

BIOTECHNOLOGY

Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice

Shuai Jin^{1,2*}, Yuan Zong^{1,2*}, Qiang Gao^{3*}, Zixu Zhu^{1,2}, Yanpeng Wang¹, Peng Qin⁴, Chengzhi Liang^{2,3}, Daowen Wang^{1,2}, Jin-Long Qiu⁵, Feng Zhang⁶, Caixia Gao^{1,2†}

Cytosine and adenine base editors (CBEs and ABEs) are promising new tools for achieving the precise genetic changes required for disease treatment and trait improvement. However, genome-wide and unbiased analyses of their off-target effects *in vivo* are still lacking. Our whole-genome sequencing analysis of rice plants treated with the third-generation base editor (BE3), high-fidelity BE3 (HF1-BE3), or ABE revealed that BE3 and HF1-BE3, but not ABE, induce substantial genome-wide off-target mutations, which are mostly the C→T type of single-nucleotide variants (SNVs) and appear to be enriched in genic regions. Notably, treatment of rice with BE3 or HF1-BE3 in the absence of single-guide RNA also results in the rise of genome-wide SNVs. Thus, the base-editing unit of BE3 or HF1-BE3 needs to be optimized in order to attain high fidelity.

Many genetic diseases and undesirable traits are due to base-pair alterations in genomic DNA (1, 2). Cytosine and adenine base editors (CBEs and ABEs), which are fusions of a nickase-type Cas9 (nCas9) protein with a deaminase domain, can catalyze the conversion of C to T (C>T) and A>G, respectively, in the target site of a single-guide RNA (sgRNA) (3–6). To investigate base-editing specificity, previous attempts focused on either the limited number of off-target sites predicted by *in silico* or *in vitro* approaches, such as Digenome-seq (7) and EndoV-seq (8), or the proximal and predictable regions of sgRNA binding sites (9–11). Because of the challenges posed by the analysis of large genomes from heterogeneous cells, it is still unclear whether these base editors introduce unwanted genome-wide off-target mutations (6). Analyzing the samples from clonally derived

systems by whole-genome sequencing (WGS) may overcome these limitations, thus yielding an objective assessment of the specificities of base editors at the whole-genome level. In this study, we performed a comprehensive investigation of genome-wide off-target mutations from three widely used base editors, the third-generation base editor (BE3), high-fidelity BE3 (HF1-BE3), and ABE (Fig. 1A), in rice (*Oryza sativa* L.), an important crop species.

A total of 14 base editor constructs targeting 11 genomic sites were transformed into rice via *Agrobacterium* transformation (Fig. 1A, table S1, and methods). Regenerated T0 (primary-transformant) plants edited by BE3, HF1-BE3, or ABE and those transformed with the base editors but without sgRNAs (BE3^{-sgRNA}, HF1-BE3^{-sgRNA}, and ABE^{-sgRNA} plants) were analyzed by WGS (Fig. 1 and fig. S1). In addition, 12 wild-type (WT)

plants were used to filter out background mutations in the rice population (methods), and nine plants that went through the transformation process but with no transfer DNA integration (designated as control plants) were used to evaluate the mutations occurring during tissue culture and transformation (Fig. 1B). To ensure high confidence in base calling, all plants were sequenced at an average depth of 60× (table S2). Genetic changes consisting of single-nucleotide variants (SNVs) and small insertions or deletions (indels) were identified in each plant by using three and two independent variant-calling programs, respectively (fig. S2). The identified mutations were confirmed by Sanger sequencing at randomly selected sites with a 98% success rate (figs. S3 and S4 and table S3). Furthermore, we confirmed efficient on-target base editing through WGS (table S4).

The SNVs identified by WGS in the base editor plants were compared with the off-target mutations predicted by using the software Cas-OFFinder (12). Only six SNVs in BE3-edited plants were found to come from three predicted off-target sites; none of the SNVs in HF1-BE3- or ABE-edited plants concurred with the predicted off-target sites (figs. S5 and S6 and methods). Additional examinations also showed that low

¹State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China. ²University of Chinese Academy of Sciences, Beijing 100864, China. ³State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China. ⁴Rice Research Institute, Sichuan Agricultural University, Chengdu 611130, China. ⁵State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. ⁶Department of Plant and Microbial Biology, Center for Plant Precision Genomics, The Microbial and Plant Genomics Institute, University of Minnesota, Minneapolis, MN 55108, USA.

*These authors contributed equally to this work.

†Corresponding author. E-mail: cxgao@genetics.ac.cn

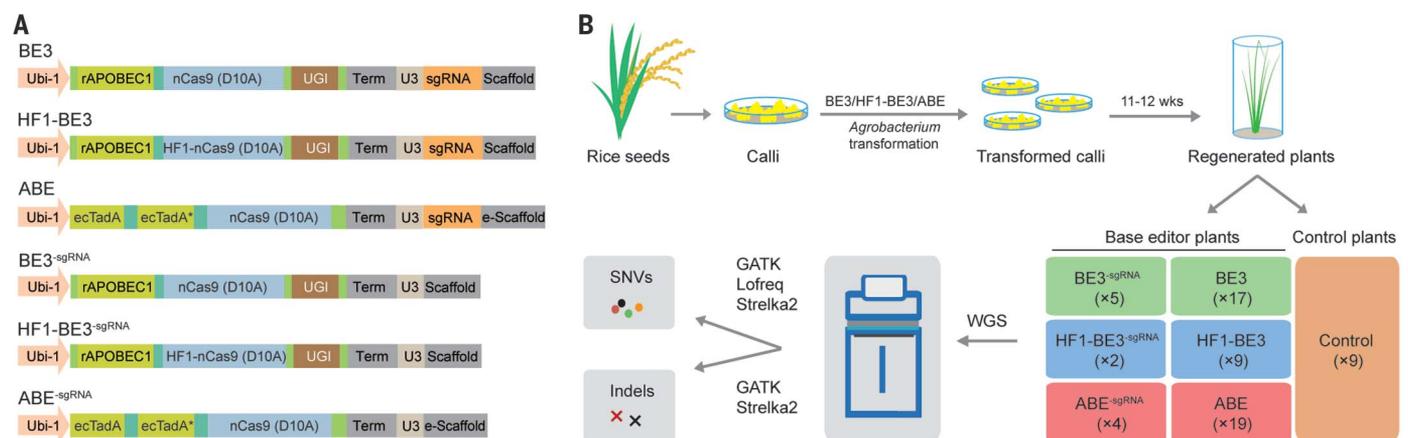


Fig. 1. BE3-, HF1-BE3-, and ABE-mediated base editing in rice.

(A) Schematic representation of the three base editors. (B) Experimental design and workflow. The values in parentheses represent numbers of independent plants used for WGS. Ubi-1, ubiquitin-1 promoter;

rAPOBEC1, rat APOBEC1; D10A, Asp¹⁰→Ala; Term, terminator; U3, rice small nuclear RNA U3 promoter; ecTadA, *E. coli* tRNA-specific adenine deaminase; ecTadA*, an evolved ecTadA variant; e-Scaffold, enhanced sgRNA scaffold; GATK, Genome Analysis Toolkit.

sequence similarity was observed between the adjacent sequences of the identified SNVs and the target sites (figs. S7 and S8 and methods), further indicating that the majority of the SNVs identified by WGS are not predictable by Cas-OFFinder.

We analyzed the indels and SNVs detected in BE3, HF1-BE3, and ABE plants compared with control plants, after removing on-target and predicted off-target SNVs (fig. S5 and tables S4 and S5). The numbers of indels in base editor groups showed no differences from the control group

(Fig. 2A and fig. S9). By contrast, the numbers of SNVs in BE3 and HF1-BE3 groups were significantly higher than those detected in ABE and control groups (Fig. 2B and fig. S10). We classified the SNVs into individual mutation types (figs. S11 and S12 and tables S6 and S7).

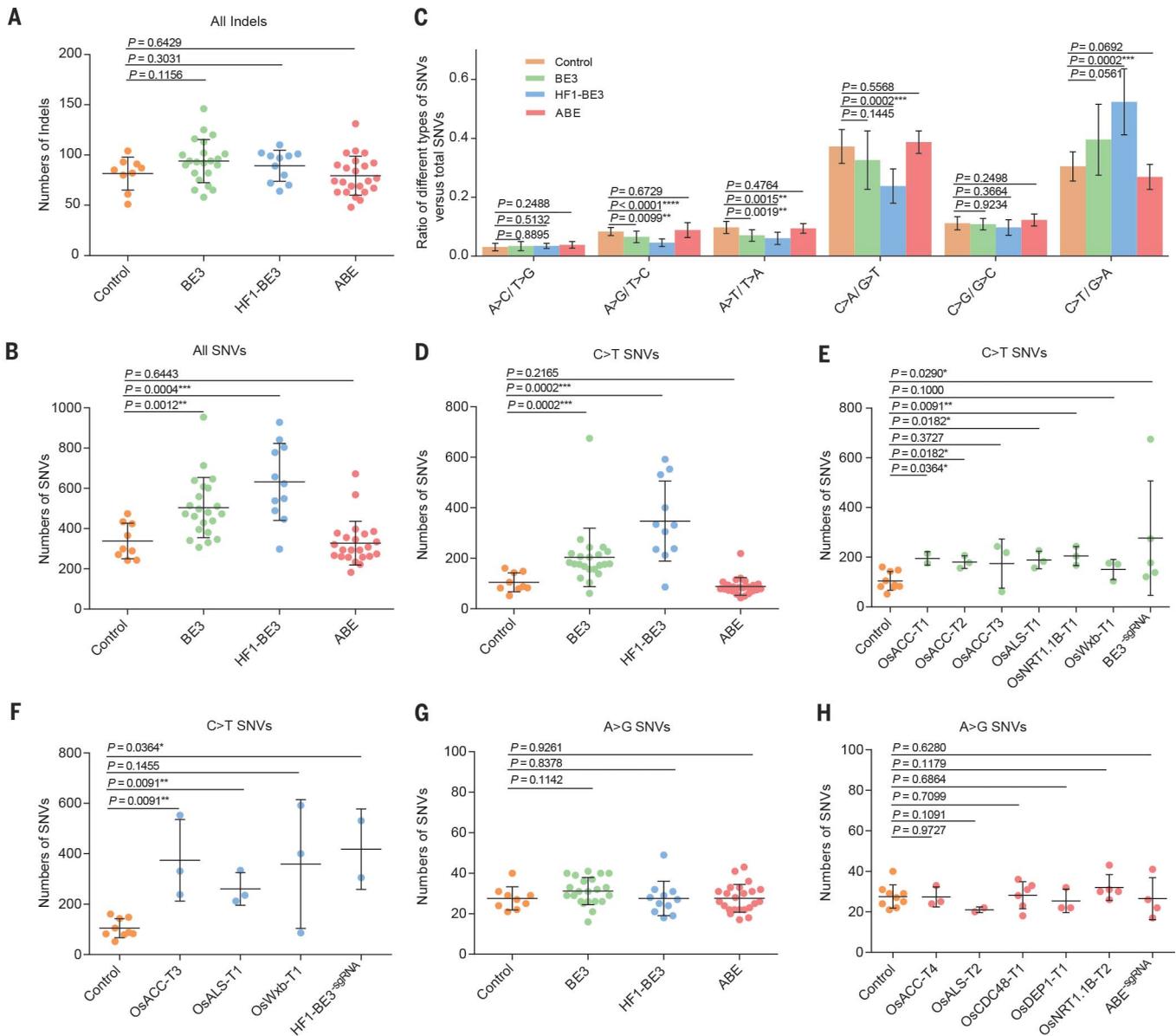


Fig. 2. Analysis of the genetic changes identified by WGS.

(**A** and **B**) Numbers of indels (**A**) and total SNVs (**B**) identified in the BE3, HF1-BE3, and ABE plants. Each dot represents the number of indels or SNVs from an individual plant. The numbers of indels in BE3, HF1-BE3, ABE, and control plants were 94, 89, 79, and 82, respectively. The numbers of total SNVs in BE3, HF1-BE3, ABE, and control plants were 504, 632, 327, and 338, respectively. (**C**) The frequencies of different types of SNVs in the plants exposed to the three base editors and in the control group. (**D**) Comparison of total C>T SNVs in the BE3, HF1-BE3, ABE, and control plants. The numbers of SNVs were 203, 347, 88, and 105, respectively. (**E** and **F**) Analysis of C>T SNVs in the BE3 plants (**E**) or the HF1-BE3 plants (**F**) according to target sites by comparison with the C>T SNVs in the control group and the individuals

treated with BE3^{-sgRNA} or HF1-BE3^{-sgRNA}. Six rice genomic sites (OsACC-T1, OsACC-T2, OsACC-T3, OsALS-T1, OsNRT1.1B-T1, and OsWxb-T1) were targeted by BE3, and three of them (OsACC-T3, OsALS-T1, and OsWxb-T1) were also targeted by HF1-BE3. (**G**) Comparison of total A>G SNVs in the BE3, HF1-BE3, ABE, and control plants. The numbers of SNVs were 31, 28, 28, and 28, respectively. (**H**) Analysis of A>G SNVs in the ABE plants according to target sites by comparison with the A>G SNVs in the control group and the individuals treated with ABE^{-sgRNA}. Five rice genomic sites (OsACC-T4, OsALS-T2, OsCDC48-T1, OsDEP1-T1, and OsNRT1.1B-T2) were targeted by ABE. *P* values were calculated by the Mann-Whitney test; *P* < 0.05 was considered significant. All values represent means ± SD. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

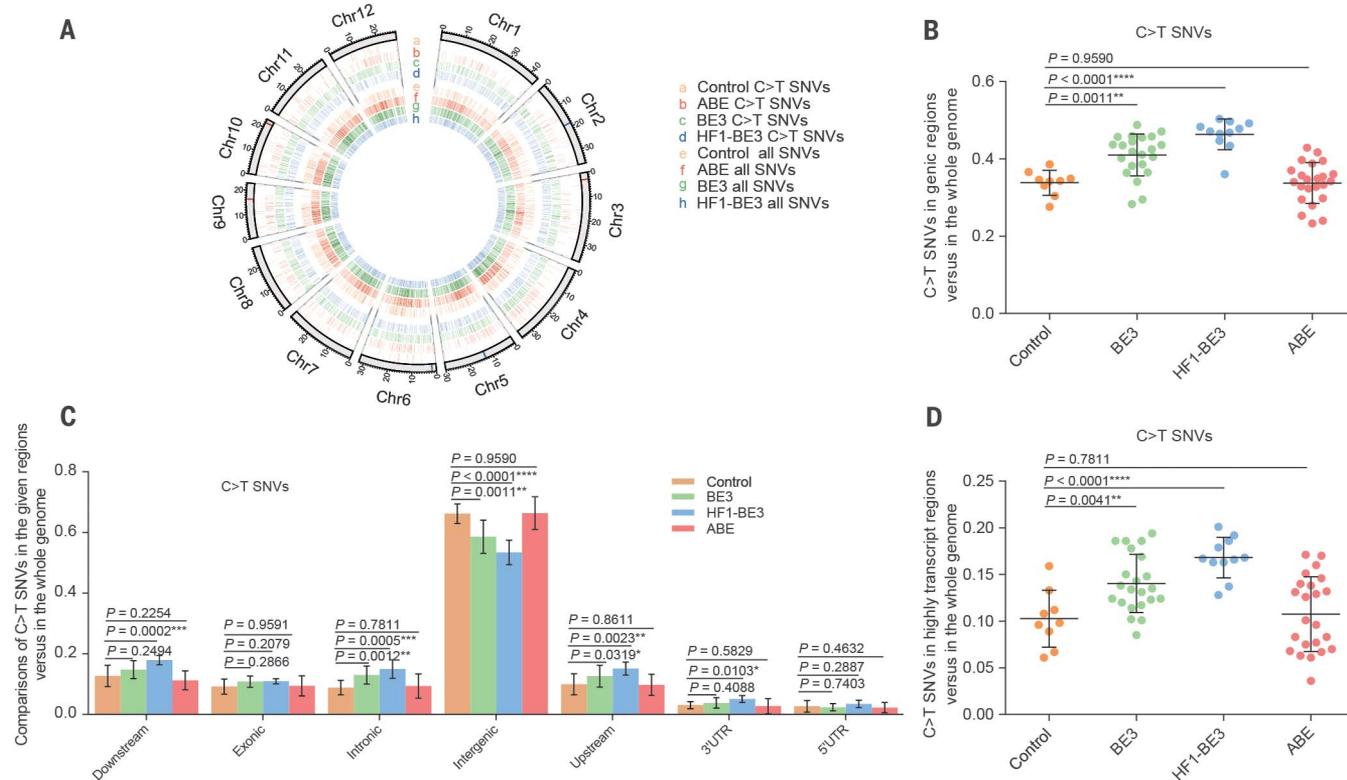


Fig. 3. Genomic distribution of the C>T SNVs identified in the BE3, HF1-BE3, and ABE plants. (A) All SNVs and C>T SNVs are randomly distributed on the 12 rice chromosomes (chr1 to chr12) in BE3, HF1-BE3, ABE, and control groups of plants. (B) C>T SNVs in genic regions versus in the whole genome, compared among BE3, HF1-BE3, ABE, and control groups. (C) Comparisons of C>T SNVs in the given regions versus in the whole genome. (D) C>T SNVs in highly transcribed regions versus in the whole genome among BE3, HF1-BE3, ABE, and control groups. *P* values were calculated by the Mann-Whitney test, and *P* < 0.05 was considered to be statistically significant in (B) to (D). **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001. All values represent means ± SD.

In BE3 and HF1-BE3 groups, the percentages of C>T (G>A) transitions were higher than the percentage obtained for the control group (Fig. 2C). In the ABE group, on the other hand, the levels of the different mutation types were all similar to those in the control group (Fig. 2C). These results suggest that SNVs in plants exposed to BE3 and HF1-BE3 were mainly C>T transitions. In addition, the average numbers of the C>T SNVs in the BE3, HF1-BE3, BE3^{-sgRNA}, and HF1-BE3^{-sgRNA} plants were higher than those found in the ABE and control plants (Fig. 2, D to F). One sample in the BE3^{-sgRNA} group had a notably high number of SNVs (Fig. 2E). Upon examining the sequencing and variant-calling data, we did not find this high-end data point to be caused by experimental error (fig. S11 and table S2). Moreover, omitting this high-end data point from the analysis did not alter the trend that the BE3 and BE3^{-sgRNA} groups had more total SNVs and C>T SNVs than the control and ABE groups (fig. S13). On the other hand, the numbers of the A>G mutations did not differ significantly across the base editor and control groups (Fig. 2, G and H). This is consistent with the results of previous studies showing that the overexpression of different deaminases results in elevated global C>T mutations in *Escherichia coli*, yeast, and humans (4, 13, 14).

Moreover, the uracil glycosylase inhibitor (UGI), present in BE3 and HF1-BE3 but not ABE (Fig. 1A), has also been reported to enhance genome-wide C>T conversion (15). Therefore, we speculate that the higher C>T mutation rates observed in BE3 and HF1-BE3 plants relative to controls may result from APOBEC1 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1) and/or UGI. By contrast, ABE is derived by fusing an nCas9 protein with an engineered RNA adenosine deaminase (5). It is possible that the engineered RNA adenosine deaminase does not show excessive DNA base editing, thus avoiding the generation of genome-wide A>G SNVs outside the sgRNA targeting windows.

We mapped the distribution of SNVs and found that total SNVs and C>T SNVs were distributed throughout the rice genome (Fig. 3A and tables S8 and S9), with no mutation hotspots detected (table S10). In addition, we found that the percentages of C>T SNVs in genic regions were significantly higher in the two BE3 and HF1-BE3 groups than in the ABE or control groups (Fig. 3, B and C; table S11; and methods). In addition, the high numbers of C>T SNVs associated with BE3 and HF1-BE3 are more likely to occur in transcribed genic regions (Fig. 3D, table S12, and methods), where single-stranded DNA is generated by active transcription.

whole genome among BE3, HF1-BE3, ABE, and control groups. 3'UTR and 5'UTR, 3' and 5' untranslated regions. (D) C>T SNVs in highly transcribed regions versus in the whole genome among BE3, HF1-BE3, ABE, and control groups. *P* values were calculated by the Mann-Whitney test, and *P* < 0.05 was considered to be statistically significant in (B) to (D). **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001. All values represent means ± SD.

In summary, BE3 and HF1-BE3, but not ABE, induce genome-wide mutations in rice. These off-target mutations, mainly C>T SNVs that are enriched in transcribed genic regions, are not predicted by current in silico approaches. A similar study also finds that BE3 but not ABE induces substantial off-target mutations in mouse embryos (16). To minimize the off-target base mutations by BE3 or HF1-BE3, functional optimization of cytidine deaminase and/or UGI components is necessary.

Additionally, improved CBEs, such as YEE-BE3, which may have lower DNA affinity than the BE3 used in this study (17), might be employed to help reduce off-target mutations.

REFERENCES AND NOTES

1. M. J. Landrum *et al.*, *Nucleic Acids Res.* **44**, D862–D868 (2016).
2. S. Henikoff, L. Comai, *Annu. Rev. Plant Biol.* **54**, 375–401 (2003).
3. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, *Nature* **533**, 420–424 (2016).
4. K. Nishida *et al.*, *Science* **353**, aaf8729 (2016).
5. N. M. Gaudelli *et al.*, *Nature* **551**, 464–471 (2017).
6. H. A. Rees, D. R. Liu, *Nat. Rev. Genet.* **19**, 770–788 (2018).
7. D. Kim *et al.*, *Nat. Biotechnol.* **35**, 475–480 (2017).
8. P. Liang *et al.*, *Nat. Commun.* **10**, 67 (2019).
9. H. K. Lee *et al.*, *Nat. Commun.* **9**, 4804 (2018).
10. H. A. Rees *et al.*, *Nat. Commun.* **8**, 15790 (2017).
11. J. M. Gehrke *et al.*, *Nat. Biotechnol.* **36**, 977–982 (2018).
12. S. Bae, J. Park, J. S. Kim, *Bioinformatics* **30**, 1473–1475 (2014).
13. R. S. Harris, S. K. Petersen-Mahrt, M. S. Neuberger, *Mol. Cell* **10**, 1247–1253 (2002).

14. I. B. Rogozin *et al.*, *Nat. Immunol.* **8**, 647–656 (2007).
15. E. H. Radany *et al.*, *Mutat. Res.* **461**, 41–58 (2000).
16. E. Zuo *et al.*, *Science* **364**, 289–292 (2019).
17. Y. B. Kim *et al.*, *Nat. Biotechnol.* **35**, 371–376 (2017).

ACKNOWLEDGMENTS

We thank F. Lu (IGDB) for advice in statistics analysis and S. Li (Sichuan Agricultural University) for providing the Zhonghua11 genome. **Funding:** This work was supported by grants to C.G. from the National Natural Science Foundation of China (31788103), the National Key Research and Development Program of China

(2016YFD0101804), and the Chinese Academy of Sciences (QYZDY-SSW-SMC030) and a start-up fund to F.Z. from the College of Biological Sciences, University of Minnesota. **Author contributions:** S.J. and Y.Z. designed and performed experiments. Q.G., S.J., and P.Q. performed WGS analysis. Z.Z. performed experiments. Y.W. designed figures. C.G. and F.Z. supervised the project. All authors wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All the sequencing data were deposited in NCBI BioProject under accession code PRJNA522656, in which CBE refers to BE3 and HF1-CBE refers to HF1-BE3.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6437/292/suppl/DC1
Materials and Methods
Figs. S1 to S13
Table S1 to S13
References (18–34)

20 January 2019; accepted 17 February 2019
Published online 28 February 2019
10.1126/science.aaw7166

Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice

Shuai Jin, Yuan Zong, Qiang Gao, Zixu Zhu, Yanpeng Wang, Peng Qin, Chengzhi Liang, Daowen Wang, Jin-Long Qiu, Feng Zhang and Caixia Gao

Science **364** (6437), 292-295.

DOI: 10.1126/science.aaw7166originally published online February 28, 2019

Spotting off-targets from gene editing

Unintended genomic modifications limit the potential therapeutic use of gene-editing tools. Available methods to find off-targets generally do not work in vivo or detect single-nucleotide changes. Three papers in this issue report new methods for monitoring gene-editing tools in vivo (see the Perspective by Kempton and Qi). Wienert *et al.* followed the recruitment of a DNA repair protein to DNA breaks induced by CRISPR-Cas9, enabling unbiased detection of off-target editing in cellular and animal models. Zuo *et al.* identified off-targets without the interference of natural genetic heterogeneity by injecting base editors into one blastomere of a two-cell mouse embryo and leaving the other genetically identical blastomere unedited. Jin *et al.* performed whole-genome sequencing on individual, genome-edited rice plants to identify unintended mutations. Cytosine, but not adenine, base editors induced numerous single-nucleotide variants in both mouse and rice.

Science, this issue p. 286, p. 289, p. 292; see also p. 234

ARTICLE TOOLS

<http://science.sciencemag.org/content/364/6437/292>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2019/02/27/science.aaw7166.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/364/6437/289.full>
<http://science.sciencemag.org/content/sci/364/6437/234.full>

REFERENCES

This article cites 33 articles, 5 of which you can access for free
<http://science.sciencemag.org/content/364/6437/292#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works