1b.

Using this protocol, we found that by sequencing base editor-treated plants and control plants that the most frequently used cytosine base editor, BE3 (the original cytosine base editor), but not the adenine base editor, ABE7.10, induced substantial numbers of genome-wide sgRNA-independent off-target C>T SNVs (Fig. 3). Moreover, we found that these off-target SNVs were enriched in genic regions, which increases concern about the specificity of base editing<sup>19</sup>. Thus, further optimization of the specificity of this cytosine base editor is desirable, and the specificities of other base editors also need to be assessed.

### An overview of the procedure

The entire workflow is summarized in Fig. 1 and essentially comprises the following elements: production of vectors (Steps 1–12), transformation of *Agrobacterium* by electroporation (Steps 13–24), harvesting of rice seeds and growing wild-type plants (Steps 25–32), growing calluses from mature seeds (Steps 33–39), infection and co-cultivation of calluses and *Agrobacterium tumefaciens* (Steps 40–47), tissue culture (Steps 48 & 49), regenerating testable seedlings (Steps 50–52), screening and identification of mutants (Steps 53–61), performing WGS (Steps 62–64), processing of raw reads (Steps 65–69), variant calling (Steps 70–72), filtering out of wild-type mutations and on-target mutations and obtaining the intersection of results using three software packages (Steps 73–78); and finally, Sanger sequencing to confirm the SNVs and indels identified by WGS (Steps 79–83).

### Comparison with other methods

Previous experimental methods used to assess the specificity of genome editing tools can be divided into three groups: methods based on in vitro digestion of genomic DNA, methods based on in vitro or in vivo DSB capture, and those based on in vivo WGS. These methods have greatly facilitated research on the specificity of genome editing but are not suitable for assessing the specificity of base editors in vivo.

Digenome-seq<sup>26,27</sup>, CIRCLE-seq<sup>29</sup>, EndoV<sup>50</sup>, and related methods<sup>51–53</sup> involve digesting genomic DNA with Cas9–sgRNA or BE–sgRNA nucleoprotein complexes in vitro. For example, Digenome-seq and its extensions can identify the genome-wide off-target effects of multiple genome editing tools (Cas9, Cpf1, CBE and ABE) by in vitro digestion of genomic DNA and computational identification of cleavage sites. However, base editors work only on ssDNA, which these methods cannot detect within the dynamic environment of the cell<sup>16,17</sup>.

DSB-capture in vivo methods include IDLV<sup>23,24</sup> (based on capture of integrase-defective lentiviral vector at DSB sites), HTGTS<sup>25</sup> (high-throughput genome-wide translocation sequencing, capturing translocations of DSB sites), BLESS<sup>54</sup> (break labeling, enrichment on streptavidin and next-generation sequencing, based on biotin labeling of DSB sites), SITE-seq<sup>28</sup> (selective enrichment and identification of adaptor-tagged DNA ends by sequencing), GUIDE-seq<sup>31</sup> (genome-wide unbiased identification of DSBs enabled by sequencing, using dsODN tags to capture DSB sites), and DISCOVER-seq<sup>32</sup> (discovery of in situ Cas off-targets and verification by sequencing). For example, DISCOVER-seq detects the effects of CRISPR-based knockouts by recruiting DNA repair factors to DSB sites. These methods focus on capturing genome-wide DSB sites but are not suitable for detecting genome-wide SNVs.

Previously reported in vivo WGS-based methods<sup>33,36</sup> can also evaluate the specificity of Cas9 and Cpf1 in many organisms in either a biased way (focusing on sgRNA-dependent sites) or an unbiased way (genome-wide evaluation). However, these approaches are difficult because of the high background mutation rates (spontaneous SNVs), especially in large, complex genomes. Recently, another unbiased in vivo tool for evaluating specificity of genome editing, namely GOTI<sup>20</sup>, has demonstrated an acceptable level of accuracy for detecting the specificity of genome editing tools by directly comparing FACS-sorted edited and non-edited cells derived from the same embryos. The method has been used in mouse embryos, yielding results for the genome-wide specificity of base editors similar to those we obtained<sup>19</sup>. The numbers of background mutations are lower in GOTI than in the plant WGS assay described in this protocol. But GOTI is used with 14.5-d mouse embryos rather than mature individual organisms and cannot be used with plant cells.

### Limitations of the method

#### Limited range of application

This method is difficult in species with large, complex genomes and heterogeneous populations, such as wheat (*Triticum aestivum*, 17 Gb), barley (*Hordeum vulgare*, 4.6 Gb) and maize (*Zea mays*, 2.1 Gb)<sup>55</sup>. Even if wild-type individuals are used to filter background SNVs, the off-target SNVs may still be masked by population heterogeneity. In species with low complexity and small genomes, such as *Arabidopsis thaliana* (125 Mb), green algae (*Chlamydomonas reinhardtii*, 110 Mb), and moss (*Physcomitrella patens*, 473 Mb)<sup>55</sup>, reducing numbers of background mutations would make them easier to analyze (the genome sizes of these species were obtained from the Phytozome database<sup>55</sup>, <https://phytozome.jgi.doe.gov/pz/portal.html#>).

#### Time

About 3–4 months is needed from *Agrobacterium*-mediated transformation to seedling regeneration.

## Experimental design

### Vector construction

Binary vectors are needed for *Agrobacterium*-mediated rice callus transformation. We use a published binary vector, pH-CBE<sup>19</sup>, as backbone vector; it contains the SpCas9-based sgRNA scaffold and the base editor BE3 (APOBEC1-linker-nSpCas9-UGI fusion). The APOBEC1 deaminase sequence or the total BE3 sequence can be replaced by Gibson cloning to create new base editors (Steps 1–9). To construct base-editing vectors with other deaminases, the deaminase sequences are codon-optimized for cereal plants and synthesized commercially (Step 1), and the fusion protein sequences are cloned into the backbone of the pH-CBE vector (Steps 2–9). These are referred to as ‘BEs without sgRNA’ (Fig. 1a). A total of 3–5 sgRNA sequences can then be designed for each base editor. Forward and reverse sgRNA oligos are synthesized and cloned into the base-editing vectors using T4 ligase (Steps 10–12). These vectors are referred to as ‘BEs with sgRNAs’ (Fig. 1a).

### Off-target edits in sgRNA-dependent regions and sgRNA-independent regions

This protocol yields information on both types of off-target edits, sgRNA-dependent and sgRNA-independent. sgRNA-dependent edits can be monitored by various approaches and minimized by engineering Cas9 protein, optimizing sgRNA sequences, using transient delivery methods and so on<sup>16,17</sup>. We use the sgRNA-dependent off-target edit prediction software Cas-OFFinder<sup>56</sup> to predict sgRNA-dependent edits and compare the locations of SNVs identified by WGS with the locations predicted by Cas-OFFinder to distinguish between the two types of edits. We recommend that at least three sgRNA targets and several sgRNA-dependent off-target regions be used to identify sgRNA-dependent edits (Fig. 1a). Genome-wide sgRNA-independent edits could not previously be comprehensively monitored in vivo, but in this protocol, the numbers of SNVs and indels in the experimental groups and control group are compared to evaluate the sgRNA-independent edits.

### Preparation of rice calluses

The constructed binary vectors are introduced into *A. tumefaciens* strain AGL1 by electroporation. AGL1 cells not harboring a binary vector are used as a control. The experimental groups and control groups of AGL1 cells should be transformed into rice calluses at the same time.

The Japonica rice variety Zhonghua 11 is used for genetic transformation in our work, but other varieties with high-quality reference genomes are also good choices. *Agrobacterium*-mediated transformation of Zhonghua 11 callus cells is conducted as previously described with modifications<sup>57</sup> (Fig. 2), and hygromycin (50 µg/ml) is used to select transgenic plants. Calluses from mature rice seeds derived from a single plant are used. Rice is a self-fertile plant and offspring are considered to be genetically identical copies of their parents, so using the seeds from a single plant should generate a homogeneous plant population. Researchers using other species as research objects should carefully consider the homogeneity of the population under study.

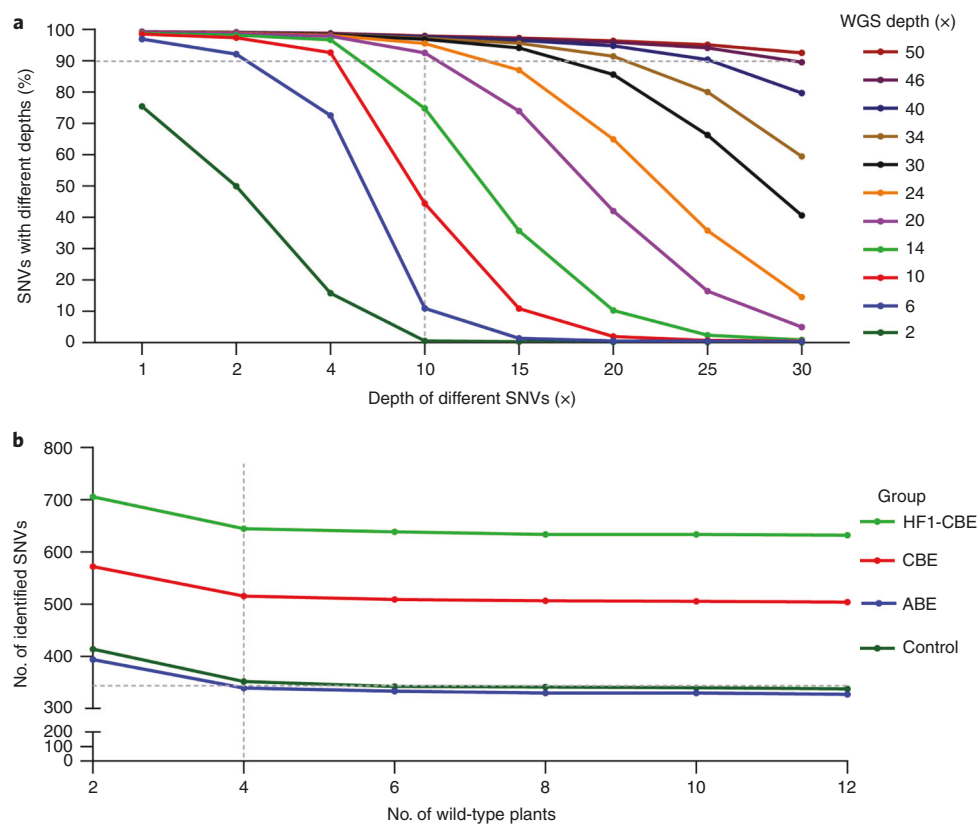
### Background mutations

Decreasing background mutations is the most important part of this method. We classify background mutations into three types:

The first are genetic differences between the tested rice plant population and the reference genome. Generally, a high level of genetic diversity between sequenced genomes and the reference genome makes SNV calling more difficult. We advise using the reference genome of the variety used to minimize genetic differences. To further decrease such genetic differences, the wild-type plants should also be sequenced using WGS. The variants called in wild-type plants are considered background mutations and should be removed in WGS analysis of sequenced control and experimental groups (Fig. 1b) (Step 69).

The second type are spontaneous mutations arising during DNA replication, and the third are mutations induced by tissue culture and transformation. The latter two types can be identified by testing the control group—which will generally harbor ~200–500 SNVs and ~50–150 indels<sup>19</sup>—and comparing the mutations found with those in the experimental groups

Transposon element mutant plants could be used to decrease mutations induced by tissue culture and transformation, for example, TOS17/19 null alleles, should decrease the transposon element ‘jumps’ and so reduce background mutations. However, Wei et al reported that researchers tend to subculture and amplify these transposon element null calli alleles for 3–6 months to obtain more calli,



**Fig. 4 | Saturation analysis of WGS depth and numbers of sequenced wild-type plants.** **a**, Percentages of identified SNVs with different depths versus the total identified SNVs at a given WGS depth (see also Supplementary Table 1). The x axis shows the different final depths of sequencing for SNVs, and the y axis shows the cumulative percentages of the total SNVs obtained at a given final depth of sequencing. The colors of the different lines indicate different depths of WGS. The vertical and the horizontal gray dotted lines show that when the WGS depth was 20 $\times$ , >90% of the total SNVs eventually identified were already detected at a sequencing depth of 10 $\times$ . The sample named CBE-sgRNA-#3 was used for this analysis. **b**, Numbers of SNVs identified by WGS in a given number of wild-type plants in CBE, ABE, HF1-CBE and Control groups. The x axis shows the different number of sequenced wild-type plants, and the y axis shows the number of identified SNVs by WGS with 50 $\times$  depth. The colors of the different lines indicate different groups. The vertical and the horizontal gray dotted lines show that when the numbers of wild-type plants are more than four, the numbers of identified SNVs do not change greatly with the increment of numbers of wild-type plants (see also Supplementary Table 2). This figure was generated by re-analyzing previously published data<sup>19</sup>.

and such prolonged growth increases the number of background mutations<sup>58</sup>. More attention should be paid to the callus induction period when these alleles are used.

### Control group

Tissue culture and the transformation process can induce SNVs and indels<sup>36</sup>. Hence, in this protocol, control plants that go through the transformation process, but without transferred DNA, are used to represent the numbers of background mutations induced during tissue culture and transformation (Fig. 1b).

### Depth of WGS and number of wild-type plants

In our work, WGS is performed at an average depth of 60 $\times$ . This high sequencing depth increases both variant-calling accuracy and cost. When researchers have conducted WGS on other species with large genomes, they have had to compromise between cost and accuracy. Therefore, we conducted a saturation analysis based on previously reported WGS data to determine the optimal sequencing depth (Fig. 4a and Supplementary Table 1)<sup>19</sup>. We varied the sequencing depth from 2 $\times$  to 50 $\times$  (Fig. 4a and Supplementary Table 1). More than 90% of the SNVs detected at 20 $\times$  sequencing depth were also detected at a sequencing depth of 10 $\times$  (Fig. 4a). Thus, we recommend a sequencing depth of  $\geq 20\times$  (Fig. 4a) in rice. The sequencing numbers of wild-type plants also influence the cost and experimental complexity of WGS, so we conducted a saturation analysis based on previously reported WGS data

to determine the optimal sequencing number for wild-type plants (Fig. 4b and Supplementary Table 2)<sup>19</sup>. We varied the wild-type plant numbers from 2 to 12, with 5 repeats, and we found  $\geq 4$  wild-type plants are required for genome-wide SNV detection in rice (Fig. 4b and Supplementary Table 2).

#### WGS data analyses

The raw reads are quality-controlled with Trimmomatic<sup>59</sup> and then mapped to the reference genome by BWA-MEM<sup>60</sup>. Using ensembles of sequence analysis tools could decrease the number of potential false positives induced by different software algorithms<sup>61</sup>. GATK, the standard variant caller, is commonly used for variant calling of WGS data on the basis of the Bayesian model<sup>62</sup>. LoFreq is a fast and ultrasensitive variant caller for calling somatic variants. It is possible to accurately call variants occurring in  $<0.05\%$  of a population<sup>63</sup>. Strelka2 is a fast and accurate small-variant caller optimized for analysis of germline and somatic variation<sup>64</sup>. To identify genome-wide de novo variants with high confidence, we call SNVs with all three of these software packages: GATK<sup>62</sup>, LoFreq<sup>63</sup> and Strelka2<sup>64</sup>, and the SNVs at the intersection of the three methods are considered true SNVs. Indels are called with two software packages—GATK<sup>62</sup> and Strelka2<sup>64</sup>—and the indels at the intersection of the two are considered true indels. We recommend using these three software methods to call variants in the rice genome; other ensembles may also work well but would need to be tested. Then the SNVs and indels identified in wild-type plants are filtered out to reduce the calculated genetic differences between the experimental population and the reference genome. At least 20 randomly selected variants should be Sanger-sequenced to evaluate the accuracy of WGS.

To assess sgRNA-dependent and sgRNA-independent edits, the SNVs and indels identified at on-target sites are filtered out. Then the locations of the predicted sgRNA-dependent off-target sites are organized by Cas-OFFinder and marked with Perl script and the SNVs and indels remaining are then classified as sgRNA-independent.

## Materials

### Biological materials

- Rice variety Zhonghua 11 (seeds are available from the authors upon request)
- DH5a *Escherichia coli* competent cells (Shanghai Weidi Biotechnology, cat. no. DL1001M)
- AGL1 *Agrobacterium* competent cells (Shanghai Weidi Biotechnology, cat. no. AE1020M)

### Plasmids and primers

- pH-CBE, the BE3 (third-generation cytosine base editor) plant expression vector, which includes an SpCas9-based sgRNA scaffold, a fusion of the rAPOBEC1 cytidine deaminase, SpCas9 nickase and UGI (Supplementary Note 1; ref. <sup>19</sup>). This plasmid can be obtained from the authors upon request.
- pH-HF1-CBE, the HF1-CBE plant expression vector, which includes the SpCas9-based sgRNA scaffold and fusion of rAPOBEC1 cytidine deaminase, HF1-SpCas9 nickase (high-fidelity SpCas9) and UGI (Supplementary Note 2; ref. <sup>19</sup>). This plasmid can be obtained from the authors upon request.
- pH-PABE-7 (refer to pH-PABE-7-sgRNA in Addgene), the adenine base editor (ABE) plant expression vector, which includes the SpCas9-based sgRNA scaffold and fusion of synthesized Tad and Tad7.10 adenine deaminase sequences and nCas9 (Supplementary Note 3; ref. <sup>10</sup>) (Addgene, cat. no. 115621)
- PCR primers (BGI, custom order; Table 1)

### Reagents

#### Production of vectors

- Agarose (Invitrogen, cat. no. 16500500)
- TAE buffer (10 $\times$ ; Cellgro, cat. no. 46-010-CM)
- PCR enzyme KOD FX (Toyobo, cat. no. KFX-101)
- dNTP (Toyobo, cat. no. NTP-201)
- FastPfu DNA polymerase (TransGen Biotech, cat. no. AP221-03)
- ClonExpress II One Step Cloning Kit (Vazyme Biotech, cat. no. C112-02)
- AxyPrep DNA Gel Extraction Kit (Axygen, cat. no. AP-GX-250)
- AxyPrep Plasmid Miniprep Kit (Axygen, cat. no. AP-MN-P-250)
- Ethanol (Sigma-Aldrich, cat. no. E7023)
- Rifampicin (Phytotech, cat. no. R501-10G)



**Table 1 | Vector construction and primers used for mutant screening**

Step	Name	Sequence (5'-3')	Purpose
2	APOBEC1-Gibson-F	TGTTACTTCTGCAGCCCTAGGATGCCGAAG AAGAAGCGCAAGGTGTCCAGCGAGACGGG	Forward primer to amplify the cereal codon-optimized rAPOBEC1 deaminase sequence
2	APOBEC1-Gibson-R	GCACCCAGAGGTGTGGGTACCTCGGCCAGT GAGCCTCGTTGCTTGGGGAGTAGTTGACG	Reverse primer to amplify the cereal codon-optimized rAPOBEC1 deaminase sequence
8	Ubi-F	TATGCAGCAGCTATATGTGGAT	Forward primer for colony PCR and sequencing
8	Cas9-R	ACCTTCGCCATCTCGTTGG	Reverse primer for colony PCR and sequencing
11	sgRNA-CBE-ACC-T1-F	GCGGAGCCATGGGAATGTAGACAA	Forward primer for ACC-T1 target site construction
11	sgRNA-CBE-ACC-T1-R	AAACTTGCTACATTCCCATGGCT	Reverse primer for ACC-T1 target site construction
11	sgRNA-CBE-ACC-T2-F	GGCGTCCACAGCTATCACACCCAC	Forward primer for ACC-T2 target site construction
11	sgRNA-CBE-ACC-T2-R	AAACGTGGGTGTGATAGCTGTGGA	Reverse primer for ACC-T2 target site construction
11	sgRNA-CBE/HF1-CBE-ACC-T3-F	GGCGTTCCTCGTGCTGGACAAGTG	Forward primer for ACC-T3 target site construction
11	sgRNA-CBE/HF1-CBE-ACC-T3-R	AAACCACTTGTCAGCAGCAGAGGAA	Reverse primer for ACC-T3 target site construction
11	sgRNA-CBE/HF1-CBE-ALS-T1-F	GGCGCAGGTCCCCCGCCGCATGAT	Forward primer for ALS-T1 target site construction
11	sgRNA-CBE/HF1-CBE-ALS-T1-R	AAACATCATGCGGCGGGGACCTG	Reverse primer for ALS-T1 target site construction
11	sgRNA-CBE-NRT1.1B-T1-F	GGCACGGCGACGGCGAGCAAGTGG	Forward primer for NRT1.1B-T1 target site construction
11	sgRNA-CBE-NRT1.1B-T1-R	AAACCCACTTGCTCGCCGTCGCCG	Reverse primer for NRT1.1B-T1 target site construction
11	sgRNA-CBE/HF1-CBE-Wxb-T1-F	GGCGTCCTCGGTACGACCAGTACA	Forward primer for Wxb-T1 target site construction
11	sgRNA-CBE/HF1-CBE-Wxb-T1-R	AAACTGTAAGTGGTCTACCGAGGA	Reverse primer for Wxb-T1 target site construction
11	sgRNA-ABE-ACC-T4-F	GGCGCATAGCACTCAATGCGGTCT	Forward primer for ACC-T4 target site construction
11	sgRNA-ABE-ACC-T4-2R	AAACAGACCGCATTGAGTGCTATG	Reverse primer for ACC-T4 target site construction
11	S2gRNA-ABE-ALS-T2-F	GGCGCCCAAGTGGGGCGCATTCA	Forward primer for ALS-T2 target site construction
11	sgRNA-ABE-ALS-T2-R	AAACTGAATGCGCCCCCACTTGGG	Reverse primer for ALS-T2 target site construction
11	sgRNA-ABE-CDC48-T1-F	GGCGTAGCACCCATGACAATGACA	Forward primer for CDC48-T1 target site construction
11	sgRNA-ABE-CDC48-T1-R	AAACTGTCATTGTCATGGGTGCTA	Reverse primer for CDC48-T1 target site construction
11	sgRNA-ABE-DEP1-T1-F	GGCGAGCACATGAGAGAACAATAT	Forward primer for DEP1-T1 target site construction
11	sgRNA-ABE-DEP1-T1-R	AAACATATTGTTCTCTCATGTGCT	Reverse primer for DEP1-T1 target site construction
11	sgRNA-ABE-NRT1.1B-T2-F	GGCGACTAGATATCTAAACCATTA	Forward primer for NRT1.1B-T2 target site construction
11	sgRNA-ABE-NRT1.1B-T2-R	AAACTAATGGTTTAGATATCTAGT	Reverse primer for NRT1.1B-T2 target site construction
12, 56	ACC-T1-F	TCCTCGTGCTGGACAAGTGTG	Forward primer to amplify the ACC-T1 target site in rice
12, 56	ACC-T1-R	AGCAAACCGTATCGCAATCTGAGT	Reverse primer to amplify the ACC-T1 target site in rice
12, 56	ACC-T2-F	GGCAACTAATGGTGTGTCCATCTTACTG	Forward primer to amplify the ACC-T2 target site in rice
12, 56	ACC-T2-R	CTGCAGCCATGGGAATGTAGAC	Reverse primer to amplify the ACC-T2 target site in rice

Table continued

**Table 1 (continued)**

Step	Name	Sequence (5'-3')	Purpose
12, 56	ACC-T3-F	CCTGCCTACATTGGTGGAC	Forward primer to amplify the ACC-T3 target site in rice
12, 56	ACC-T3-R	AGGTTCTCAACAATAGTCGAGCC	Reverse primer to amplify the ACC-T3 target site in rice
12, 56	ALS-T1- F	CGTGGCTGTGTGCATCCTATCCCAC	Forward primer to amplify the ALS-T1 target site in rice
12, 56	ALS-T1-R	CACGGCCATCTGCTGCTGGATGTCC	Reverse primer to amplify the ALS-T1 target site in rice
12, 56	NRT1.1B-T1-F	GATGTCACCTGATGATCTGAAGTAGC	Forward primer to amplify the NRT1.1B-T1 target site in rice
12, 56	NRT1.1B-T1-R	ATGATGGTGGTCGCCAGAT	Reverse primer to amplify the NRT1.1B-T1 target site in rice
12, 56	Wxb-T1-F	CACAACTTCGATCGCTCGTC	Forward primer to amplify the Wxb-T1 target site in rice
12, 56	Wxb-T1-R	CGTTGCAGACGAACACAACA	Reverse primer to amplify the Wxb-T1 target site in rice
12, 56	ACC-T4-F	CATGATGCCAACTATCCCTGCTGAC	Forward primer to amplify the ACC-T4 target site in rice
12, 56	ACC-T4-R	GAAAGATCGTGATTCTCCAGTCCA	Reverse primer to amplify the ACC-T4 target site in rice
12, 56	ALS-T2- F	CTGTCTTCGGCTGGTCTGGG	Forward primer to amplify the ALS-T2 target site in rice
12, 56	ALS-T2-R	TGCCAAGCACATCAAACAAGTAAA	Reverse primer to amplify the ALS-T2 target site in rice
12, 56	CDC48-T1- F	TCTTTCTGATTAATGGCCCGAG	Forward primer to amplify the CDC48-T1 target site in rice
12, 56	CDC48-T1 -R	GGATGCATTGGAGAGCAGCC	Reverse primer to amplify the CDC48-T1 target site in rice
12, 56	DEP1-T1- F	TCAGCCTGCAGTACTGAATTATC	Forward primer to amplify the DEP1-T1 target site in rice
12, 56	DEP1-T1 -R	GGGCCTAAGTGTGACATACAAG	Reverse primer to amplify the DEP1-T1 target site in rice
12, 56	NRT1.1B-T2-F	AGCTAGGAGTAGAGAACGAGACATATAC	Forward primer to amplify the NRT1.1B-T2 target site in rice
12, 56	NRT1.1B-T2-R	GTTGGGAGAATAGCTGAAGCTATCGG	Reverse primer to amplify the NRT1.1B-T2 target site in rice
56	CBE-ALS-T1-#1-OT-F	CCGCCAAGCAAACACGCGTA	Forward primer to amplify an off target site with 1 mismatch with CBE-ALS-T1-#1 in rice
56	CBE-ALS-T1-#1-OT-R	CGACAGGCGTGGCACCTCTTA	Reverse primer to amplify an off target site with 1 mismatch with CBE-ALS-T1-#1 in rice
56	CBE-ALS-T1-#3-OT-F	TCATGCGAAGACGCGTACCT	Forward primer to amplify an off-target site with 2 mismatches with CBE-ALS-T1-#3 in rice
56	CBE-ALS-T1-#3-OT-R	CAGCAGGCGTGTGACTTCT	Reverse primer to amplify an off-target site with 2 mismatches with CBE-ALS-T1-#3 in rice
56	CBE-NRT1.1B-T1-#3-OT-F	CGCACTGGAACACTCGTGG	Forward primer to amplify an off-target site with 4 mismatches with CBE-NRT1.1B-T1-#3 in rice
56	CBE-NRT1.1B-T1-#3-OT-R	CTCAAACGCCGAGCTCCTTCT	Reverse primer to amplify an off-target site with 4 mismatches with CBE-NRT1.1B-T1-#3 in rice
81	Variants-test-F	GTTCCGGTATCAGGGGTGAAA	Example forward primer for indel and SNV validation in rice
81	Variants-test-R	CACTGTTGGCATTCCAGCTC	Example reverse primer for indel and SNV validation in rice

- Kanamycin (TransGen, cat. no. GG201-01)
- Tryptone (Oxoid, cat. no. LP0042)
- NaCl (IGDB, cat. no. YC-SJ02027)
- Yeast extract (Oxoid, cat. no. LP0021)
- Commercially synthesized deaminase and linker sequences (Genewiz, custom order)

**A. *tumefaciens*-mediated rice transformation**

- Phytagel (Sigma-Aldrich, cat. no. P8169-5Kg)
- Sucrose (Sigma-Aldrich, cat. no. V900116)
- NB basal medium (Duchefa, cat. No. DU1158.0050)
- 2,4-Dichlorophenoxyacetic acid (2,4-D; Sigma-Aldrich, cat. no. D7299)
- $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  (Sigma-Aldrich, cat. no. C3036)
- KOH (Sigma-Aldrich, cat. no. 484016)
- Naphthaleneacetic acid (NAA; Sigma-Aldrich, cat. no. N0640)
- Murashige and Skoog (MS) salts, including vitamins (Phytotech, cat. no. M519)
- Casein hydrolysate (Duchefa, cat. no. C1301.0500)
- Inositol (Sangon, cat. no. A600536-0500)
- Proline (Sangon, cat. no. A600923-0500)
- L-Glutamine (Sigma, cat. no. G3126-100g)
- Hygromycin (Phytotech, cat. no. H370-20 ML)
- NaOH (Sigma-Aldrich, cat. no. 306576)
- $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$  (Santa Cruz, cat. no. sc-234269)
- $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  (Sigma, cat. no. RES0089M-A702X)
- $(\text{NH}_4)_2\text{SO}_4$  (Sigma-Aldrich, cat. no. 1012165000)
- $\text{K}_2\text{SO}_4$  (Sigma, cat. no. RES0797P-A702X)
- $\text{KH}_2\text{PO}_4$  (Biorbyt, cat. no. orb65575)
- Boric acid (Sigma-Aldrich, cat. no. B0252)
- $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$  (Sigma, cat. no. 63535)
- $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}$  (JSENB, cat. no. UE277908)
- $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$  (Sigma-Aldrich, cat. no. 1088835000)
- $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$  (Merck Millipore, cat. no. 1.08454.1000)
- $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  (Sigma, cat. no. F8263)
- KCl (Mreda, cat. no. M026379)
- $\text{NaH}_2\text{PO}_4$  (Sigma-Aldrich, cat. no. 1063700050)
- $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  (Dogesce, cat. no. 3148)
- $\text{MnSO}_4 \cdot \text{H}_2\text{O}$  (Sigma-Aldrich, cat. no. 1059990500)
- KI (Targetmol, cat. no. T0357)
- $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  (Sigma-Aldrich, cat. no. 1065210250)
- $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  (Sigma-Aldrich, cat. no. 1025390100)
- L-aspartic acid (MedChemExpress, cat. no. HY-N0666)
- L-arginine (MedChemExpress, cat. no. HY-N0455)
- L-glycine (Solarbio, cat. no. IG0720)
- Acetosyringone (MedChemExpress, cat. no. HY-W009884)
- Glucose (Macklin, cat. no. D823520)
- NaClO (Lablead, cat. no. clsn01)
- Mannitol (Sigma, cat. no. M1902-5KG)
- N6-benzyladenine (Cayman, cat. no. 21711)
- Kinetin (Sigma, cat. no. K3378-5G)
- Ethyl alcohol (Aladdin, cat. no. E111993-4x4L)
- Carbenicillin (Yeasen Biotech, cat. no. 60202ES60)

**Mutant screening and identification**

- DNAquick Plant System (Tiangen, cat. no. DP321-03)
- 2x Rapid Taq Master Mix (Vazyme Biotech cat. no. P222-03)

**Equipment**

- PCR thermocycler (Veriti 96-well thermal cycler; Applied Biosystems, cat. no. 9902)
- Heating water bath (Memmert, cat. no. WNB22L4L5)
- UV-visible light spectrophotometer (Thermo, model no. NanoDrop ND-2000)
- Centrifuge series (Eppendorf, model nos. 5810R, 5424 and 5417R)
- High-speed centrifuge (Beckman Coulter, model no. Avanti JXN-30)
- Incubator shaker (Eppendorf, model no. Innova42)
- Rice husker (<https://www.chinaricemill.com/MLGT-Rice-Husker-pd80899.html>, MLGT25)

- MicroPulser (Bio-Rad, model no. 1652100)
- Gel electrophoresis apparatus (Bio-Rad, model no. 1704489EDU or equivalent)
- Digital gel imaging system (BioDoc-It; UVP, cat. no. 97-0256-02)
- Fluorescence microscope (Olympus, model no. SZX16)
- High-throughput tissue homogenizer (SPEX SamplePrep, model no. 2010 Geno/Grinder)
- Centrifuge tubes (50 ml)
- Petri dishes (Greiner Bio-One, cat. no. 633180)
- Standard microcentrifuge tubes (1.5 ml; Eppendorf, cat. no. 0030 125.150)
- Single-edge razor blades (Feiyang, cat. no. 9875814)
- Nylon mesh (40- $\mu$ m ; BD Falcon, cat. no. F613461)
- Syringe sterilization filter (0.22- $\mu$ m; Sartorius Stedim Biotech, cat. no. 21423103)
- Eppendorf tubes (1.5 and 2.0 ml; Eppendorf)
- Graduated cylinder
- Sequencing system (Illumina, model no. NovaSeq 6000)

#### Operating systems

- Linux or Mac OS

#### Software

- targetDesign (<http://skl.scau.edu.cn/targetdesign/>)
- Cas-OFFinder (<http://www.rgenome.net/cas-offinder/>)
- Perl v.5.10 or higher (<https://www.perl.org/get.html>)
- BWA v.0.7.17 (<https://github.com/lh3/bwa>)
- BMAP (<https://www.geneious.com/plugins/bbm/>)
- Samtools v.1.7 (<http://www.htslib.org/>)
- GATK v.3.4 (<https://gatk.broadinstitute.org/hc/en-us>)
- LoFreq (<http://csb5.github.io/lofreq/>)
- Strelka2 v.2.9.10 (<https://github.com/Illumina/strelka>)
- IGV (<http://www.igv.org/>)
- Gel quantification software (ImageJ; NIH: <http://rsbweb.nih.gov/ij/>)
- Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>)

#### Example datasets

- All the data shown in this protocol are deposited in NCBI BioProject under accession code [PRJNA522656](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA522656). Because the dataset is too big (hundreds of gigabytes in total), we generated a small dataset to demo the pipeline and deposited it in GitHub at <https://github.com/ReiGao/GWSBE/tree/master/ExampleData>.

#### Reagent setup

##### 1 M KOH (100 ml)

Add 5.6 g KOH to 90 ml sterile ddH<sub>2</sub>O, and make up to 100 ml. Store at room temperature (20–25 °C) for up to 6 months.

##### 1M NaOH (100 ml)

Add 4 g NaOH to 90 ml sterile ddH<sub>2</sub>O, and make up to 100 ml. Store up to 6 months at room temperature. **! CAUTION** NaOH is corrosive! Gloves are definitely needed when setting up and pipetting. If the NaOH touches your skin, rinse it off quickly and thoroughly with cool water.

##### LB medium (1 liter)

Add 10 g tryptone, 5 g yeast extract and 10 g NaCl to 1 liter ddH<sub>2</sub>O and autoclave for 20 min at 121 °C. Store at 4 °C for up to 1 month.

##### LB agar plates (1 liter)

Add 14 g agar to 1 liter of LB medium and autoclave for 20 min at 121 °C. Adjust volume to 1 liter. After sterilization and cooling to 50 °C, add antibiotics such as kanamycin and rifampicin from sterile stock solutions to final concentrations of 50 mg/liter and 25 mg/liter, respectively. Pour ~35 ml of LB agar per sterile Petri dish. Store at 4 °C for up to 1 month.

**K1 stock solution (1 liter, 1,000×)**

Add 18.5 g KNO<sub>3</sub> and 86.2 g Ca(NO<sub>3</sub>)<sub>2</sub> · 4H<sub>2</sub>O to 900 ml sterile ddH<sub>2</sub>O, and make up to 1 liter. Store at 4 °C for up to 3 months.

**K2 stock solution (1 liter, 1,000×)**

Add 135.0 g MgSO<sub>4</sub> · 7H<sub>2</sub>O and 48.2 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> to 900 ml sterile ddH<sub>2</sub>O and make up to 1 liter. Store at 4 °C for up to 3 months.

**K3 stock solution (1 liter, 1,000×)**

Add 15.9 g K<sub>2</sub>SO<sub>4</sub> to 900 ml sterile ddH<sub>2</sub>O and make up to 1 liter. Store at 4 °C for up to 3 months.

**K4 stock solution (1 liter, 1,000×)**

Add 24.8 g KH<sub>2</sub>PO<sub>4</sub> to 900 ml sterile ddH<sub>2</sub>O and make up to 1 liter. Store at 4 °C for up to 3 months.

**K5 stock solution (1 liter, 1,000×)**

Add 2.86 g H<sub>3</sub>BO<sub>3</sub>, 1.81 g MnCl<sub>2</sub> · 4H<sub>2</sub>O, 0.098 g (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub> · 4H<sub>2</sub>O and 0.22 g ZnSO<sub>4</sub> · 7H<sub>2</sub>O to 900 ml sterile ddH<sub>2</sub>O, and make up to 1 liter. Store at 4 °C for up to 3 months.

**K6 stock solution (1 liter, 1,000×)**

Add 0.08 g CuSO<sub>4</sub> · 5H<sub>2</sub>O to 900 ml sterile ddH<sub>2</sub>O and make up to 1 liter. Store at 4 °C in the dark for up to 3 months.

**Fe-EDTA stock solution (1 liter, 1,000×)**

Add 7.45 g Na<sub>2</sub>EDTA to 450 ml sterile distilled H<sub>2</sub>O with constant stirring in a beaker; add 5.57 g FeSO<sub>4</sub> · 7H<sub>2</sub>O to 450 ml sterile distilled H<sub>2</sub>O with constant stirring in another beaker. Bring the FeSO<sub>4</sub> · 7H<sub>2</sub>O solution to a boil, pour the Na<sub>2</sub>EDTA solution into the FeSO<sub>4</sub> · 7H<sub>2</sub>O solution slowly, with constant stirring, and make up to 1 liter. Store for up to 6 months in the dark at –20 °C. Store at 4 °C in the dark for up to 3 months.

**Modified Kimura B solution (1 liter)**

Add 1 ml each of the K1, K2, K3, K4, K5, K6, and Fe-EDTA stock solutions to 900 ml sterile ddH<sub>2</sub>O and make up to 1 liter. Adjust pH to 6.0 with 1 M KOH. Store at 4 °C in the dark for up to 3 months.

**! CAUTION** KOH is corrosive! Gloves are definitely needed when setting up and pipetting. If the KOH touches your skin, rinse it off quickly and thoroughly with cool water.

**NAA stock solution (100 ml)**

Add 50 mg NAA (0.5 mg/ml final) to some absolute ethyl alcohol (1–2 ml) to dissolve; add sterile distilled H<sub>2</sub>O to 100 ml and then filter-sterilize with a 0.22-µm filter. Store up to 6 months at –20 °C.

**AA major salts stock solution (1 liter, 10×)**

Add 29.5 g KCl, 2.5 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 1.7 g NaH<sub>2</sub>PO<sub>4</sub> and 1.5 g CaCl<sub>2</sub>·2H<sub>2</sub>O to 900 ml sterile distilled H<sub>2</sub>O and make up to 1 liter; filter-sterilize using a 0.22-µm filter. Store up to 3 months at 4 °C.

**AA minor salts stock solution (1 liter, 1,000×)**

Add 3 g H<sub>3</sub>BO<sub>3</sub>, 10 g MnSO<sub>4</sub>·H<sub>2</sub>O, 2 g ZnSO<sub>4</sub>·7H<sub>2</sub>O, 0.75 g KI, 0.25 g Na<sub>2</sub>MoO<sub>4</sub>, 0.025 g CuSO<sub>4</sub>·5H<sub>2</sub>O and 0.025 g CoCl<sub>2</sub>·6H<sub>2</sub>O to 900 ml sterile distilled H<sub>2</sub>O and make up to 1 liter; then filter-sterilize using a 0.22 µm filter. Store up to 3 months at 4 °C.

**AA vitamins stock solution (1 liter, 10×)**

Add 8.76 g L-glutamine, 2.66 g L-aspartic acid, 1.74 g L-arginine and 0.075 g L-glycine to 900 ml sterile distilled H<sub>2</sub>O and make up to 1 liter; then filter-sterilize using a 0.22-µm filter. Store up to 3 months at 4 °C.

**AS200 stock solution (25 ml, 10×)**

Add 1 g acetosyringone (200 mM final concentration) to 25.45 ml absolute ethyl alcohol and filter-sterilize using a 0.22-µm filter. Store up to 6 months at –20 °C

**Liquid AAM medium (1 liter)**

Add 100 ml AA major salts stock solution (Reagent setup), 1 ml AA minor salts stock solution (Reagent setup), 100 ml AA vitamins stock solution (Reagent setup), 4.4 g MS salts (including vitamins), 0.5 g casein hydrolysate, 0.1 g inositol, 68.5 g sucrose, and 36 g glucose to 900 ml distilled H<sub>2</sub>O and make up to 1 liter. Adjust pH to 5.2 with 1 M NaOH (Reagent setup), autoclave for 20 min at 121 °C, and add 500 µl AS200 stock solution after cooling to 50 °C. Store up to 3 months at 4 °C.

**Carb stock solution (50 ml, 1,000×)**

Add 10 g carbenicillin (200 mg/ml final conc.) to 40 ml sterile distilled H<sub>2</sub>O and make up to 50 ml; then filter-sterilize using a 0.22-µm filter. Store up to 6 months at -20 °C.

**Kinetin stock solution (100 ml)**

Add 100 mg kinetin (1 mg/ml final) to 40 ml sterile distilled H<sub>2</sub>O and add some 1 M KOH to dissolve; make up to 100 ml and then filter-sterilize using a 0.22-µm filter. Store up to 6 months at -20 °C.

**Solid M1 medium (1 liter)**

Add 4.1 g NB basal medium, 0.5 g casein hydrolysate, 0.1 g inositol, 2.8 g proline, 0.5 g L-glutamine, 30 g sucrose, 2 mg 2,4-D and 3.75 g Phytigel to 900 ml sterile distilled H<sub>2</sub>O and make up to 1 liter; then use 1 M NaOH to adjust the pH to 5.8 and autoclave for 20 min at 121 °C. After cooling to 50 °C, pour ~30-ml aliquots into Petri dishes. Store up to 6 weeks at 4 °C in dark containers.

**Solid M2 medium (1 liter)**

Add 4.1 g NB basal medium, 0.5 g casein hydrolysate, 0.1 g inositol, 2.8 g proline, 0.5 g L-glutamine, 30 g sucrose, 2 mg 2,4-D, and 3.75 g Phytigel to 900 ml sterile distilled H<sub>2</sub>O and make up to 1 liter; then use 1 M NaOH to adjust pH to 5.8 and autoclave for 20 min at 121 °C. After cooling to 50 °C, add 50 mg hygromycin and 1 ml Carb stock solution and pour ~35 ml into Petri dishes. Store up to 6 weeks at 4 °C in dark containers. **▲ CRITICAL** Add antibiotics after cooling to 50 °C; otherwise, the antibiotics will be inactivated by high temperature.

**Solid M3 medium (1 liter)**

Add 4.4 g MS salts (including vitamins), 0.5 g casein hydrolysate, 0.1 g inositol, 30 g sucrose, 20 g mannitol, 2 mg N6-benzyladenine, 200 µl Kinetin stock solution and 4.6 g Phytigel to 900 ml sterile distilled H<sub>2</sub>O and make up to 1 liter. Adjust pH to 5.8 with 1 N NaOH and autoclave for 20 min at 121 °C. After cooling to 50 °C, add 50 mg hygromycin, 1 ml Carb stock solution and 400 µl NAA stock solution and pour ~35 ml per Petri dish. Store up to 6 weeks at 4 °C in dark containers.

**Solid M4 medium (1 liter)**

Add 2.2 g MS salts (including vitamins), 0.1 g inositol, 30 g sucrose and 3.0 g Phytigel to 900 ml sterile distilled H<sub>2</sub>O and make up to 1 liter. Adjust pH to 5.8 with 1 N NaOH and autoclave for 20 min at 121 °C. After cooling to 50 °C, add 100 µl NAA stock solution and 50 mg hygromycin and pour ~35 ml per Petri dish. Store at 4 °C in dark containers up to 6 weeks.

**Procedure****Production of vectors ● Timing 10 d**

**▲ CRITICAL** Steps 1–8 describe how to construct your base-editing vector of choice, using pH-CBE (APOBEC1-BE3) as an example. Alternatively, the ready-made vectors pH-CBE, pH-HF1-CBE and pH-PABE-7 are listed in the Materials section and can be purchased ready for use, in which case you can skip to Step 9.

- 1 To construct different base-editing vectors, commercially synthesize a cereal codon-optimized deaminase and linker sequences with 5' AvrII and 3' KpnI enzyme sites (Genewiz).
- 2 Taking pH-CBE (APOBEC1-BE3) as an example, PCR-amplify the synthesized sequences from Step 1 as follows:

Component	Volume ( $\mu\text{l}$ )	Final Conc.
PCR enzyme KOD FX (1.0 U/ $\mu\text{l}$ )	1	1 U/50 $\mu\text{l}$
2 $\times$ PCR buffer for KOD FX	25	1 $\times$
dNTPs (2 mM)	10	0.4 mM
APOBEC1-Gibson-F (10 pmol/ $\mu\text{l}$ ; Table 1)	1.5	0.3 $\mu\text{M}$
APOBEC1-Gibson-R (10 pmol/ $\mu\text{l}$ ; Table 1)	1.5	0.3 $\mu\text{M}$
Synthesized sequences from Step 1	>1	200 ng
ddH <sub>2</sub> O	$\leq$ 50	

**▲ CRITICAL STEP** We recommend using a high-fidelity PCR polymerase such as KOD FX or FastPfu DNA polymerase to decrease the PCR error rate.

- Perform PCR using the following conditions and run all the PCR products (50  $\mu\text{l}$  of each) on a 1.2% (wt/vol) agarose gel in TAE buffer at 150 V for  $\sim$ 10 min; single bright bands are expected.

Cycle no.	Denaturation	Annealing	Extension
1	95 $^{\circ}\text{C}$ , 3 min		
2-32	95 $^{\circ}\text{C}$ , 15 s	$T_m$ $^{\circ}\text{C}$ , 15 s	72 $^{\circ}\text{C}$ , 1 min
33			72 $^{\circ}\text{C}$ , 5 min

$T_m$ , melting temperature.

- Linearize the pH-CBE vector with AvrII and KpnI at 37  $^{\circ}\text{C}$  for 30 min and run the enzyme-digested products (20  $\mu\text{l}$  of each) on a 1.2% (wt/vol) agarose gel in TAE buffer at 150 V for  $\sim$ 10 min. Two bands, 18,000 and 750 bp, are expected.
- Purify the 750-bp product from Step 3 and the 18,000-bp product from Step 4 with an AxyPrep DNA Gel Extraction Kit according to the manufacturer's instructions. The concentration of the gel-extracted products should be  $\geq$ 30 ng/ $\mu\text{l}$  as measured by NanoDrop spectrophotometer.
- Clone the purified 750-bp product from Step 5 into the linearized pH-CBE vector (the 18,000-bp product purified in Step 5) with the ClonExpress II One Step Cloning Kit, using the following reaction mix, and incubate at 37  $^{\circ}\text{C}$  for 30 min.

Component	Volume ( $\mu\text{l}$ )	Final conc.
Exnase II	2	
CE Buffer (5 $\times$ )	4	1 $\times$
750-bp fragment from Step 3, purified in Step 5	Variable	150-300 ng
Digested vector (18,000-bp product from Step 4, purified in Step 5)	Variable	30 ng
ddH <sub>2</sub> O	To 20	

- Transfer 10  $\mu\text{l}$  of the product from Step 6 to 100  $\mu\text{l}$  DH5a *E. coli* competent cells as described in the manufacturer's instructions.
- Prepare colony PCR reactions for six to eight clones, using the following reaction mix with primers Ubi-F and Cas9-R (Table 1):

Component	Volume ( $\mu\text{l}$ )	Final conc.
Rapid Taq Master Mix (2 $\times$ )	25	1 $\times$
Ubi-F (10 pmol/ $\mu\text{l}$ ; Table 1)	1	0.3 $\mu\text{M}$
Cas9-R (10 pmol/ $\mu\text{l}$ ; Table 1)	1	0.3 $\mu\text{M}$
Single colony of bacteria	—	—
ddH <sub>2</sub> O	21	—

Perform PCR using the following conditions and run all the PCR products (50  $\mu\text{l}$  of each) on a 1.2% (wt/vol) agarose gel in TAE buffer at 150 V for  $\sim$ 10 min; single bright bands are expected.

Cycle no.	Denaturation	Annealing	Extension
1	95 °C, 3 min		
2-32	95 °C, 3 min	$T_m$ °C, 15 s	72 °C, 30 s
33			72 °C, 5 min

$T_m$ , melting temperature.

- Extract the plasmid DNA using a Plasmid Miniprep Kit, following the manufacturer’s instructions, and perform Sanger sequencing to identify the positive clones. These vectors are referred to as ‘BE without sgRNA’ (Fig. 1a).

**■ PAUSE POINT** The plasmid DNA can be stored at –20 °C for several months.

- For each base editor vector, design 3–5 sgRNA sequences for on-target edits using the targetDesign software as described in the guidelines (<http://skl.scau.edu.cn/help/>) and predict sgRNA-dependent off-target sites for each sgRNA using Cas-OFFinder and allowing up to 5-nt mismatches as in the following format:

```
C:\cas-offtarget\ZH11_genome_chr.cor.fasta
NNNNNNNNNNNNNNNNNNNNNNNGG
CAGGTCCCCGCCGCATGAT 5
```

In this example input file, the first line represents the location of the reference genome file, the second line represents the crRNA and PAM pattern. The third line represents the on-target sequence and maximal number of tolerated mismatches between on-target and off-target sequences, separated by a space. The example target sequence shown here is that of CBE-ALS-T1, an sgRNA sequence targeting CBE (cytosine base editor) to the *OsALS* gene. At least three sgRNA targets should be selected. For exploring sgRNA-dependent off-target effects more comprehensively, we advise that at least one sgRNA harbor off-target sites with one or two mismatches.

**▲ CRITICAL STEP** The portable version of Cas-OFFinder supports the user-provided reference genome sequence, and the instructions for downloading and running the portable version can be found at <http://www.rgenome.net/cas-offinder/portable>.

- Commercially synthesize the primers for cloning the sgRNAs designed in Step 10 with 5'-GGCG and 5'-AAAC adaptors added to the forward and reverse primers, respectively (Table 1), enabling them to be ligated into the constructed base editor vectors from Step 9 using T4 ligation. These procedures can be carried out according our previously published protocol<sup>65,66</sup>. Then extract the plasmid DNA using a Plasmid Miniprep Kit, following the manufacturer’s instructions. These vectors are referred to as ‘BE with sgRNA’ (Fig. 1a).

**■ PAUSE POINT** The plasmid DNA can be stored at –20 °C for several months.

- (Optional) Validate the sgRNA activities using the rice protoplast transient transformation assay, following our previously published protocol<sup>65,66</sup>.

**▲ CRITICAL STEP** Optionally, using transient protoplast transformation and amplicon deep sequencing assays one could evaluate the sgRNA-dependent and sgRNA-independent mutation efficiencies; the detailed protocols are given in our previous publications<sup>65,66</sup>.

**Transformation of *Agrobacterium* by electroporation ● Timing 3 d**

- Place the electroporator’s 0.1-cm electric shock cup on a fresh filter and open the lid for 10 min to drain the water and evaporate the ethanol.
- Place the cup on ice after draining the water and evaporating the ethanol.
- Thaw the *Agrobacterium* competent cells, which should be stored in a –80 °C refrigerator on ice.
- Add the 500-ng plasmid from Step 9 to 100 µl of *Agrobacterium* competent cells and add the 500-ng plasmid from Step 11 to 100 µl of *Agrobacterium* competent cells. Mix gently by tapping 4–5 times; add 5 µl sterile water to 100 µl of the competent cells as a control.
- Pipette the cells with plasmid added from Step 16 into chilled electric shock cups and tap the cups to bring the liquid to the bottom of the cuvettes.

**▲ CRITICAL STEP** Do not introduce bubbles because arcing may occur due to the high concentration of salts or air bubbles.



- 18 Start the MicroPulser and select the pre-programmed setting for *A. tumefaciens* (voltage = 2.2 kV, number of pulses = 1, and no time constant).
- 19 Place the electric shock cup into the slide of the shocking chamber and press the yellow 'Pulse' button.
- 20 Withdraw the slide from the chamber and quickly insert it into ice.
- 21 Add 700  $\mu$ l LB medium without antibiotics to the electric shock cup and transfer the mixture to an empty 2.0-ml Eppendorf tube.
- 22 Culture the cells at 28 °C for 2–3 h with shaking at 200 r.p.m.
- 23 Spread 200  $\mu$ l cells on an LB agar plate containing 25  $\mu$ g/ml rifampicin and 50  $\mu$ g/ml kanamycin; spread the control group cells on an LB plate containing only 15  $\mu$ g/ml rifampicin.
- 24 Incubate the plates at 28 °C for 2 d until bacterial colonies become visible.

### Harvesting rice seeds and growing wild-type plants ● Timing 15 d

- 25 Grow rice variety Zhonghua 11 in a field under normal growing conditions<sup>67</sup>. Harvest the mature seeds manually from individual plants for growing calluses.  
**! CAUTION** Although rice is a primarily self-pollinated species, rice outcrossing can occur with low frequency when rice is grown in open field. Covering the flower heads with bags can prevent natural outcrossings.  
**▲ CRITICAL STEP** About 1,500–2,500 seeds would be harvested from one rice plant. The seeds from each individual plant should be kept separately. At least 100 seeds are needed for one vector's transformation. Each rice plant is enough for transformation of 15–25 vectors.  
**▲ CRITICAL STEP** The calluses used in each experiment should be derived from single plants to decrease the heterogeneity of the regenerated plants.
- 26 Dehusk the mature seeds with a rice husker and wash them in 20 ml 70% (vol/vol) ethanol for 1 min with shaking and then in 25 ml 3% (wt/vol) sodium hypochlorite for 30 min with shaking, then wash the seeds using sterile water for 5 times with shaking.
- 27 Put the seeds into a conical flask with 50 ml sterile water and put the conical flask into an incubator chamber at 37 °C for 2 d, changing the water every 12 h.
- 28 After 2 d, the seeds will germinate.
- 29 Place the germinated seeds into 96-well plates. Use scissors to cut off the bottoms of the wells in order to let the roots steep in water and cultivate in modified Kimura B solution (Supplementary Fig. 1) in the incubator chamber at 28 °C for 2 weeks, changing the medium every day (Fig. 1b).
- 30 Transfer healthy green seedlings to 18-cm-long glass containers with ~60 ml M4 medium without hygromycin for further development at 28 °C in the light, until the mutants grow to ~18 cm.
- 31 Extract the genomic DNA using a DNAquick Plant System, following the user manual instructions; add 60  $\mu$ l ddH<sub>2</sub>O or TE buffer (from the DNAquick Plant System) for elution.
- 32 Perform WGS of 8–12 individual plants at an average coverage of 50 $\times$  with a NovaSeq 6000.

### Growing calluses using mature seeds ● Timing ~1 month

- 33 Dehusk the mature seeds from Step 24 manually or with a rice husker; collect 100 mature seeds that are clean and free of plaque into a 50-ml tube.
- 34 Wash the seeds in 20 ml 70% (vol/vol) ethanol for 1 min with shaking and then in 25 ml 3% (wt/vol) sodium hypochlorite for 30 min with shaking.
- 35 Wash the seeds in 30 ml sterile water 4 or 5 times.
- 36 Dry the seeds on sterile filter paper in a fume hood.
- 37 Cultivate 15–20 of the dry seeds per plate with 30 ml solid M1 medium at 28 °C in the dark in a growth incubator for 10 d (Fig. 2a).  
**▲ CRITICAL STEP** At least 100 seeds are needed for one vector's transformation; therefore, 5–7 plates are needed for each vector.
- 38 Developing calluses should be visible on the mature seeds 10 d later (Supplementary Fig. 2); one callus develops per seed; use forceps to transfer the yellow and compact callus from the seed to a new plate with 30 ml solid M1 medium (Fig. 2b,c), 15–20 calluses per plate. Culture calluses in the dark in a growth incubator for 2 weeks.  
**▲ CRITICAL STEP** After 2 weeks, the calluses can be used for transformation or subcultured by using forceps to place them into a new plate with 30 ml solid M1 medium to maintain the activity of the calluses.

### ? TROUBLESHOOTING

- 39 Four or five days before transformation (at the end of the 2 weeks of culture in the previous step), transfer the yellow and compact calluses to a new plate with 30 ml solid M1 medium, 10–15 calluses per plate.  
**▲ CRITICAL STEP** At least 4 d are required for calluses to adapt to the new growth environment.

**Infection and co-cultivation of calluses and *A. tumefaciens* ● Timing 3–6 d**

- 40 Pour 2 ml liquid AAM medium into the *Agrobacterium* incubation plates from Step 23 and transfer the *Agrobacterium*-containing liquid AAM medium to a 50-ml centrifuge tube; then make volume up to 10 ml with liquid AAM medium.  
 41 Incubate the *Agrobacterium* in liquid AAM medium at 28 °C for 1–3 d, until it reaches an absorbance at 600 nm ( $A_{600}$ ) value of 1.0–2.0 as measured by a UV–visible light spectrophotometer. Centrifuge for 5 min at 13,400g and 25 °C, discard the supernatant and resuspend the pellet in ~30 ml of liquid AAM medium ( $A_{600} = 0.1$ ).  
 42 Select 80–100 colonies of healthy calluses from Step 39 and place them into a new empty plate on a clean bench.  
 43 Add 30 ml of *Agrobacterium* suspension from Step 41 to the plate with ~100 callus colonies from Step 42.  
 44 Incubate the calluses and *A. tumefaciens* suspension for 30 min on a clean bench (Fig. 2d).  
 45 Use a pipette to remove the *A. tumefaciens* suspension.  
 46 Dry the calluses on sterile filter paper (Fig. 2e).  
 47 Transfer all the calluses to one Petri dish with filter paper; incubate in the dark for 3 d at 25 °C.

**Tissue culture ● Timing 6 weeks**

- 48 After 3 d, use forceps to carefully transfer the calluses to a plastic Petri dish containing 35 ml solid M2 medium. About 15–20 calluses can be placed in one plate.  
**▲ CRITICAL STEP** About 80% of hygromycin-resistant callus colonies can be obtained from embryogenic calluses. After regeneration, ~50% of the resistant colonies will generate seedlings. Considering the different target mutation efficiencies of base editors, ≥50 embryogenic callus colonies are required for each vector to guarantee that 10 individual mutants will be obtained for WGS.  
**? TROUBLESHOOTING**  
 49 Incubate the plates in a dark growth incubator at 28 °C for 2 weeks for each selection cycle; transfer healthy, yellowish hygromycin-resistant calluses to fresh solid M2 medium for the next round of selection. Three rounds of selection are needed (Fig. 2f).  
**▲ CRITICAL STEP** Do not spread the calluses when you transfer them; keep them grouped together and keep the groups separated. The sequenced regenerated plants need to come from different transgenic events; thus, we advise sequencing only one plant for each callus colony.  
**? TROUBLESHOOTING**

**Regenerate testable seedlings ● Timing 4–6 weeks**

- 50 After 3 rounds of selection cycles, place yellow, actively dividing hygromycin-resistant calluses into 9-cm Petri dishes with 35 ml solid M3 medium, 5–10 calluses per dish.  
 51 Incubate the plates in a lit growth incubator at 28 °C in normal light (16 h:8 h photoperiod) for ~4–5 weeks for regeneration until the seedlings grow to 1 cm (Fig. 2g).  
**? TROUBLESHOOTING**  
 52 Transfer whole calluses bearing multiple seedlings to new plates containing 30 ml solid M4 medium, 5–10 seedlings per plate, and leave them intact; incubate the plates in a lit growth incubator at 28 °C in normal light (16 h:8 h photoperiod) for 7 d (Fig. 2h).

**Screening and identification of mutants ● Timing 3 d**

- 53 Cut off a piece of leaf (~10 × 5 mm) aseptically from each plantlet for each callus colony.  
 54 Extract genomic DNA using a DNAquick Plant System, following the user manual instructions; add 60 μl ddH<sub>2</sub>O or TE buffer (from the DNAquick Plant System) for elution.  
 55 Design primers for screening the mutant plants. We recommend using the NCBI Primer-BLAST tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) to design primers with default parameters and input sequences from 500 bp upstream to 500 bp downstream of the target site.  
**! CAUTION** Multiple pairs of primers may be needed to get the desired PCR products. We advise that two pairs of primers be tested.

- 56 PCR-amplify the genomic DNA using primers across target sites.

Component	Volume ( $\mu$ l)	Final conc.
Rapid Taq Master Mix (2 $\times$ )	25	1 $\times$
On-target-test-F (10 pmol/ $\mu$ l; Table 1)	1	0.3 $\mu$ m
On-target-test-R (10 pmol/ $\mu$ l; Table 1)	1	0.3 $\mu$ m
Genomic DNA from Step 54	2	-
ddH <sub>2</sub> O	21	-

- 57 Perform PCR using the following conditions:

Cycle no.	Denaturation	Annealing	Extension
1	95 °C, 3 min		
2–32	95 °C, 15 s	$T_m$ °C, 15 s	72 °C, 1 min
33			72 °C, 5 min

$T_m$ , melting temperature.

- 58 Run 5  $\mu$ l of the PCR products on a 1.2% (wt/vol) agarose gel in TAE buffer at 160 V for 10 min to confirm that the amplified fragment is unique (that a single band of the expected size is obtained). Single and bright bands are expected.
- 59 Purify 20  $\mu$ l PCR products with an AxyPrep DNA Gel Extraction Kit and subject to Sanger sequencing.
- ▲ CRITICAL STEP** At least ten plants regenerated from different callus colonies are needed for Sanger sequencing per vector. For a high-efficiency base-editing system (such as BE3), ten plants from ten callus colonies are enough. If the editing efficiency of the base editor is low (no mutant plants were identified across ten sequenced plants), sequencing more plants will be necessary.
- 60 Calculate the on-target mutation rates of the base editor-treated plants as follows:  $a/b \times 100\%$ , where  $a$  is the number of mutant plants edited at the target site identified by Sanger sequencing, and  $b$  is the number of total plants sequenced
- 61 Transfer the mutants with on-target edits to 18-cm-length glass containers with ~60 ml solid M4 medium for further development at 28 °C in normal light (16 h:8 h photoperiod) until the mutants grow to ~18 cm (Fig. 2i), one mutant per container.

### Performing WGS ● Timing 7 d

- 62 Cut off leaves (~10  $\times$  5 mm) aseptically from each mutant plant and mix 3–5 leaves from each plant as one sample.
- 63 Extract genomic DNA using a DNAquick Plant System and following the user manual instructions; add 60  $\mu$ l ddH<sub>2</sub>O or TE buffer (from the DNAquick Plant System) for elution.
- 64 Perform WGS at an average coverage of 50 $\times$ , using the NovaSeq 6000.
- ▲ CRITICAL STEP** It normally takes 3 d to obtain the results. We compared the effects of different sequence depths on the accuracy of the WGS method, and we found that  $\geq 20\times$  sequence depth is needed (Fig. 4).

### Processing of raw reads ● Timing 20 h

- 65 Use Trimmomatic to trim adaptors and filter low-quality reads (reads with >10% low-quality bases (quality <10) are filtered) with the following command:

```
java -jar ./Software/Trimmomatic-0.36/trimmomatic-0.36.jar PE
-threads $core -phred33 $R1 $R2 $outdir/R1_trim_clean.fq.gz
$outdir/R1_trim_unpaired_clean.fq.gz $outdir/R2_trim_clean.fq.gz
$outdir/R2_trim_unpaired_clean.fq.gz ILLUMINACLIP:./Software/
Trimmomatic-0.36/adapters/TruSeq3-PE.fa:2:30:3 LEADING:2 TRAILING:10
SLIDINGWINDOW:1:2 MINLEN:75
#1.2 filter Quality < 10 over 10% reads;
```

```

my $QUAL=10;
my $PROPER=10;
gunzip -c $outdir/R2_trim_paired.clean.fq.gz|fastq_quality_filter -v
-q $QUAL -p $PROPER -z -Q 33 -o $outdir/R1.fastx.fq.gz
gunzip -c $outdir/R2_trim_paired.clean.fq.gz|fastq_quality_filter -v
-q $QUAL -p $PROPER -z -Q 33 -o $outdir/R2.fastx.fq.gz
#1.3 fix mispair;
perl Script/fixPaired.pl $outdir/R1.fastx.fq.gz $outdir/R2.fastx.fq.
gz $project $outdir

#1.4 gzip final fastq file

pigz -p $score $outdir/R1.paired.fastq
pigz -p $score $outdir/R2.paired.fastq

```

The detailed code for this Step is available in GitHub at <https://github.com/ReiGao/GWSBE/blob/master/1.Clean.sh>.

- 66 Use the BWA-MEM algorithm to map the reads to the reference genome with the following command:

```

SampleName=WT-1
CORE=10
mkdir tmp_pipe_data
bwa mem -M -t $CORE -R "@RG\tID:$SampleName\tLB:$SampleName\tSM:$Sample-
Name\tPL:illumina\tPU:$SampleName" $REF $R1 $R2|samtools view -bS - >
tmp_pipe_data/$SampleName.source.bam

```

**▲ CRITICAL STEP** The detailed code for Steps 65–69 is available in GitHub at <https://github.com/ReiGao/GWSBE/blob/master/2.bwa.sh>. The ZH11 reference genome used in this study can be downloaded from <http://mbkbase.org/ZH11/>.

- 67 Use Samtools to view, sort and index the BAM files with the following command:

```

samtools view -h -b -q30 tmp_pipe_data/$SampleName.source.bam > tmp_
pipe_data/$SampleName.bam
samtools sort -m lg -@ $CORE tmp_pipe_data/$SampleName.bam -o tmp_
pipe_data/$SampleName.sorted.bam
samtools index tmp_pipe_data/$SampleName.sorted.bam
samtools flagstat tmp_pipe_data/$SampleName.source.bam > tmp_pipe_
data/$SampleName.source.mapinfo

```

- 68 Use Picard to mark duplicate reads of BAM files with the following command:

```

java -Xmx10g -Djava.io.tmpdir=/tmp -jar ./Software/picard-tools-1.102/
FixMateInformation.jar I=tmp_pipe_data/$SampleName.sorted.bam O=tmp_
pipe_data/$SampleName.fix.bam SO=coordinate VALIDATION_STRINGENCY=
LENIENT CREATE_INDEX=true
java -Xmx10g -Djava.io.tmpdir=/tmp -jar $picard/MarkDuplicates.jar
I=tmp_pipe_data/$SampleName.fix.bam O=tmp_pipe_data/$SampleName.
markdup.bam METRICS_FILE=tmp_pipe_data/$SampleName.markdup.metricsFile
CREATE_INDEX=true VALIDATION_STRINGENCY=LENIENT REMOVE_DUPLICATES=
true ASSUME_SORTED=true

```

- 69 Use GATK to realign the reads near indels (-T RealignerTargetCreator and -T IndelRealigner commands) with the following command:

```

java -Xmx10g -Djava.io.tmpdir=/tmp -jar ./Software/GenomeAnalysisTK-
3.4-46/GenomeAnalysisTK.jar -nt $CORE -T RealignerTargetCreator -R $REF
-o tmp_pipe_data/$SampleName.intervals.list -I tmp_pipe_data/$Sample-
Name.markdup.bam

```

```
java -Xmx10g -Djava.io.tmpdir=/tmp -jar ./Software/GenomeAnalysisTK-3.4-46/GenomeAnalysisTK.jar -T IndelRealigner -R $REF -I tmp_pipe_data/$SampleName.markdup.bam --targetIntervals tmp_pipe_data/$SampleName.intervals.list -o tmp_pipe_data/$SampleName.realigned.bam
```

The mapping results are deposited into the tmp\_pipe\_data folder. The \*.realigned.bam files are the BAM files used for variant calling.

### Variant calling ● Timing 30 h

70 Use GATK to call SNVs and indels; all the BAM files in the tmp\_pipe\_data folder will be used to call SNVs and indels using the following command:

```
java -Xmx35g -Djava.io.tmpdir=$dir/tmp -jar ./Software/GenomeAnalysisTK-3.4-46/GenomeAnalysisTK.jar -nt $score -glm BOTH -T UnifiedGenotyper -R $refdb $input -o $dir/all_last.vcf -metrics $dir/all.UniGenMetrics -stand_call_conf 50.0 -stand_emit_conf 10.0 -dcov 1000 -A Coverage -A AlleleBalance
java -Xmx35g -Djava.io.tmpdir=$dir/tmp -jar ./Software/GenomeAnalysisTK-3.4-46/GenomeAnalysisTK.jar -R $refdb -T VariantFiltration -V:VCF $dir/all_last.vcf -o $dir/all_last.filtered.vcf --clusterWindowSize 10 --filterExpression "MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1)" --filterName "HARD_TO_VALIDATE" --filterExpression "DP < 5" --filterName "LowCoverage" --filterExpression "QUAL < 30.0" --filterName "VeryLowQual" --filterExpression "QUAL > 30.0 && QUAL < 50.0" --filterName "LowQual" --filterExpression "QD < 1.5" --filterName "LowQD"
rm -Rf $dir/tmp
```

71 Use LoFreq to call SNVs according to the following command:

```
core=10refdb=./zh11_genome.fa
refdb=./zh11_genome.fa
./Software/lofreq call-parallel --pp-threads $score -f $refdb -o lofreq/$name.vcf INPUTBAM
```

72 Use Strelka2 to call SNVs and indels according to the following command:

```
core=10refdb=./zh11_genome.fa
./Software/strelka-2.9.10.centos6_x86_64/bin/configureStrelkaGermlineWorkflow.py \
--bam INPUTBAM \
--referenceFasta $refdb \
--runDir ./strelka/SampleName
cd ./strelka/$name
./runWorkflow.py -m local -j $score
```

### Filter out wild-type mutations and on-target mutations and obtain the intersection of the results of the three software packages ● Timing 0.5 h

73 Filter out all the variants (SNVs and indels) identified in wild-type plants using AnaGATK.pl, AnaLoVcf.pl and AnaStr2Vcf.pl. The output files for this step are \*.filter.vcf (LoFreq/Strelka2) and filtered.vcf (GATK). The third columns of the \*.filter.vcf and filtered.vcf files show the genetic differences between wild-type plants and the reference genome. In this column, 0 means that all wild-type plants and reference genomes are the same, and 1 means that the wild-type plants and reference genomes are different, which are filtered out. Use the following command:

```
cd Lofreq
perl ../Script/AnaLoVcf.pl
cd ..
cd strelka
```

```
perl ../Script/AnaStr2Vcf.pl
cd ..
cd GATK
perl ../Script/AnaGATK.pl
cd ..
```

▲ **CRITICAL STEP** The detailed code for this step is available in GitHub at <https://github.com/ReiGao/GWSBE/tree/master/Script>.

- 74 Intersect the GATK, LoFreq and Strelka2 filtered results. The script “ALLSITE.pl” is used to read the \*.filter.vcf (LoFreq/Strelka2) and filtered.vcf (GATK) files, and to produce the output files “ALLSite.vcf”. Use the following command:

```
perl ./Script/ALLSITE.pl
```

▲ **CRITICAL STEP** The detailed code for this step is available in GitHub at <https://github.com/ReiGao/GWSBE/tree/master/Script>.

- 75 Use BbMap (<https://github.com/BioInfoTools/BbMap>) to obtain the positions of the sgRNAs, using the following command:

```
bbmap/bbmap.sh in=Target.fa ref=./ZH11_genome_chr.cor.fasta
out=bbmap.sam maxindel=100 k=10 slow.
```

- 76 Convert the output files ending in ‘bbmap.sam’ to files ending in ‘range.txt’ in order to obtain the following format:

```
cat range.txt
qName tName tStrand tStart tEnd
PBE-ACC-T3_P2_JS-2_HP2/0_23 Chr5 1 13098162 13098185
PBE-ACC-T1_ACCB/0_23 Chr5 0 13097961 13097984
PBE-ACC-T2_ACCC/0_23 Chr5 0 13098250 13098273
PBE-WXB_P3_JS-5_HP3/0_23 Chr6 0 1765048 1765071
.
```

- 77 Obtain the on-target variants in all the samples; the output file is ALL.vcf.GeneSample.pos.txt, which contains all on-target information for all samples. Use the following commands:

```
perl getGeneRange.pl ALLSite.vcf ./range.txt
awk '{print $2"\t"$3"\t"$3}' ALL.vcf.GeneSample.pos.txt > target.txt
```

- 78 Filter out on-target sites using the script ALLSITEnotarget.pl to read the ‘target.txt’ file:

```
perl ALLSITEnotarget.pl
```

This produces the following final results files:

- ALLnotarget.vcf: The intersection .vcf file without on-target mutations; it has the same format as the GATK .vcf file.
- AllSNP.jiaoji.summary.txt: This contains SNV statistics files for all samples.
- Mutation.\*.txt: This contains statistics for the different mutation types for each software package (JJ name means intersection; an example of output data is shown in Supplementary Table 3).
- VennALLSITE: This directory contains specific information on the SNVs generated by each package for all the samples; this can be used to draw Venn diagrams later.

▲ **CRITICAL STEP** The detailed code for this step is available in GitHub at <https://github.com/ReiGao/GWSBE/tree/master/Script>

**Sanger sequencing to confirm the SNVs and indels identified by WGS ● Timing 2 d**

- 79 Randomly select 20–50 SNVs and indels in the ALLnotarget.vcf file, one of the output files from Step 78.

80 Design primers to PCR-amplify the selected SNVs and indels using the Primer-BLAST tool on the NCBI website (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>); the recommended product size is 500 bp.

**! CAUTION** Multiple pairs of primers may be needed to get the desired PCR products.

81 Prepare the following reaction to PCR-amplify genomic DNA from Step 63 from both mutant and wild-type plants (at least ten mutant plants and ten wild-type plants are needed to test), using primers designed in Step 80.

Component	Volume ( $\mu$ l)	Final conc.
Rapid Taq Master Mix (2 $\times$ )	25	1 $\times$
Variants-test-F (10 pmol/ $\mu$ l; Table 1)	1	0.3 $\mu$ m
Variants-test-R (10 pmol/ $\mu$ l; Table 1)	1	0.3 $\mu$ m
Genomic DNA from Step 63	2	—
ddH <sub>2</sub> O	21	—

82 Perform PCR, using the following conditions:

Cycle no.	Denaturation	Annealing	Extension
1	95 °C, 3 min		
2–34	95 °C, 15 s	$T_m$ °C, 15 s	72 °C, 10 s
35			72 °C, 5 min

$T_m$ , melting temperature.

### ? TROUBLESHOOTING

83 Assess the accuracy of variant calling; a correct rate of  $\geq 90\%$  is acceptable.

### ? TROUBLESHOOTING

## Troubleshooting

Troubleshooting advice can be found in Table 2.

**Table 2 | Troubleshooting**

Step	Problem	Possible reason	Solution
38	No visible calluses on mature seeds	Lack of seed vigor, medium composition	Do not use old seeds with low vigor. H <sub>2</sub> O <sub>2</sub> is needed for newly harvested seeds to break the seed dormancy. Check the medium composition
48	The calluses are contaminated in the first round of selection (with visible fungal or bacterial colony)	The <i>Agrobacterium</i> may be contaminated	Re-transform the <i>Agrobacterium</i> and ensure aseptic processing procedures
49	The calluses are contaminated in the second round of selection	Incorrect antibiotic concentrations	Increase the carbenicillin antibiotic concentration to 150–200% (300–400 mg/liter final conc.) in the solid M2 medium, which is an efficient method of suppressing contamination
51	No green seedlings appear	Incorrect medium composition or the calluses were too wet	Check the solid M3 medium composition to guarantee adequate nutrients. Keep the calluses in the Petri dishes dry
82	No PCR product or incorrect length or sequence	Incorrect PCR conditions or PCR primers. Incorrect variants identified by WGS	Improve PCR conditions. Use multiple pairs of primers. Variants in highly repetitive regions of the genome are difficult to amplify specifically
83	A low accuracy of variant calling (<80%)	Incorrect parameters in variant-calling process	Check the parameter setting in variant-calling process, shown in Steps 70–72. Check for the presence of single-primer binding sites in the genome to avoid nonspecific amplification, especially in highly repetitive regions of the genome

## Timing

- Steps 1–12, production of vectors: 10 d
- Steps 13–24, transformation of *Agrobacterium* by electroporation: 3 d
- Steps 25–32, harvesting rice seeds and growing wild-type plants: 15 d
- Steps 33–39, growing calluses from mature seeds: 1 month
- Steps 40–47, infection and co-cultivation of calluses and *A. tumefaciens*: 3–6 d
- Steps 48 & 49, tissue culture: 6 weeks
- Steps 50–52, regenerate testable seedlings: 4–6 weeks
- Steps 53–61, screening and identification of mutants: 3 d
- Steps 62–64, performing WGS: 7 d
- Steps 65–69, processing of raw reads: 20 h
- Steps 70–72, variant calling: 30 h
- Steps 73–78, filtering out wild-type mutations and on-target mutations, and obtain the intersection of the results of the three software packages: 0.5 h
- Steps 79–83, Sanger sequencing to confirm the SNVs and indels identified by WGS: 2 d

## Anticipated results

Using this protocol, we have analyzed the off-target changes made by a third-generation base editor (BE3), high-fidelity BE3 (HF1-BE3), and ABE. Sixty-five regenerated plants and 12 wild-type plants were sequenced, and the intersections of the products identified by the three software programs were identified as true variants (Fig. 3a). The numbers of indels in the three base editor groups did not differ from those of the control group (Fig. 3b). When we compared the numbers of SNVs in the three base editor groups, we found that BE3 and HF1-BE3, but not ABE, generated significantly more SNVs than were detected in the control group ( $P = 0.0012$  and  $0.0004$ , respectively) (Fig. 3c). Six SNVs were identified by WGS, as well as by Cas-OFFinder (Fig. 3d). Because these six SNVs are predictable off-target edits with one, two, or four mismatches with regard to the on-target sequences (Supplementary Table 4), the SNVs whose numbers increased were sgRNA-independent edits (Fig. 3d).

### Reporting Summary

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

### Data availability

All the sequence data have been deposited in NCBI BioProject under accession code [PRJNA522656](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA522656), in which CBE refers to BE3 and HF1-CBE to HF1-BE3.

### Code availability

All the code used in this protocol is available on GitHub at: <https://github.com/ReiGao/GWSBE>. The code in this protocol has been peer-reviewed.

## References

1. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
2. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
3. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, 6305 (2016).
4. Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
5. Liang, P. et al. Effective gene editing by high-fidelity base editor 2 in mouse zygotes. *Protein Cell* **8**, 601–611 (2017).
6. Liu, Z. Q. et al. Highly efficient RNA-guided base editing in rabbit. *Nat. Commun.* **9**, 2717 (2018).
7. Tanaka, S. et al. In vivo targeted single-nucleotide editing in zebrafish. *Sci. Rep.* **8**, 11423 (2018).
8. Zhang, Y. H. et al. Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat. Commun.* **8**, 118 (2017).



9. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
10. Li, C. et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* **19**, 59 (2018).
11. Zong, Y. et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**, 950–953 (2018).
12. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
13. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
14. Jiao, R. & Gao, C. Anything impossible with CRISPR/Cas9? *Sci. China Life. Sci.* **60**, 445–446 (2017).
15. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
16. Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat. Rev. Genet.* **17**, 300–312 (2016).
17. Tycko, J., Myer, V. E. & Hsu, P. D. Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell.* **63**, 355–370 (2016).
18. Zhang, F. Development of CRISPR-Cas systems for genome editing and beyond. *Q. Rev. Biophys.* **52**, E6 (2019).
19. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
20. Zuo, E. W. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
21. Lee, H. K., Smith, H. E., Liu, C., Willi, M. & Hennighausen, L. Cytosine base editor 4 but not adenine base editor generates off-target mutations in mouse embryos. *Commun. Biol.* **3**, 19 (2020).
22. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **3**, 620–628 (2020).
23. Gabriel, R. et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* **29**, 816–823 (2011).
24. Wang, X. L. et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* **33**, 175–178 (2015).
25. Frock, R. L. et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* **33**, 179–186 (2015).
26. Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).
27. Kim, D., Kim, S., Kim, S., Park, J. & Kim, J. S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res.* **26**, 406–415 (2016).
28. Cameron, P. et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods* **14**, 600–606 (2017).
29. Tsai, S. Q. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).
30. Crosetto, N. et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).
31. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
32. Wienert, B. et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286–289 (2019).
33. Li, J. et al. 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 (2019).
34. Willi, M., Smith, H. E., Wang, C., Liu, C. & Hennighausen, L. Mutation frequency is not increased in CRISPR-Cas9-edited mice. *Nat. Methods* **15**, 756–758 (2018).
35. Iyer, V. et al. No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice. *PLoS Genet* **14**, e1007503 (2018).
36. Tang, X. et al. 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 (2018).
37. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
38. Jackson, S. A. Rice: The first crop genome. *Rice* **9**, 14 (2016).
39. Shukla, V. K. et al. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* **459**, 437–441 (2009).
40. Townsend, J. A. et al. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature* **459**, 442–445 (2009).
41. Li, T., Liu, B., Spalding, M. H., Weeks, D. P. & Yang, B. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* **30**, 390–392 (2012).
42. Wang, Y. P. et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947–951 (2014).

43. Wood, A. J. et al. Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307–307 (2011).
44. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
45. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
46. Yang, L. H. et al. Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nat. Commun.* **5**, 5507 (2014).
47. Feng, Z. Y. et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **111**, 4632–4637 (2014).
48. Feng, C. et al. High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize. *Plant Biotechnol. J.* **16**, 1848–1857 (2018).
49. Nekrasov, V. et al. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* **7**, 482 (2017).
50. Liang, P. et al. Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat. Commun.* **10**, 67 (2019).
51. Kim, D., Kim, D. E., Lee, G., Cho, S. I. & Kim, J. S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).
52. Kim, D. et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
53. Kim, D. & Kim, J. S. DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. *Genome Res.* **28**, 1894–1900 (2018).
54. Crosetto, N. et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).
55. Goodstein, D. M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
56. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
57. Zhang, Y., Li, J. & Gao, C. Generation of stable transgenic rice (*Oryza sativa* L.) by *Agrobacterium*-mediated transformation. *Curr. Protoc. Plant Biol.* **1**, 235–246 (2016).
58. Wei, F. J. et al. Somaclonal variation does not preclude the use of rice transformants for genetic screening. *Plant J.* **85**, 648–659 (2016).
59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
60. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
61. Genomes Project Consortium . et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
62. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
63. Wilm, A. et al. 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 (2012).
64. Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
65. Shan, Q., Wang, Y., Li, J. & Gao, C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* **9**, 2395–2410 (2014).
66. Shan, Q. et al. Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Mol. Plant* **6**, 1365–1368 (2013).
67. Hiei, Y. & Komari, T. *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nat. Protoc.* **3**, 824–834 (2008).

### Acknowledgements

We thank Y. Zhang (Department of Plant and Environmental Sciences, University of Copenhagen) for critical revision of the manuscript. This work was supported by grants from the National Natural Science Foundation of China (31788103), the Strategic Priority Research Program of the Chinese Academy of Sciences (Precision Seed Design and Breeding, XDA24020102), the National Key Research and Development Program of China (2016YFD0101804) and the R&D Program in Key Areas of Guangdong Province (2018B020202005).

### Author contributions

S.J. performed the experiments; S.J. designed figures; C.G. supervised the project; S.J., Q.G. and C.G. wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41596-020-00423-y>.

Correspondence and requests for materials should be addressed to C.G.

**Peer review information** *Nature Protocols* thanks Keiji Nishida, Pengcheng Wei and Huanbin Zhou for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

Received: 18 February 2020; Accepted: 24 September 2020;  
Published online: 21 December 2020

#### Related links

##### Key reference using this protocol

Jin, S. et al. *Science* **364**, 292–295 (2019): <https://doi.org/10.1126/science.aaw7166>

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of all covariates tested   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

The raw reads are quality controlled by Trimmomatic, and then mapped to the reference genome by BWA-MEM. Indels are called with two software packages: GATK and Strelka2. SNV are called with three software packages: GATK, Lofreq, and Strelka2. The WGS analysis pipeline is available in GitHub at <https://github.com/ReiGao/GWSBE>

Data analysis

Graphpad prism 6 was used to analyze the data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the sequence data has been deposited in NCBI BioProject under accession code PRJNA522656, in which CBE refers to BE3 and HF1-CBE to HF1-BE3. All the codes used in this protocol are available on Github at: <https://github.com/ReiGao/GWSBE>.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	65 BE-treated plants and 12 wild-type plants were analyzed by WGS.
Data exclusions	No data exclusion.
Replication	All attempts at replication were successful.
Randomization	The identified mutations by WGS were confirmed by Sanger sequencing at randomly selected sites.
Blinding	Samples were not blinded as the data did not require blinding.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging