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 Wang[1], Peng Qin[4], Chengzhi Liang[2,3], Daowen Wang[1,2], Jin-Long Qiu[5], Feng Zhang[6], Caixia Gao[1,2]†

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. [2]University of Chinese Academy of Sciences, Beijing 100864, China. [3]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. [4]Rice Research Institute, Sichuan Agricultural University, Chengdu 611130, China. [5]State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. [6]Department 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

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 (WGS) analysis of rice plants treated with 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 inferior 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→T (C>T) and A>G in the target site of a single guide RNA (sgRNA), respectively (*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*). It is still unclear whether these base editors may introduce unwanted genome-wide off-target mutations due to the challenges posed by analysis of large genomes from heterogeneous cells (*6*). Analyzing the samples from clonally derived systems by whole genome sequencing could 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 of BE3, HF1-BE3, and ABE in rice, an important crop species.

We chose three widely used base editors, BE3, HF1-BE3, and ABE (Fig. 1A). 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 plants edited by BE3, HF1-BE3, or ABE and those transformed with the base editors but without sgRNAs 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 T-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 SNVs and small insertions/deletions (indels) were identified in each plant 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 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 to the predicted off-target sites (figs. S5 and S6 and methods). Additional examinations also showed that low sequence similarity was observed between the adjacent sequences of the identified SNVs and the target sites (figs. S7 and S8 and methods), further supporting 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 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). In contrast, the numbers of SNVs per 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). In BE3 and HF1-BE3 groups, the percentages of C>T (G>A) transition 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 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). We noted that one sample in the BE3^{-sgRNA} group had a considerably high number of SNVs (Fig. 2E). This high-end data point was not caused by experimental error upon examining the sequencing and variant calling data (fig. S11 and table S2). Moreover, omitting this high-end data point in 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 previous studies showing that overexpression of different deaminases results in elevated global C>T mutations in *Escherichia coli*, yeast, and human (*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 and/or UGI. By contrast, ABE is derived by fusing a 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 owing to active transcription.

In summary, BE3 and HF1-BE3, but not ABE, induce genome-wide mutations in rice. These off-target mutations, being mainly C>T SNVs and 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 BE3 used in this study (*17*), might be employed to help reduce off-target mutations.

REFERENCES AND NOTES
1. M. J. Landrum, J. M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein, D. R. Maglott, ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016). doi:10.1093/nar/gkv1222 Medline
2. S. Henikoff, L. Comai, Single-nucleotide mutations for plant functional genomics. *Annu. Rev. Plant Biol.* **54**, 375–401 (2003). doi:10.1146/annurev.arplant.54.031902.135009 Medline
3. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016). doi:10.1038/nature17946 Medline
4. K. Nishida, T. Arazoe, N. Yachie, S. Banno, M. Kakimoto, M. Tabata, M. Mochizuki, A. Miyabe, M. Araki, K. Y. Hara, Z. Shimatani, A. Kondo, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016). doi:10.1126/science.aaf8729 Medline
5. N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, D. R. Liu, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017). doi:10.1038/nature24644 Medline
6. H. A. Rees, D. R. Liu, Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018). doi:10.1038/s41576-018-0059-1 Medline
7. D. Kim, K. Lim, S.-T. Kim, S. H. Yoon, K. Kim, S.-M. Ryu, J.-S. Kim, Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017). doi:10.1038/nbt.3852 Medline
8. P. Liang, X. Xie, S. Zhi, H. Sun, X. Zhang, Y. Chen, Y. Chen, Y. Xiong, W. Ma, D. Liu, J. Huang, Z. Songyang, Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat. Commun.* **10**, 67 (2019). doi:10.1038/s41467-018-07988-z Medline
9. H. K. Lee, M. Willi, S. M. Miller, S. Kim, C. Liu, D. R. Liu, L. Hennighausen, Targeting fidelity of adenine and cytosine base editors in mouse embryos. *Nat. Commun.* **9**, 4804 (2018). doi:10.1038/s41467-018-07322-7 Medline
10. H. A. Rees, A. C. Komor, W.-H. Yeh, J. Caetano-Lopes, M. Warman, A. S. B. Edge, D. R. Liu, Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017). doi:10.1038/ncomms15790 Medline
11. J. M. Gehrke, O. Cervantes, M. K. Clement, Y. Wu, J. Zeng, D. E. Bauer, L. Pinello, J. K. Joung, An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018). Medline
12. S. Bae, J. Park, J. S. Kim, Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014). doi:10.1093/bioinformatics/btu048 Medline
13. R. S. Harris, S. K. Petersen-Mahrt, M. S. Neuberger, RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10**, 1247–1253 (2002). doi:10.1016/S1097-2765(02)00742-6 Medline
14. I. B. Rogozin, L. M. Iyer, L. Liang, G. V. Glazko, V. G. Liston, Y. I. Pavlov, L. Aravind, Z. Pancer, Evolution and diversification of lamprey antigen receptors: Evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol.* **8**, 647–656 (2007). doi:10.1038/ni1463 Medline
15. E. H. Radany, K. J. Dornfeld, R. J. Sanderson, M. K. Savage, A. Majumdar, M. M.

Seidman, D. W. Mosbaugh, Increased spontaneous mutation frequency in human cells expressing the phage PBS2-encoded inhibitor of uracil-DNA glycosylase. *Mutat. Res.* **461**, 41–58 (2000). doi:10.1016/S0921-8777(00)00040-9 Medline

16. E. Zuo, Y. Sun, W. Wei, T. Yuan, W. Ying, H. Sun, L. Yuan, L. M. Steinmetz, Y. Li, H. Yang, Cytosine base editor generates substantial off-target single nucleotide variants in mouse embryos. *Science* 10.1126/science.aav9973 (2019).

17. Y. B. Kim, A. C. Komor, J. M. Levy, M. S. Packer, K. T. Zhao, D. R. Liu, Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* 35, 371–376 (2017). doi:10.1038/nbt.3803

18. Y. Zong, Y. Wang, C. Li, R. Zhang, K. Chen, Y. Ran, J.-L. Qiu, D. Wang, C. Gao, Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017). doi:10.1038/nbt.3811 Medline

19. C. Li, Y. Zong, Y. Wang, S. Jin, D. Zhang, Q. Song, R. Zhang, C. Gao, Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* **19**, 59 (2018). doi:10.1186/s13059-018-1443-z Medline

20. Q. Shan, Y. Wang, K. Chen, Z. Liang, J. Li, Y. Zhang, K. Zhang, J. Liu, D. F. Voytas, X. Zheng, Y. Zhang, C. Gao, Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Mol. Plant* **6**, 1365–1368 (2013). doi:10.1093/mp/sss162 Medline

21. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010). doi:10.1093/bioinformatics/btp698 Medline

22. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009). doi:10.1093/bioinformatics/btp352 Medline

23. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi:10.1101/gr.107524.110 Medline

24. A. Wilm, P. P. K. Aw, D. Bertrand, G. H. T. Yeo, S. H. Ong, C. H. Wong, C. C. Khor, R. Petric, M. L. Hibberd, N. Nagarajan, 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). doi:10.1093/nar/gks918 Medline

25. S. Kim, K. Scheffler, A. L. Halpern, M. A. Bekritsky, E. Noh, M. Källberg, X. Chen, Y. Kim, D. Beyter, P. Krusche, C. T. Saunders, Strelka2: Fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018). doi:10.1038/s41592-018-0051-x Medline

26. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013). doi:10.1093/bib/bbs017 Medline

27. B. Bushnell, poster presented at the Ninth Annual DOE Joint Genome Institute User Meeting, Walnut Creek, CA, 18 to 20 March 2014.

28. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi:10.1093/bioinformatics/btq033 Medline

29. J. P. Zhang, Y. Yu, Y.-Z. Feng, Y.-F. Zhou, F. Zhang, Y.-W. Yang, M.-Q. Lei, Y.-C. Zhang, Y.-Q. Chen, MiR408 regulates grain yield and photosynthesis via a phytocyanin protein. *Plant Physiol.* **175**, 1175–1185 (2017). doi:10.1104/pp.17.01169 Medline

30. Y. Tian, X. Ma, C. Lv, X. Sheng, X. Li, R. Zhao, Y. Song, T. Andl, M. V. Plikus, J. Sun, F. Ren, J. Shuai, C. J. Lengner, W. Cui, Z. Yu, Stress responsive miR-31 is a major modulator of mouse intestinal stem cells during regeneration and tumorigenesis. *eLife* **6**, e29538 (2017). doi:10.7554/eLife.29538 Medline

31. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015). doi:10.1038/nmeth.3317 Medline

32. P. Cingolani, A. Platts, L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012). doi:10.4161/fly.19695 Medline

33. M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, M. A. Marra, Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009). doi:10.1101/gr.092759.109 Medline

34. H. Chen, P. C. Boutros, VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**, 35 (2011). doi:10.1186/1471-2105-12-35 Medline

ACKNOWLEDGMENTS
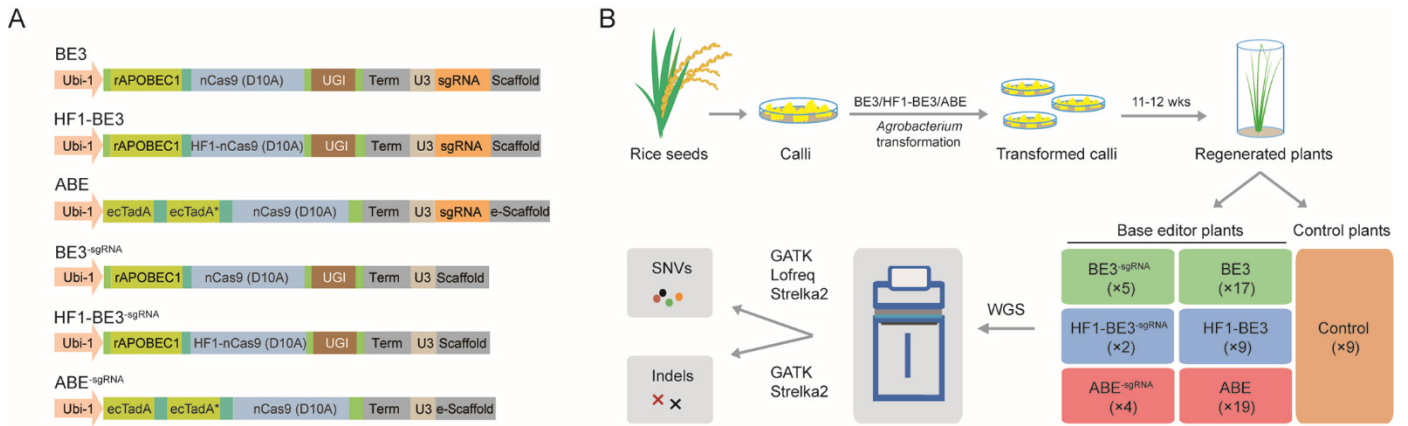
SUPPLEMENTARY MATERIALS

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.
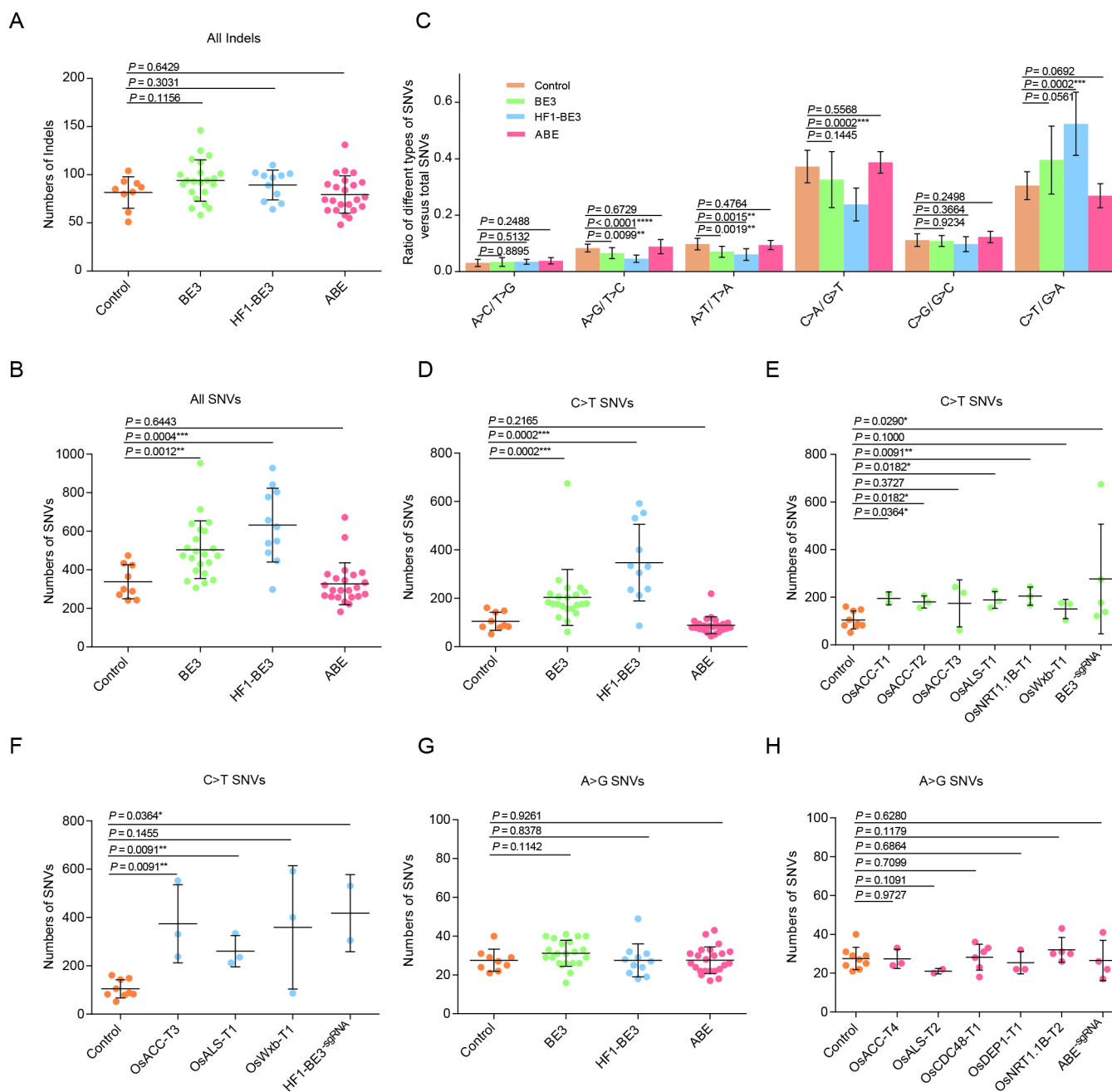
Fig. 2. Analysis of the genetic changes identified by WGS. (**A** and **B**) Numbers of indels (A) and total SNVs (B) identified in the BE3, HF1-BE3, and ABE plants. Each dot represents the number of indels or SNVs from an individual plant. The numbers of indels of BE3, HF1-BE3, ABE, and control were 94, 89, 79, and 82, respectively. The numbers of total SNVs of BE3, HF1-BE3, ABE, and control 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 by BE3[-sgRNA] or HF1-BE3[-sgRNA]. (**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 by ABE[-sgRNA]. *P* values were calculated by the Mann-Whitney test; *P* < 0.05 was considered significant in (A) to (H). All values represent means ± SD.
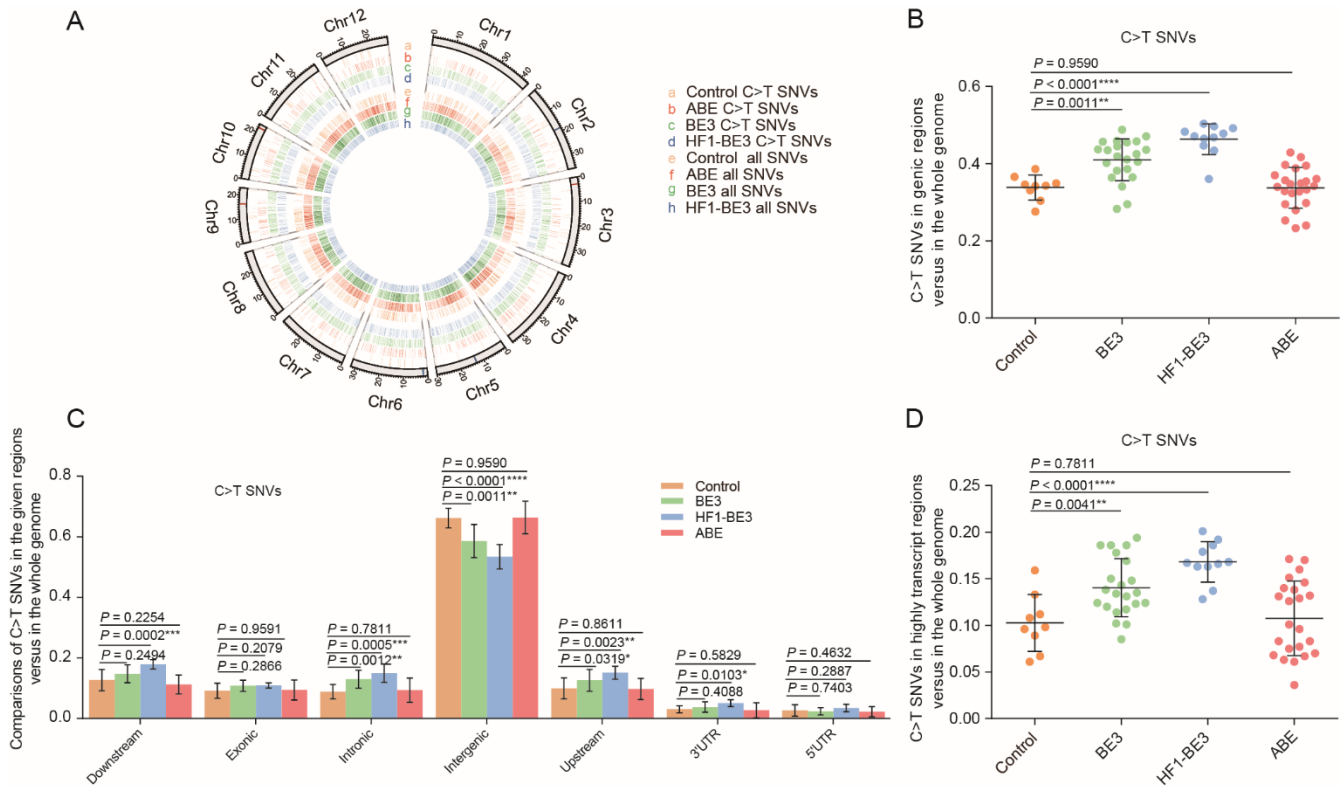
Fig. 3. Genomic distribution of the C>T SNVs identified in the BE3, HF1-BE3, and ABE plants. (A) All SNVs and C>T SNVs are randomly distributed on the 12 rice chromosomes in BE3, HF1-BE3, ABE, and control groups of plants. (B) C>T SNVs in genic regions versus in the whole genome compared among BE3, HF1-BE3, ABE, and control groups of plants. (C) Comparisons of C>T SNVs in the given regions versus in the whole genome among BE3, HF1-BE3, ABE, and control groups of plants. (D) C>T SNVs in highly transcribed regions versus in the whole genome among BE3, HF1-BE3, ABE, and control groups of plants. P values were calculated by the Mann-Whitney test, and $P < 0.05$ was considered to be statistically significant in (B) to (D). All values represent means ± SD.

Science

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

ARTICLE TOOLS	http://science.sciencemag.org/content/early/2019/02/27/science.aaw7166
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2019/02/27/science.aaw7166.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/early/2019/02/27/science.aav9973.full
REFERENCES	This article cites 32 articles, 4 of which you can access for free http://science.sciencemag.org/content/early/2019/02/27/science.aaw7166#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions