

An efficient CRISPR–Cas12a promoter editing system for crop improvement

Received: 8 September 2022

Accepted: 2 March 2023

Published online: 06 April 2023

 Check for updates

Jianping Zhou , Guanqing Liu , Yuxin Zhao¹, Rui Zhang¹, Xu Tang¹, Ling Li¹, Xinyu Jia¹, Yachong Guo¹, Yuechao Wu , Yangshuo Han , Yu Bao , Yao He¹, Qinjin Han¹, Han Yang¹, Xuelian Zheng , Yiping Qi , Tao Zhang & Yong Zhang

Promoter editing represents an innovative approach to introduce quantitative trait variation (QTV) in crops. However, an efficient promoter editing system for QTV needs to be established. Here we develop a CRISPR–Cas12a promoter editing (CAPE) system that combines a promoter key-region estimating model and an efficient CRISPR–Cas12a-based multiplexed or singular editing system. CAPE is benchmarked in rice to produce QTV continuums for grain starch content and size by targeting *OsGBSS1* and *OsGS3*, respectively. We then apply CAPE for promoter editing of *OsD18*, a gene encoding GA3ox in the gibberellin biosynthesis pathway. The resulting lines carry a QTV continuum of semidwarfism without significantly compromising grain measures. Field trials demonstrated that the *OsD18* promoter editing lines have the same yield performance and antilodging phenotype as the Green Revolution *OsSD1* mutants in different genetic backgrounds. Hence, promoter editing of *OsD18* generates a quantitative Green Revolution trait. Together, we demonstrate a CAPE-based promoter editing and tuning pipeline for efficient production of useful QTV continuum in crops.

The rapid development of genome editing technologies has been fuelling innovations in crop breeding to provide timely solutions to sustainably feed 10 billion people in the shadow of climate change¹. Given that most agronomic traits are quantitative, the introduction of quantitative trait variation (QTV) in crops by genome editing holds great promise for revolutionizing and fast-tracking crop breeding. This concept was first demonstrated in the landmark report of clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR associated protein 9 (Cas9)-based multiplexed promoter editing in

tomato². CRISPR–Cas9-based promoter editing was also demonstrated in staple crops such as rice^{3–6} and maize⁷. With the capability of fine-tuning key agronomic traits such as flowering time and plant architecture⁸, promoter editing is a powerful approach for crop improvement.

At present, a major challenge of promoter editing is to efficiently achieve the desired editing outcomes, which are quantitative in nature. However, unlike editing of the protein-coding region, which often leads to the identifiable variation with gene loss-of-function, the phenotypic

¹Department of Biotechnology, School of Life Sciences and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, China. ²Chongqing Key Laboratory of Plant Resource Conservation and Germplasm Innovation, Integrative Science Center of Germplasm Creation in Western China (Chongqing) Science City, School of Life Sciences, Southwest University, Chongqing, China. ³Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding/Jiangsu Key Laboratory of Crop Genetics and Physiology, Agricultural College, Yangzhou University, Yangzhou, China. ⁴Key Laboratory of Plant Functional Genomics of the Ministry of Education/Joint International Research Laboratory of Agriculture and Agri-Product Safety, The Ministry of Education of China, Yangzhou University, Yangzhou, China. ⁵Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou, China. ⁶Department of Plant Science and Landscape Architecture, University of Maryland, College Park, Rockville, MD, USA. ⁷Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD, USA. ⁸These authors contributed equally: Jianping Zhou, Guanqing Liu.  e-mail: Yiping@umd.edu; zhangtao@yzu.edu.cn; zhangyong916@uestc.edu.cn

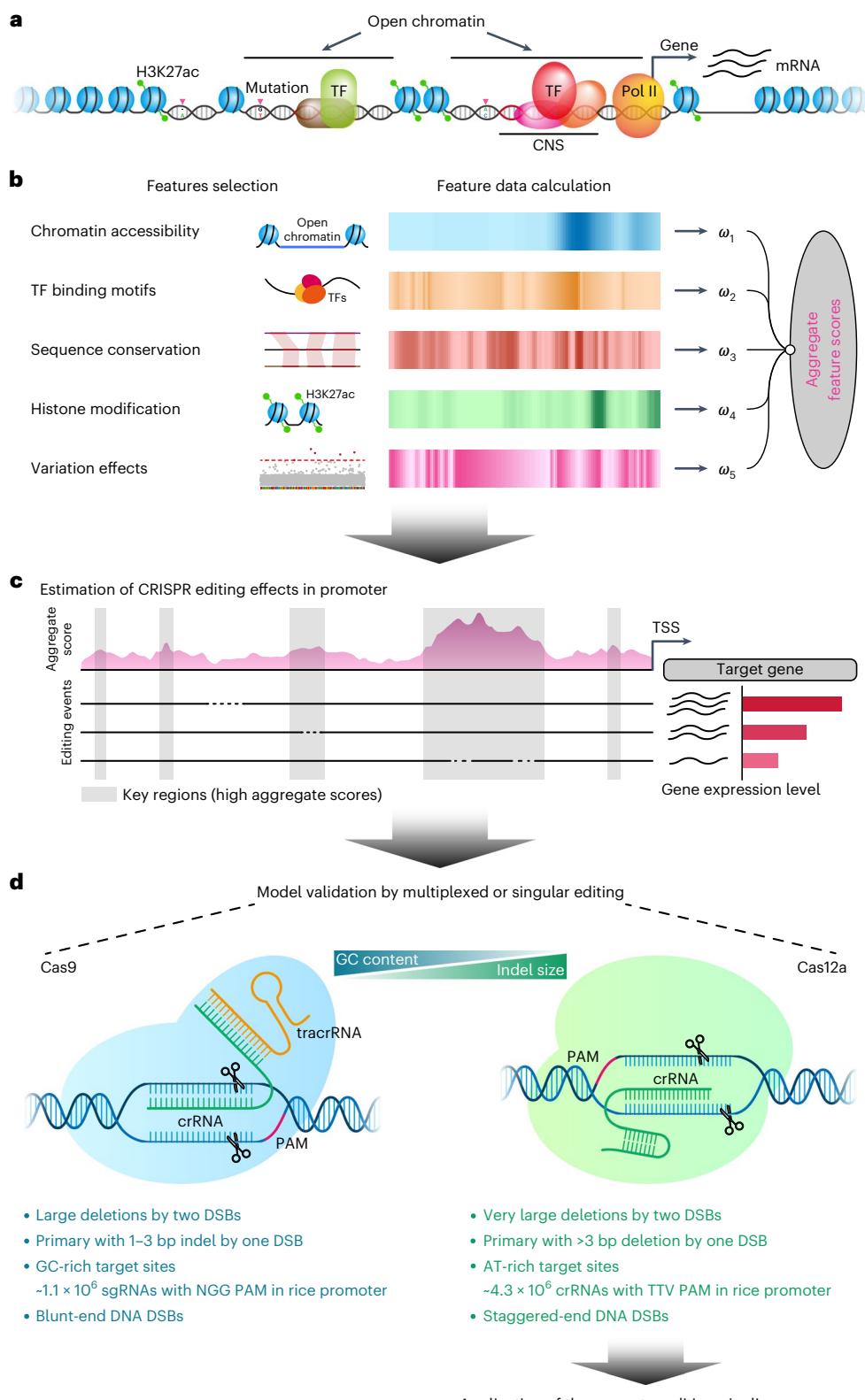


Fig. 1 | Rational design of the CAPE system. **a**, Chromatin architecture within the upstream promoter region that orchestrates the expression of the gene. **b**, Selection of genomic features for building the estimating model. Five features (chromatin accessibility, TF binding sites, sequence conservation, H3K27ac histone modification and variation effects) were selected as input features. Data from all features within the promoter region were chunked into 20-bp bins and normalized. Processed data were aggregated with given weights. Each bin

was assigned an aggregate score, which represents the putative effect of the bin on the target-gene expression. **c**, Schematic of the putative effects on the target gene by CRISPR-based promoter editing. The aggregate scores of the promoter regions are marked in pink. The predicted KRs with high aggregate scores are shaded in grey. **d**, Characteristic comparisons of the two most popular CRISPR systems supports CRISPR–Cas12a as a promising promoter editing tool. TracrRNA, trans-activating CRISPR RNA.

outcomes of promoter region editing are hard to predict^{9,10}. This is largely because massive *cis*-regulatory elements (CREs), along with functionally neutral sequences, are harboured in the promoter regions in which CREs orchestrate with various transcription factors (TFs) to regulate spatiotemporal gene expression^{11,12}. Hence, multiple factors associated with CREs—such as chromatin accessibility^{13–15}, histone post-transcriptional modifications (PTMs)^{11,16}, TF binding motifs^{12,17}, conserved noncoding sequences (CNSs)^{18–20} and variation effects (the effects of genetic variations on phenotypes)^{21–23}—should be considered. In addition, CREs are commonly enriched in open chromatin regions, where chromatin has high accessibility^{14,24}. Active plant promoters are characterized by active histone modifications such as H3K4me3 and H3K27ac, which can be used for identifying CRE activity^{14,25,26}. Considering all these factors, it is imperative to build a biocomputational model that can assign quantitative values to different regions within a promoter, which will in turn effectively guide promoter editing experiments to fine-tune gene expression and generate desirable quantitative traits in crops.

To optimize promoter editing, it is necessary to use the most appropriate and efficient CRISPR system. Previous promoter editing studies had relied on the Cas9 nuclease, which generates blunt-end DNA double-strand breaks (DSBs), leading to small insertions and deletions (indels)²⁷. However, small indels may be unable to abolish or affect the functions of CREs in many cases. Unlike Cas9, Cas12a induces staggered-end DSBs, which lead to much larger deletions^{28–31}. Hence, Cas12a seems to be a better tool than Cas9 to destroy CREs and render promoter editing effects. Furthermore, CRISPR–Cas12a only requires short CRISPR RNAs (crRNAs) for DNA targeting and is capable of self-processing a crRNA array due to its intrinsic RNase activity³². As highly multiplexable CRISPR–Cas12a systems have been established in plants^{30,33–37}, CRISPR–Cas12a represents a promising tool for promoter editing to introduce QTV in crops.

Here we developed an efficient promoter editing system, CRISPR–Cas12a promoter editing, henceforth referred to as CAPE. This CAPE system combines the estimation of key regions (KR)s of promoters and effective multiplexed Cas12a-mediated promoter editing to robustly create QTV in plants. As a proof of concept, we demonstrated the use of CAPE to efficiently generate QTV continuums in the starch content and grain size of rice by targeting the *OsGBSS1* (ref. 38) and *OsGS3* (ref. 39) promoters, respectively. Furthermore, by fine-tuning plant height through promoter editing of key genes in plant hormone biosynthesis pathways, we successfully generated promoter-edited crops with a desirable semidwarfism and antilodge trait without a substantial yield penalty. In the past, this Green Revolution trait in crops was achieved through the qualitative loss-of-function of *SD1* (also known as *GA20ox*)^{8,40,41}. Here we demonstrated CAPE as an efficient system for creating a quantitative Green Revolution trait and hence greatly expanded the molecular breeding resources.

Fig. 2 | Generation of a QTV continuum of the starch content of rice by promoter editing of *OsGBSS1*. **a**, Allele analysis of the *OsGBSS1* gene and its promoter for 4,594 cultivars from seven rice subspecies. All of the mutation sites recorded by MBKbase are marked with a dashed line. The stacked bar chart displays the usage preference of each allele in each rice population. The pie charts and the bottom bar plots show the usage ratio and their corresponding amylose content, respectively, for seven natural *Wx* alleles from Huang and collagues³⁸. The green, grey and blue boxes in the gene schematic indicate promoter, 5' UTR and CDS, respectively. The amylose content was tested in triplicate and the exact values are provided on the bar plots; data represent the mean \pm s.d. **b**, Feature and aggregate scores calculated by the estimating model. The KRs are shaded in grey. **c**, Schematic depicting the genotypes of 47 CRISPR–Cas12a-based *OsGBSS1* promoter editing (*OsGBSS1-PE*) lines along with the WT and null-mutant (*OsGBSS1-m01*) controls (left). The amylose content of all plants is provided (right); data represent the mean \pm s.d. of three biological

Results

Creating the CAPE system

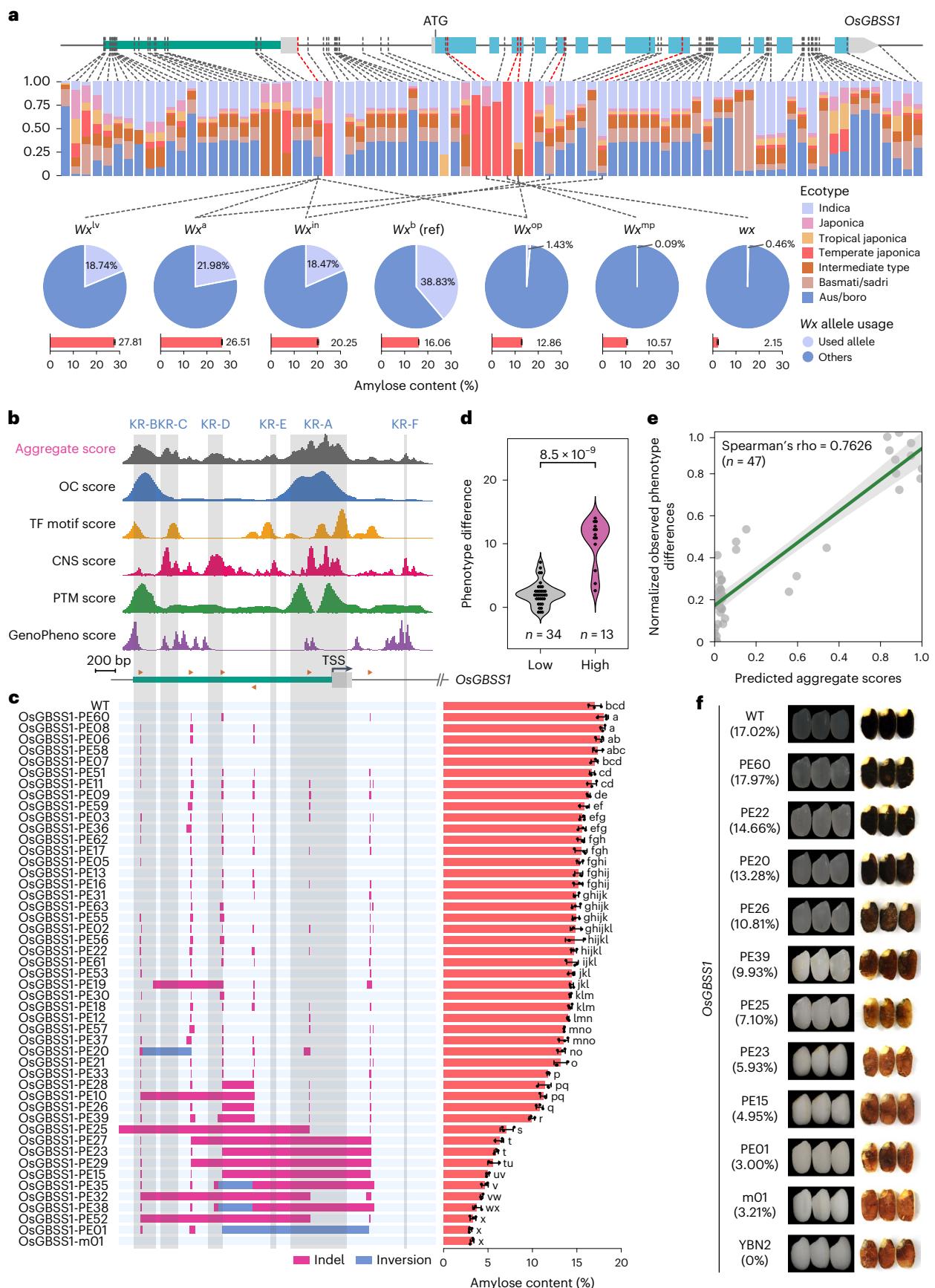
We aspired to have a biocomputational design tool to guide promoter editing. A major aspect of this design tool is to estimate the importance of different promoter subregions and their potential impact on gene expression. Thus, we developed an estimating model based on five genomic or epigenomic features that are associated with CREs—that is, chromatin accessibility, TF binding motifs, sequence conservation, histone modification (H3K27ac) and variation effects (Fig. 1a and Supplementary Fig. 1). The estimating model assigns scores to each feature (hereafter we refer to these feature scores as OC, TF motif, CNS, PTM and GenoPheno scores, respectively) within the promoter regions. All the feature scores were then aggregated with the respective weights to determine the putative influence of the regulatory regions on gene expression (Fig. 1b and Methods). We assumed that the aggregate scores are correlated with the activity of CREs, which play a key role in the regulation of target-gene expression: higher aggregate scores indicate more KRs that when deleted or mutated may give more pronounced phenotypes due to remarkable changes in gene expression (Fig. 1c). To validate our model, we analysed the recently published promoter editing data for *SIWOX9* in tomato⁴². We calculated and visualized the aggregate scores in the *SIWOX9* promoter (Supplementary Fig. 2a) and marked the genotypes of 33 promoter editing lines (Supplementary Fig. 2b). Promoter editing lines with mutations that highly overlapped with the KRs clearly had more significant phenotypes with increased branches compared with the wild-type (WT; Supplementary Fig. 2b). Our estimating model showed a strong predictive power as the phenotype is highly correlated with the genotype, and the phenotype difference among mutated regions with low and high aggregate scores was highly statistically significant (Supplementary Fig. 2c). We also analysed two promoter editing datasets in maize at *ZmCLE7* and *ZmFCP1* (ref. 7). As expected, mutations in regions with higher aggregate scores induced more phenotypic changes than mutations in regions with low aggregate scores for both the *ZmCLE7* (Supplementary Fig. 3a,b) and *ZmFCP1* mutants (Supplementary Fig. 3d,e). These results support the predictive power of our model, although statistically significant differences were not detected due to small sample sizes (Supplementary Fig. 3c,f).

Based on previous studies, Cas12a is potentially more advantageous than Cas9 for promoter editing given that it generates larger deletions and targets AT-rich sites (with the TTV protospacer adjacent motif (PAM) requirements) that are prevalent in promoters and KRs as estimated by our model (Fig. 1d and Supplementary Fig. 4). By combining promoter KR estimation and multiplexed Cas12a editing, we developed the CAPE system for efficient promoter editing in plants (Fig. 1). We set to comprehensively test and apply CAPE in rice, a major crop. To facilitate the use of the system, we predicted the KRs associated with gene expression in each promoter, which can be visualized in a genome browser. The crRNAs were pre-designed for every annotated gene in the rice genome and crRNAs targeting regions

replicates. Different letters indicate significant differences, determined using a one-way analysis of variance (ANOVA; $P < 0.05$; Duncan test). A schematic of the *OsGBSS1* gene is provided (top); the green segment indicates the promoter region and the orange arrowheads represent crRNAs carried by the pGEL590 vector. TSS, transcription start site. **d**, Comparison of phenotype changes after promoter editing of mutants with low or high aggregate scores. Phenotype changes were defined as the difference of the measured phenotype between the WT and mutant. Individual data points are marked as open circles; n , number of mutants. Significance analysis was done using a one-tailed Mann–Whitney *U*-test. **e**, Correlation between the predicted aggregate scores and the normalized phenotype differences. Phenotype differences were normalized using the *Z*-score method. The grey error band represents the 95% confidence interval. **f**, Images of different *OsGBSS1-PE* rice following milling (left) and grain starch iodine staining (right). WT, Nipponbare; YBN2, a local waxy rice variety. The percentages below the sample names represent the amylose content.

with different levels of aggregate scores were colour-coded for easy selection (Supplementary Fig. 1; <https://zhangtaolab.org/cape>). Meanwhile, we also developed an efficient CRISPR–Cas12a cloning

system for multiplexed editing in both dicotyledons and monocotyledons (Supplementary Fig. 5). In this study we included crRNAs targeting some promoter regions of lower aggregate scores so that we



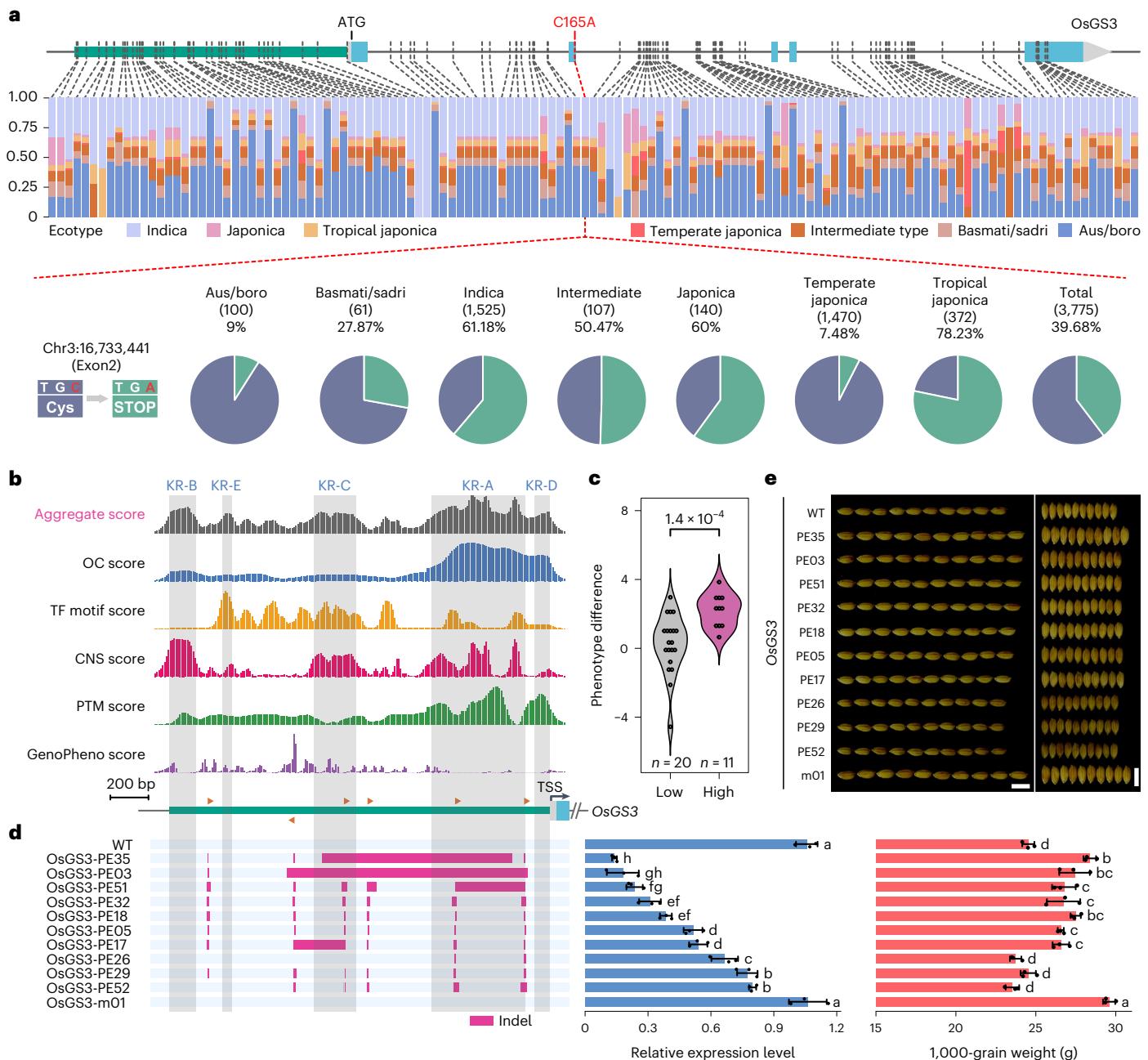


Fig. 3 | Generation of a QTV continuum of rice grain size by promoter editing of *OsGS3*. **a**, Allele analysis of the *OsGS3* gene and its promoter for 3,775 cultivars from seven rice subspecies. All of the mutation sites recorded by MBKbase are marked with a dashed line. The stacked bar charts display the usage preference of each allele in each rice population. The pie charts show the allele usages of a reported natural mutation (C165A; indicated in red on the schematic of the gene), which introduces premature transcription termination, in different rice populations. **b**, Feature and aggregate scores calculated by the estimating model. The KRs are shaded in grey. **c**, Comparison of phenotype changes in all mutants after promoter editing of mutants with low or high aggregate scores. Phenotype changes were defined as the difference of the measured phenotype between the WT and the mutant. Individual data points are marked as open circles; *n*, number

of mutants. Significance analysis was done using a one-tailed Mann–Whitney *U*-test. **d**, Schematic depicting the genotypes of ten CRISPR–Cas12a-based *OsGS3* promoter editing (*OsGS3-PE*) lines along with the WT and null-mutant controls (*OsGS3-m01*; left). The relative expression level of *OsGS3* (middle) and 1,000-grain weight (right) of all plants are provided; data represent the mean ± s.d. of three biological replicates. Different letters indicate significant differences, determined using a one-way ANOVA (*P* < 0.05; Duncan test). A schematic of the *OsGS3* gene is provided (top); the green segment indicates the promoter region and orange arrowheads represent crRNAs carried by the pGEL592 vector; the grey box indicates the transcription start site (TSS). **e**, Seed length (right) and width (right) of the *OsGS3-PE* mutants and controls. Scale bars, 1 cm.

could assess the design model in an unbiased way while generating QTV continuums.

Efficient creation of amylose content variation in rice

The cooking and taste characteristics of small grains such as rice, maize and wheat are highly influenced by their amylose content.

Granule-bound starch synthase I (*GBSS1*, also known as *Waxy/Wx*) is responsible for amylose biosynthesis in grains^{38,43}. We analysed the *OsGBSS1* locus of 4,594 cultivars from seven subspecies of rice and identified many natural mutations throughout the coding sequence and promoter region. However, among these abundant germplasm resources, only six natural variation mutations (five in the coding

sequence and one in the 5' untranslated region) controlling amylose content seem to be selected during rice breeding and the resulting amylose contents of these cultivars do not form a fine-scale QTV continuum (Fig. 2a). Our CAPE estimating model identified five *OsGBSS1* promoter KRAs with high aggregate scores above threshold in the *OsGBSS1* default promoter region (2,000 base pair (bp) upstream of the transcription start site). We ranked these KRAs from KR-A to KR-E based on their aggregate scores (Fig. 2b). We then conducted CAPE with six Cas12a crRNAs that are within or spanning these regions for multiplexed editing in the japonica rice cultivar Nipponbare, and one Cas9 construct was used to generate gene knockout (Supplementary Fig. 6a). Analysis of 39 T0 lines revealed editing at multiple target sites, with many being large deletions (Supplementary Fig. 6b,c). The editing frequencies of six crRNAs ranged from 59.0% to 94.9% (Supplementary Fig. 6b). We randomly tested 47 transgene-free homozygous T1 promoter editing (PE) offspring lines that carried deletions of variable sizes and occasionally chromosomal inversions, and these lines collectively confer a QTV continuum of amylose content from the WT level (approximately 17%) to the null-mutant level (approximately 3%; Fig. 2c). The gradual reduction of the amylose content in these lines is consistent with a reverse QTV continuum in gel consistency (Supplementary Fig. 7). Next, we analysed the association between our model and phenotype changes. As anticipated, more severe phenotypes (as reduced amylose content) were observed in lines with edited KRAs (Fig. 2d). For example, mutations in KR-A with the highest aggregate score (for example, *OsGBSS1*-PE15 to -PE01 editing lines) resulted in a large reduction in the amylose content. However, variations in KR-B, KR-C and KR-D (PE19, PE20 and PE10 editing lines) resulted in moderate phenotype changes, indicating that KR-A has a greater impact on the expression of *OsGBSS1*. These results suggested a strong predictive power of our model (Fig. 2e) as well as the ability of the CAPE system to efficiently generate QTV. Decreased levels of amylose in the selected promoter editing lines could be easily visualized in the rice grains, with and without iodine staining (Fig. 2f). The promoter editing lines with the highest reduction in amylose content had similar levels to that of the null mutant that was generated by Cas9 (Supplementary Fig. 6d), which was nearly comparable to YBN2, a local waxy rice cultivar (Fig. 2f).

We further characterized nine promoter editing lines and a null-mutant control (Supplementary Fig. 8a). The expression levels of *OsGBSS1* in the promoter editing lines (Supplementary Fig. 8b) were very consistent with the levels of amylose content (Supplementary Fig. 8c). By contrast, the null mutant had similar levels of *OsGBSS1* expression to the WT, despite the minimal amylose content (Supplementary Fig. 8b,c). Further analysis of two promoter editing lines (PE01 and PE20) demonstrated that the WT and null mutant had comparable phenotypes—plant height, length of the main panicle, seed morphology, seed length, seed width and 1,000-grain weight (Supplementary Fig. 8d–j). These data suggest

that fine-tuned levels of amylose in the promoter editing lines are not associated with any potential yield penalty. Hence, using CAPE, we could efficiently fine-tune the expression of *OsGBSS1* to generate a desirable QTV.

Generation of quantitative trait continuum for rice grain size

We next assessed CAPE for generating QTV in grain size by targeting *OsGS3* (refs. 39,44,45). Comparative analysis of 3,775 cultivars from seven rice subspecies revealed widespread sequence variation in the promoter and coding sequence of *OsGS3*. Strikingly, only a C-to-A nonsense mutation in the second exon has been selected by rice breeding (Fig. 3a), which is however consistent with the evolution history of *OsGS3* (ref. 46). Hence, generation of QTV in the grain size of rice cultivars by *OsGS3* promoter editing would drastically diversify the gene pool for variable grain sizes. To demonstrate this, we simultaneously targeted the promoter of *OsGS3* with six crRNAs (Fig. 3b) with a multiplexed Cas12a construct (Supplementary Fig. 9a). In parallel, a Cas9 construct was used to generate *OsGS3*-null mutants by editing the coding sequence (Supplementary Fig. 9a). Analysis of 34 T0 lines indicated that nearly all lines carried multiplexed edits and the editing efficiencies for the crRNAs ranged from 70.6% to 100% (Supplementary Fig. 9b). Transgene-free homozygous T1 lines were obtained, with large deletions across the target sites for promoter editing by Cas12a (Supplementary Fig. 9c) and the small deletion for the coding sequence editing by Cas9 (Supplementary Fig. 9d). All promoter editing lines carried different mutations in the two promoter regions with high aggregate scores (Supplementary Fig. 10a,b). A QTV continuum of 1,000-grain weight was observed among these lines (Supplementary Fig. 10b), and the clusters of low and high 1,000-grain weight measures were largely predictable based on the promoter editing genotypes (Fig. 3c and Supplementary Fig. 10c). A QTV continuum was also observed for the seed length (Supplementary Fig. 11a) and width (Supplementary Fig. 11b). We further analysed ten representative promoter editing lines and found an *OsGS3* expression continuum among them, which is consistent with the 1,000-grain weight phenotype (Fig. 3d) and grain size (Fig. 3e). Furthermore, analysis of two representative promoter editing lines showed no difference in plant morphology, height or the length of the main panicle between these lines and the WT or null mutant (Supplementary Fig. 11c–e). This work not only further validated the power of promoter editing to introduce QTV mutations in plants but also provided abundant new alleles with expression variation of *OsGS3*, which do not exist in the rice germplasms available at present.

Promoter editing of hormone genes for Green Revolution traits

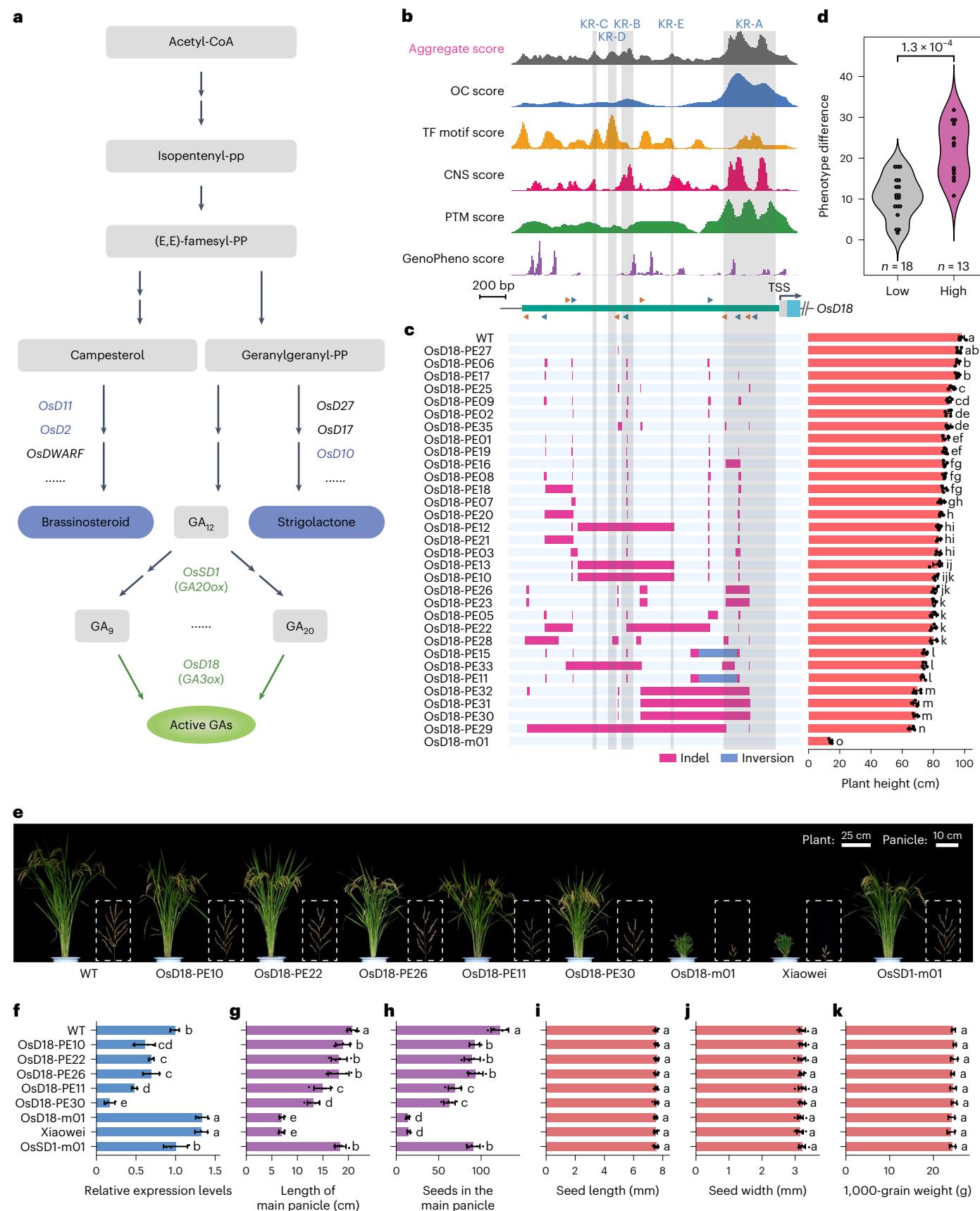
Semidwarf and antilodging traits are highly valuable in small grain crops, which historically contributed to yield increase since the Green Revolution^{8,40,47}. Plant height is controlled by several hormones, including brassinosteroid, gibberellin and strigolactone^{8,40,48}. We investigated

Fig. 4 | Generation of a QTV continuum of rice plant height by promoter editing of *OsD18*. **a**, Schematic depicting the synthesis pathways of brassinosteroid, strigolactone and gibberellin. The genes shown in blue and green text (*OsDII*, *OsD2*, *OsSD1*, *OsD18* and *OsD10*) were chosen for genome editing in this study. **b**, Feature and aggregate scores calculated by the estimating model. The KRAs are shaded in grey. **c**, Schematic depicting the genotypes of 31 CRISPR–Cas12a-based *OsD18* promoter editing (*OsD18*-PE) lines along with the WT and null-mutant (*OsD18*-m01) controls (left). Plant heights (right) are indicated; data represent mean \pm s.d. of five biological replicates. A schematic of the *OsD18* gene is provided (top); the green segment indicates the promoter region, and the blue and orange arrowheads represent crRNAs carried by the pGEL602 and pGEL603 vectors, respectively. **d**, Comparison of phenotype changes after promoter editing of mutants with low or high aggregate scores. Phenotype changes were defined as the difference of the measured phenotype

between the WT and mutant. Individual data points are marked as open circles; *n*, number of mutants. Significance analysis was done using a one-tailed Mann–Whitney *U*-test. **e**, Plant (left) and panicle (right; dashed-white box) phenotypes of *OsD18*-PE mutants and control plants grown in the green house. **f**, The relative expression levels of *OsD18* in *OsD18*-PE lines, *OsD18*-m01 mutant, *xiaowei*, *OsSD1*-m01 mutant and WT controls were measured by quantitative PCR with reverse transcription (RT–qPCR) with three biological and three technical replicates per biological sample. **g**, Main panicle length. Five panicles were measured per genotype. **h**, Seed number of main panicles. **g,h**, Data represent the mean \pm s.d. of five replicates. **i**, Seed length. **j**, Seed width. **i,j**, Data represent the mean \pm s.d. of ten biological replicates. **k**, 1,000-grain weight. Data represent the mean \pm s.d. of three biological replicates. **c,f–k**, Different letters indicate significant differences, determined using a one-way ANOVA ($P < 0.05$; Duncan test).

whether it is possible to explore promoter editing to generate novel semidwarf Green Revolution traits by targeting some of the biosynthesis genes for these hormones. We selected *OsD2* (ref. 49) and *OsD11*

(ref. 50) in the brassinosteroid biosynthesis pathway, *OsD10* (ref. 51) in the strigolactone biosynthesis pathway, and *OsSD1* (ref. 40) and *OsD18* (ref. 52) in the gibberellin biosynthesis pathway (Fig. 4a).



We first generated a *OsD2*-null mutant using Cas9 (Supplementary Fig. 12a,b) and observed dwarfism in both the seedling and mature plant stages (Supplementary Fig. 12c,d), resulting in severely reduced grain yield as there were few seeds per panicle (Supplementary Fig. 12e). Two CRISPR–Cas12a T-DNA vectors, each multiplexing six crRNAs (Supplementary Fig. 13a), were used to generate the promoter editing lines (Supplementary Fig. 13b,c). Expression analysis showed a gradual reduction of *OsD2* expression in these lines (Supplementary Fig. 13d), consistent with a gradual reduction of the height of these lines (Supplementary Fig. 13e) and their semidwarf phenotypes (Supplementary Fig. 13f). However, reduced grain-yield parameters—including length and seed number of the main panicle (Supplementary Fig. 13g,h), grain size (Supplementary Fig. 13i–k) and the 1,000-grain weight (Supplementary Fig. 13l)—were also observed in the promoter editing lines. Promoter editing of *OsD11* led to similar observations (Supplementary Fig. 14a–m). These results demonstrated that promoter editing of genes in the brassinosteroid biosynthesis pathway could generate semidwarf phenotypes, which are associated with a severe yield penalty.

We also used Cas9 to generate a null mutant of *OsD10* (Supplementary Fig. 15a,b). Severe dwarfism was observed in the mutant (Supplementary Fig. 15c,d). It had a very low seed setting rate (<10%) (Supplementary Fig. 15e) and a very high number of panicles per plant (approximately 18) compared with the WT (approximately four; Supplementary Fig. 15f). These data suggest that promoter editing of *OsD10* would probably generate semidwarf plants with yield penalties. Consequently, we did not pursue promoter editing of *OsD10*.

The well-known rice Green Revolution trait is based on knockout of *OsSD1*, a gene involved in gibberellin biosynthesis⁴⁰ (Fig. 4a). We applied CAPE to generate promoter editing lines and used Cas9 to generate knockout lines (Supplementary Fig. 16a–d). The two promoter editing lines resulted in intermediate plant phenotypes (Supplementary Fig. 16e) with a minor reduction in plant height (Supplementary Fig. 16f) and the length of the main panicle (Supplementary Fig. 16g). There were no differences in seed size (Supplementary Fig. 16h–j) or 1,000-grain weight (Supplementary Fig. 16k) between the promoter editing lines and the WT and null-mutant controls. However, because the *OsSD1*-null mutant showed moderate semidwarfism compared with the WT and has been widely accepted for efficient breeding practice, it is a reasonable assumption that there is less room to create a continuum in semidwarfism by editing the promoter of *OsSD1*.

By contrast, *OsD18* knockout resulted in severe dwarfism without compromising the grain size⁵², making it a promising target for promoter editing to realise a Green Revolution trait like *OsSD1* knockout. We generated two multiplexed CRISPR–Cas12a constructs, with each expressing six crRNAs for editing the promoter of *OsD18* (Supplementary Fig. 17a). A total of 31 T0 lines and their 31 transgene-free homozygous T1 offspring lines of both constructs were analysed and confirmed to carry multiplexed genome editing events (Supplementary Fig. 17b,c). These T1 homozygous lines collectively form a semidwarf continuum from 64.0 cm to 98.0 cm (compared to 96.5–101.0 cm for WT plants; Fig. 4b,c). Consistent with our earlier observations, editing of the KRPs with high aggregate scores resulted in more pronounced phenotypes (Fig. 4b,c), which again supported the predictive power of our model (Fig. 4d). We selected five independent promoter editing lines that represent a semidwarfism continuum on the more severe end (Fig. 4c) and compared them with the WT, a null *OsD18*-m01 mutant (Supplementary Fig. 17d), xiaowei (a previously reported *OsD18*-null mutant in a slightly different Nipponbare background)⁵² and a *OsSD1*-null mutant (Fig. 4e). As expected, expression analysis showed a reduction of *OsD18* in these promoter editing lines (Fig. 4f). Interestingly, *OsD18* expression was elevated in the knockout lines carrying coding sequence mutations (Fig. 4f), suggesting a transcription-feedback regulation. Intermediate phenotypes between the WT and the *OsD18* knockout mutant in terms of the length of the main panicle (Fig. 4g) and

seed number of the main panicle (Fig. 4h) were observed for the five promoter editing lines. However, all these lines had the same measures for seed length, seed width and the 1,000-grain weight (Fig. 4i–k and Supplementary Fig. 17e). Interestingly, all these measures indicated that the *OsD18* promoter editing lines display a very similar phenotype to *OsSD1* mutants, suggesting that promoter editing of *OsD18* probably leads to a Green Revolution trait.

Molecular characterization of *OsD18* promoter editing lines

We further investigated the molecular basis of three promoter editing lines (*OsD18*-PE11, -PE22 and -PE30) that represent gene expression and phenotype continuums (Fig. 4c–h). We examined different gibberellins in the whole plant of 80-d-old promoter editing lines and different controls (the WT, a *OsD18*-null mutant and a *OsSD1*-null mutant). We could detect nearly all bioactive gibberellins except gibberellin 1 (GA1; Fig. 5a–e). GA6 had the highest level among all detected gibberellins and its level was reduced to between one-third and half of the WT in the promoter editing lines as well as in the *OsD18*-null and *OsSD1*-null mutants (Fig. 5a). Interestingly, the levels of GA3 in all three promoter editing lines were significantly increased compared with the WT (Fig. 5b). By contrast, the null mutants of *OsD18* and *OsSD1* had substantially reduced GA3 levels (Fig. 5b). The promoter editing lines and the *OsD18*-null mutant had similar levels of GA5 to the WT but GA5 was significantly reduced in the *OsSD1*-null mutant (Fig. 5c). No difference was observed for GA4 in mutants and the WT (Fig. 5d). The null mutants of *OsD18* and *OsSD1* both had significantly reduced levels of GA7, whereas intermediate GA7 levels (between the WT and the null mutant) were observed for the *OsD18* promoter editing lines (Fig. 5e). The consistent measures of bioactive gibberellins among all three promoter editing lines suggest that the semidwarfism of these lines is probably fine-tuned by the level of bioactive gibberellins.

We next conducted transcriptome analysis by messenger RNA sequencing (mRNA-seq) using the whole plant (the aboveground part) of 80-d-old promoter editing lines and different controls (the WT, a *OsD18*-null mutant and a *OsSD1*-null mutant). Over 6,000 genes were differentially regulated in the *OsD18*-null mutant (Fig. 5f). The *OsD18*-PE30 line had similar numbers of differentially expressed genes (approximately 4,000) to the *OsSD1*-null mutant but more than the other two promoter editing lines (*OsD18*-PE11 and *OsD18*-PE22; Fig. 5f and Supplementary Table 1), consistent with the different *OsD18* expression levels in the promoter editing lines (Fig. 4f). Clustering analysis based on 44 gibberellin genes in a Kyoto Encyclopedia of Genes and Genomes (KEGG) signalling pathway analysis placed all three promoter editing lines in the same group as the *OsSD1*-null mutant, whereas the *OsD18*-null mutant fell outside this group (Fig. 5g and Supplementary Table 2), which is consistent with the phenotypes of these plants. Some plant height- and gibberellin pathway-related genes⁵³ were downregulated in the *OsD18*-null mutant and promoter editing lines (Supplementary Fig. 18). We also analysed the expression of 43 genes involved in gibberellin biosynthesis (Fig. 5h, Supplementary Fig. 19 and Supplementary Table 3). *OsD18* expression was substantially increased in the null mutant of *OsD18* and reduced in the *OsD18* promoter editing lines (Supplementary Fig. 19a), suggesting that a loss of functional *OsD18* protein results in strong activation of *OsD18* transcription. This strong transcription-regulation feedback was also observed for the gibberellin pathway genes upstream of *OsD18*—that is, seven of nine ent-kaurene synthase genes (Supplementary Fig. 19c), all four ent-kaurene oxidase genes (Supplementary Fig. 19d), two GA13-oxidase genes (Supplementary Fig. 19f) and three GA20-oxidase genes, including *OsSD1* (Supplementary Fig. 19g). By contrast, loss of functional *OsD18* in the null mutant resulted in reduced expression of many gibberellin pathway genes downstream of *OsD18*, including eight of 12 GA2-oxidase genes (Supplementary Fig. 19h). These negative- and positive-feedback regulations of gibberellin biosynthesis genes at the transcriptional level were only

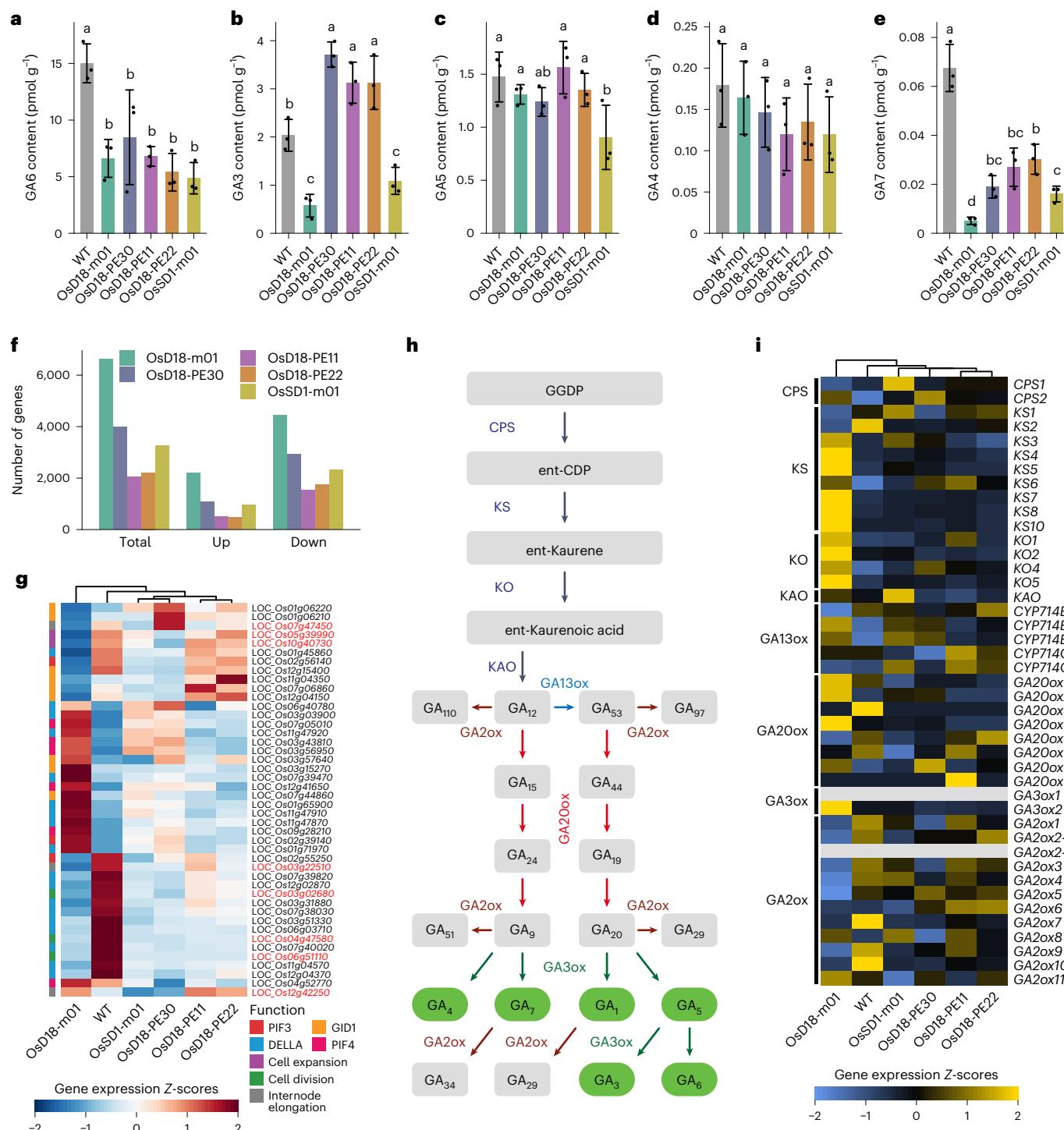


Fig. 5 | Comparison of the levels of active gibberellins and differentially expressed genes between three OsD18-PE lines and controls. a–e, Levels of active gibberellin for GA6, GA3, GA5, GA4 and GA7, respectively. Data represent the mean \pm s.d. of three biological replicates. Different letters indicate significant differences, determined using a one-way ANOVA ($P < 0.05$; Duncan test). **f**, Number of genes that are differentially expressed in the editing lines compared with the WT. Threshold value, fold change ≥ 2 and $P < 0.05$. **g**, Heat map showing the expression of gibberellin-response genes. Each row represents one gene. Lower levels of expression are represented in blue and higher expression in red. PIF3 and PIF4, phytochrome interacting factors; GID1, gibberellin-insensitive dwarf1. The LOC of gibberellin-response genes that control cell expansion/

division and internode elongation⁵³ is marked in red. LOC is used to indicate one rice gene in rice database (<http://rice.uga.edu/>). **h,i**, Schematic overview of gibberellin metabolism (**h**) alongside a heat map showing the expression of the relevant genes (**i**). Each row represents one gene. Lower levels of expression are represented in blue and higher expression in yellow. Genes that are not expressed are indicated in grey. GGDP, geranyl-geranyl diphosphate; CPS, copalyl-phosphate synthase; ent-CDP, ent-copalyl diphosphate; KS, ent-kaurene synthase; KO, ent-kaurene oxidase; KAO, ent-kaurenoic acid oxidase; GA13ox, GA13 oxidase; GA20ox, GA20 oxidase; GA3ox, GA3 oxidase; GA2ox, GA2 oxidase; and CYP714, cytochrome P714.

observed in the *OsD18*-null mutant, not in the *OsD18* promoter editing lines or the *OsSD1*-null mutant (Fig. 5*h* and Supplementary Fig. 19). KEGG analysis further supported this idea as the *OsD18* promoter editing lines were clustered with the *OsSD1*-null mutant and not with the *OsD18*-null mutant (Fig. 5*i*). Together, the gibberellin quantification and transcriptome analyses provide molecular insights on the semidwarfism of *OsD18* promoter lines, which phenocopy the *OsSD1*-null mutant.

Field trials of *OsD18* promoter editing rice in diverse backgrounds

The molecular analyses provided support for the idea that promoter editing of *OsD18* could generate the semidwarf and antilodging Green Revolution trait, as was done with *OsSD1* knockout⁴⁰. To this end, we conducted a field trial and compared the yield performance of the three *OsD18* promoter editing lines, alongside the Nipponbare WT and the *OsSD1*-null (*OsSD1*-m01) mutant. Mature *OsD18*-PE22 and *OsD18*-PE11 plants showed comparable height to *OsSD1*-m01, which was shorter than the WT but taller than the *OsD18*-PE30 plants (Fig. 6*a,b*). The same trend was observed for the number of grains per panicle (Fig. 6*c*). However, all promoter editing lines and *OsSD1*-m01 had higher numbers of panicles per plant than the WT (Fig. 6*d*). All lines had the same measures for seed length (Fig. 6*e*), seed width (Fig. 6*f*), 1,000-grain weight (Fig. 6*g*) and the seed setting rate (Fig. 6*h*). Consistent with these phenotypes, the *OsD18*-PE22 and *OsD18*-PE11 lines had the same field yield as *OsSD1*-m01 (approximately 0.75 kg m⁻²), which was slightly lower than the WT and slightly higher than *OsD18*-PE30, which had a more pronounced reduction of *OsD18* expression (Fig. 6*i*). The field trial data benchmarked some *OsD18* promoter editing lines (for example, *OsD18*-PE22 and PE11) whose field performance was comparable to Green Revolution *OsSD1*-null mutants.

However, the natural semidwarf phenotype of the Nipponbare cultivar had prevented us from investigating the full potential of introducing semidwarfism by promoter editing of *OsD18*. We sought to further validate this promoter editing strategy in two landrace waxy rice varieties, YBN2 (YiBin Nuo-2) and BBN1 (BeiBei Nuo-1). These two varieties have been cultivated in the Sichuan province of China for many years, with high-quality soft glutinous grains suitable for various uses such as steaming and brewing. However, mature plants of both varieties can reach a height of about 1.5 m, making them prone to lodging. The promoter of *OsD18* in YBN2 is very conserved compared with Nipponbare (Fig. 6*j*). We conducted promoter editing of *OsD18* in YBN2 and generated *OsD18*- and *OsSD1*-knockout lines in YBN2 as controls (Fig. 6*j* and Supplementary Fig. 20a–c). Three representative *OsD18* promoter editing lines with different degrees of semidwarfism were selected (Fig. 6*k*). As expected, these promoter editing lines had the same grain size as the WT as well as the *OsD18*-null and *OsSD1*-null mutants in the same background (Supplementary Fig. 20d). Similarly, we generated two *OsD18* promoter editing lines (Supplementary Fig. 21a), two *OsSD1*-knockout mutants (Supplementary Fig. 21b) and one *OsD18*-knockout mutant (Supplementary Fig. 21c) in the BBN1 background. As expected, similar to the *OsSD1*-knockout mutant, semidwarfism was observed in the *OsD18* promoter editing lines grown in a

growth chamber without compromising the grain quality and yield of the plants (Supplementary Fig. 21d–j).

For further testing in a field trial, we focused on the *OsD18* promoter editing lines and controls in the YBN2 background. In this field trial, most of the WT YBN2 plants showed the obvious lodging phenotype, whereas the *OsD18* promoter editing lines were lodging-resistant (Fig. 6*l*), which was attributed to their semidwarf statures resembling the *OsSD1* mutants (Fig. 6*m*). Although the WT YBN2 plants had more grains per panicle (Fig. 6*n*), the number of panicles per plant was significantly lower than the *OsD18* promoter editing lines and the *OsSD1* mutant (Fig. 6*o*). These lines had the same measures for seed length (Fig. 6*p*), seed width (Fig. 6*q*), 1,000-grain weight (Fig. 6*r*) and seed setting rate (Fig. 6*s*). Overall, all three promoter editing lines showed similar yield performance to the *OsSD1*-mutant plants (Fig. 6*t*). By contrast, the yield of the WT YBN2 plants was significantly reduced (Fig. 6*t*), largely attributed to the lodging problem. Together, we demonstrated that promoter editing of *OsD18* could generate QTV in semidwarfism without significant yield penalties. Due to its QTV nature, we term this as a quantitative Green Revolution trait.

Discussion

Traditional genetic knockout by targeting the coding sequences cannot generate continuous and tuneable gene expression. Similarly, deletion of an entire promoter usually results in damage to the three-dimensional genomic architecture and produces severe phenotype changes⁵⁴. It is difficult to use these methods to introduce QTV with practical values in crops. Earlier studies of *OsGBSS1* (Waxy) and *OsIPA1* promoter editing successfully generated rice seeds with a range of amylose content and enhanced productivity, respectively^{4–6}. However, these studies used CRISPR–Cas9 and did not evaluate the importance of different genomic regions before performing promoter editing, which explains their low efficacy in generating QTV in crops (Supplementary Table 4). By contrast, our CAPE system streamlines the promoter editing process to introduce QTV in crops and to greatly enrich the genetic resources for crop breeding. Our results on multiple target genes suggested a strong predictive power of the editing outcomes by our estimating model. CAPE uses an efficient Pol II promoter-based multiplexed Cas12a system^{28,29,34,36}, which presumably requires fewer guide RNAs in the KR of the promoter to efficiently introduce QTV, representing a powerful genome editing technology that is much more efficient than previous promoter editing systems (Supplementary Table 4). For example, with CAPE, we generated nearly complete QTV continuums of amylose content and grain sizes by editing *OsGBSS1* and *OsGS3*, respectively.

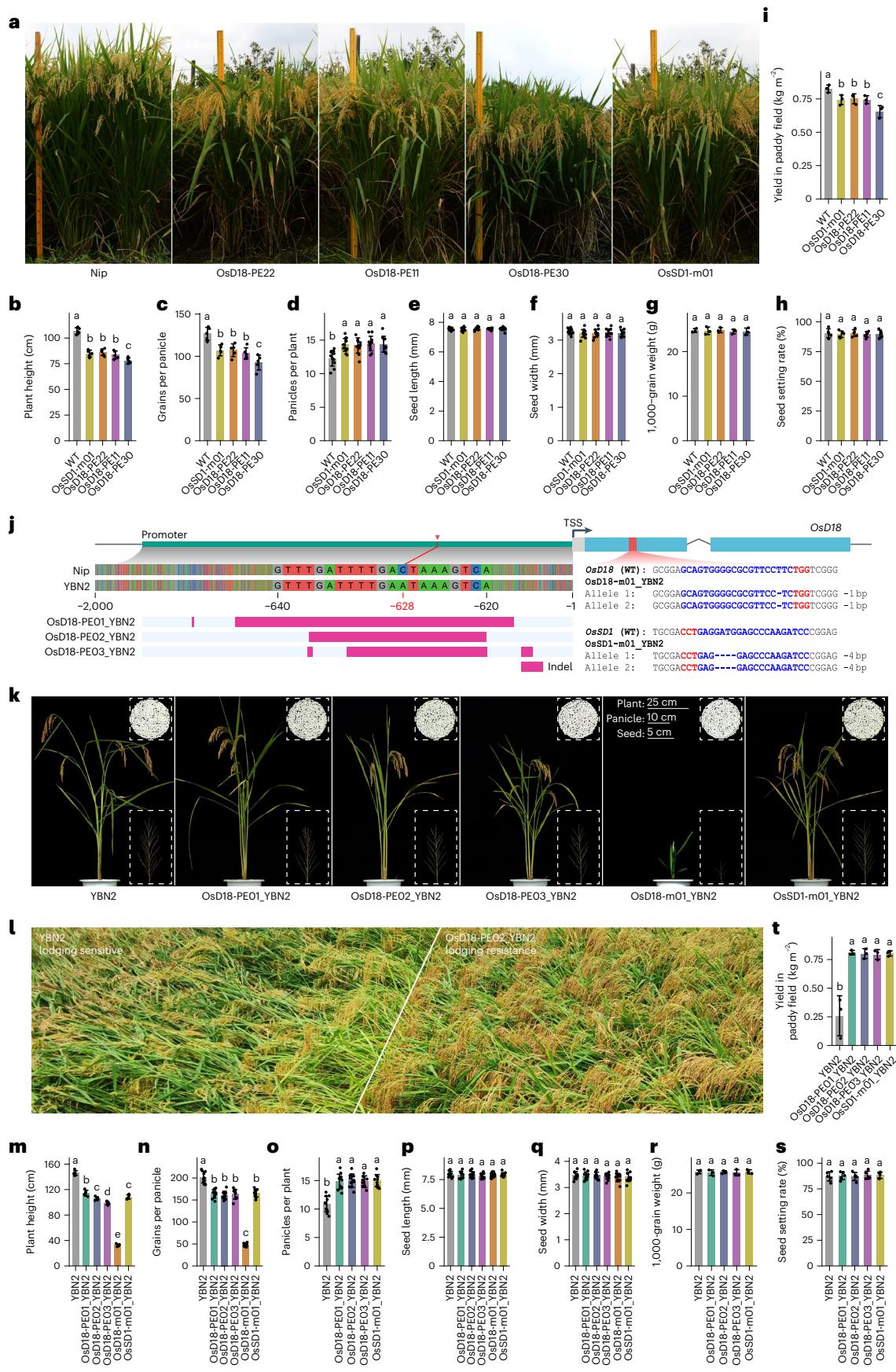
The semidwarf and antilodging traits of small grains enabled Green Revolution in agriculture^{8,40,41}. Semidwarfism without yield compromise is still considered to be one of the two most substantial agronomic traits that will benefit more crops beyond small grains, especially when introduced by efficient genome editing tools⁸. To assess whether promoter editing can be practically used to generate semidwarfism in rice, we applied CAPE to edit key enzyme genes in the biosynthesis pathways of three major plant hormones: brassinosteroid, strigolactone and gibberellin. Interestingly, although semidwarfism could be readily achieved by editing many genes in these pathways,

Fig. 6 | Field trials of *OsD18* promoter editing lines in different genetic backgrounds. **a–i**, Field trial of *OsD18*-PE and control lines in a Nipponbare (Nip) background. Mature plant phenotype (**a**), plant height (**b**; *n* = 5), grain number (**c**; *n* = 5), panicle number (**d**; *n* = 10), seed length, (**e**; *n* = 10), seed width (**f**; *n* = 10), 1,000-grain weight (**g**; *n* = 3), seed setting rate (**h**; *n* = 5) and yields in 2021 (**i**; *n* = 3 independent plots) were assessed. **j**, Schematic of *OsD18* sequence alignment in Nip and YBN2 along with three CRISPR–Cas12a-based *OsD18* promoter editing lines in YBN2 (left), and *OsD18*- and *OsSD1*-knockout mutants in YBN2 by CRISPR–Cas9 (right). The red arrowhead indicates a single nucleotide polymorphism in the *OsD18* promoters of the YBN2 genome (compared with Nip). Nucleotides in

blue indicate protospacers and those in red indicate PAM for Cas9-based editing (right). **k**, Key phenotypes—plant height (main image), architecture (bottom inset) and milled grains (top inset)—of *OsD18* promoter editing lines and control plants grown in a growth chamber. **l–t**, Field trial of *OsD18*-PE and control lines in a YBN2 background. Mature plant lodging/antilodging phenotypes (**l**), plant height (**m**; *n* = 5), grain number (**n**; *n* = 5), panicle number (**o**; *n* = 10), seed length (**p**; *n* = 10), seed width (**q**; *n* = 10), 1,000-grain weight (**r**; *n* = 3), seed setting rate (**s**; *n* = 5) and yields in 2021 (**t**; *n* = 3 independent plots) were assessed. **b–i,m–t**, Data represent the mean ± s.d.; different letters indicate significant differences, determined using a one-way ANOVA (*P* < 0.05; Duncan test).

only promoter editing of *OsD18* in the gibberellin pathway resulted in semidwarfism without significant yield penalty (Fig. 7), as proven by the field trials. The *OsD18*-null mutant had more severe dwarfism than

the *OsSD1*-null mutants, which provides a great opportunity to generate a much wider range of QTV continuum. For this reason, promoter editing of *OsD18* can generate a quantitative Green Revolution trait in



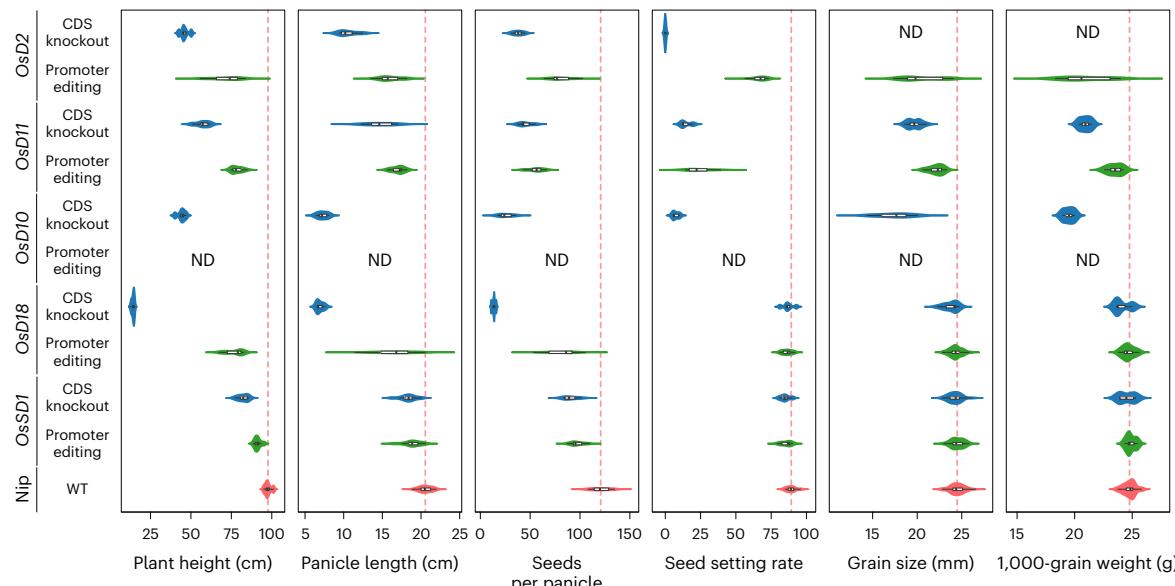


Fig. 7 | Summary of genome editing outcomes in this study towards engineering semidwarfism in rice. Comparison of rice agronomic traits influenced by coding sequence (CDS) knockout and promoter editing strategies. Biosynthesis genes for the plant hormones brassinosteroid, gibberellin and strigolactone were edited and key agronomic traits were measured in the resulting lines. ND, no data to test. For CDS knockout: *OsD2* mutants, $n = 5$ samples were measured for each of the first four traits; *OsD11* mutants, $n = 5, 5, 3, 10$ and 3 samples; *OsD10* and *OsD18* mutants, $n = 5, 5, 5, 10$ and 3 samples;

and *OsSD1* mutants, $n = 10, 10, 5, 5, 20$ and 6 samples. For promoter editing: *OsD2* mutants, $n = 15, 15, 15, 15, 30$ and 9 samples; *OsD11* mutants, $n = 10, 10, 10, 6, 20$ and 6 samples; *OsD18* mutants, $n = 25, 25, 25, 25, 50$ and 15 samples; and *OsSD1* mutants, $n = 10, 10, 10, 10, 20$ and 6 samples; promoter editing was not performed for *OsD10*. For the WT (Nip) plants, $n = 25, 20, 10, 8, 40$ and 12 samples. The middle line of the box plot represents the median. The box edges represent the first and third quartiles, and whiskers extend to the minimum and maximum. Red dashed lines represent the median of the traits in the WT plants.

rice (Fig. 7). *OsD18* is a GA3ox enzyme gene that is highly conserved in plants. Hence, we postulate that promoter editing of the GA3ox enzyme gene(s) is a promising approach to create QTV in crop statures that can be further selected for optimal performance based on different genetic backgrounds and environmental conditions or needs.

The estimating model in CAPE should inform users to conduct a more focused promoter editing experiment to have a predictable phenotype. To further test this, we targeted the KR-A (390 bp) of the highest aggregate score in the *OsD18* promoter with five crRNAs (Supplementary Fig. 22a) and obtained two homozygous T1 lines with multiplexed editing (Supplementary Fig. 22b). Both lines (*OsD18-fPE1* and *OsD18-fPE12*; group VI) indeed showed a strong and similar semi-dwarf phenotype, more so than the lines with multiplexed edits that largely resided outside KRs (group I) or large deletion lines that mostly impacted KRs of lower aggregate scores (group II; Supplementary Fig. 22c). These results demonstrate that desirable phenotypes can be efficiently obtained using the prediction model in CAPE. To further evaluate the prediction model and the nuclease choice in CAPE, we compared the *OsD18*-promoter-editing ability of Cas9 and Cas12a using singular guide RNA (sgRNA). KR-A and its adjacent non-key (NK) region were each targeted by several guide RNAs with both Cas9 and Cas12a at nearly overlapping targeting sites. For this experiment, we constructed four small promoter editing libraries (Cas9-KR, Cas12a-KR, Cas9-NK and Cas12a-NK). Among the resulting *OsD18* promoter editing lines, we found that editing events by Cas9 or Cas12a in NK regions achieved little change in the expression level of the target gene (Supplementary Fig. 23). On the contrary, promoter editing events with larger deletions in KR-A created by Cas12a showed significant reduction in the expression of *OsD18*. Given that the Cas9-edited lines all had small indels, they failed to affect *OsD18* expression despite KR-A being targeted (Supplementary Fig. 23). This further illustrates the importance of the prediction model and the use of Cas12a, not Cas9, in the CAPE system. Given the diverse spatiotemporal expression patterns of different genes, which largely reflect their molecular basis of pleiotropic

function, utilizing CAPE with a crRNA library as demonstrated here will undoubtedly enable more precise and efficient fine-tuning of gene expression.

Many plant genes are regulated in a spatiotemporal manner. Understanding the regulatory landscape of these genes requires an efficient tool to dissect CREs. We previously characterized three intronic enhancers that are critical for fine-tuning tissue- and development-specific expression of their cognate genes in *Arabidopsis*⁵⁵. CRISPR–Cas9-based promoter editing has been used to dissect the functions of CREs in tomato and other plants^{42,56}. Based on existing knowledge, we know that the epigenomic features of promoters—including multiple types of DNA, RNA and histone modifications—can observably affect the differential expression of genes, and these epigenetic modifications themselves have a certain degree of spatiotemporal dynamic changes. To avoid the influences of these dynamic epigenomic features, we also added genetic information such as sequence conservation and TF binding motifs in CAPE. In this study our CAPE pipeline successfully estimated the impacts of the KRs of promoters on gene expression through the use of both genetic information and epigenetic data from leaf tissue. It is worth noting that the three edited rice genes, *OsGBSSI*, *OsGS3* and *OsD18*, have notable spatiotemporal expression patterns and they each regulate or form part of different molecular metabolic pathways to affect various biological traits. However, the KRs in the promoters of all these genes were still clearly detected by CAPE and editing of these regions efficiently induced the novel phenotype variations. Thus, these results could reliably demonstrate the effectiveness and reliability of the CAPE pipeline developed here. We note the current CAPE pipeline may have certain limitations in providing comprehensive prediction and understanding of the functional roles for each target-gene promoter. We anticipate further optimization of the precision, scope of application and effectiveness of the CAPE system in the future. This would be aided with the availability of big data on spatiotemporal epigenetic modifications, differential expression patterns, genome diversities and phenomics information in diverse

plants. We envision our CAPE system will greatly enable the dissection of CREs in plants.

In conclusion, we developed a powerful CAPE system for engineering QTV in crops. With CAPE, we efficiently introduced QTV continuums in the starch content, grain size and plant height of rice. Thus, this study not only provided a technological breakthrough for promoter editing but also further validated the power of promoter editing to introduce agriculture-relevant QTV in crops^{2,742}. We hope this CAPE system, when applied in more crops, will greatly aid crop breeding by efficiently producing genetic variation in elite germplasms. Notably, although the demonstrations were done in rice, we believe that the quantitative Green Revolution trait engineering by promoter editing of a key GA3ox gene is applicable to many other crops.

Methods

Plant materials and field trials

The rice cultivar Nipponbare (*Oryza sativa* L. japonica) was used as the WT control and transformation host for most of the experiments. The landrace waxy rice varieties YBN2 (local name: YiBin Nuo-2) and BBN1 (local name: BeiBei Nuo-1; *Oryza sativa* L. japonica) used for *OsD18* editing are from YiBin and MianYang, Sichuan province in China, respectively. The field trials of the *OsD18*-PE and control lines in the Nipponbare background were conducted at Chengdu, Sichuan province (30° 43' N, 103° 52' E) and the field trials of *OsD18*-PE and control lines in the YBN2 background were conducted at MianYang, Sichuan province (31° 30' N, 104° 25' E) in 2021 in natural environments.

Estimating model construction for promoter editing in rice

We generated an estimating model to quantify the potential contribution of different regions of a gene promoter to gene expression. Five genetic or epigenetic datasets—open chromatin (DNase I hypersensitive sites sequencing, DNase-seq), TF binding sites (data from JASPAR⁵⁷), conserved noncoding sequences (data from PlantRegMap⁵⁸), H3K27ac histone modification (chromatin immunoprecipitation with sequencing, ChIP-seq) and variation effects (data from MBKbase⁵⁹)—were selected as features to build the model.

First, we mapped all of the data to the rice reference genome⁶⁰. For the DNase-seq and ChIP-seq, the raw data were trimmed using cutadapt v3.5 (ref. 61). For the DNase-seq, clean reads were aligned to the rice reference genome using bowtie v1.2 (bowtie -n 0 -m 1 -best -strata)⁶². Popera (<https://github.com/forrestzhang/Popera>) was used to call DNase I hypersensitive sites and generate the coverage tracks in bigWig format⁶³. For the ChIP-seq, clean reads were mapped to the reference genome using bowtie2 v2.3.2 with default parameters⁶⁴. High-quality reads were retained to generate the coverage using bamCoverage from deepTools v3.5.1 (bamCoverage -bs 5 –normalizeUsingRPKM –skipNAs –minMappingQuality 15)⁶⁵.

Genome-wide prediction of TF binding sites in rice was conducted using FIMO v4.11.2 (ref. 66) based on the non-redundant position frequency matrix data from the JASPAR 2020 Core Plants collections. Binding sites with $P < 1 \times 10^{-5}$ and β -value < 1 were kept for further analysis. We downloaded the precalculated sequence conservation scores (PhastCons) from the PlantRegMap database. The bedGraphToBigWig script was then used to convert the format of the score file.

To quantify the putative effects of single nucleotide polymorphisms/indels on the phenotypes in the rice genome, we retrieved the promoter regions of the target genes from all the genotypes in the MBKbase database. Haplotypes were detected using the Biopython package⁶⁷. We also collected the phenotype information associated with the target genes⁵⁹. To connect the haplotypes with phenotype changes, we calculated the differences of the phenotype values between the reference haplotype (WT) and other haplotypes. For each mutation site among the haplotypes, we gathered all of the phenotype differences and performed a Kruskal–Wallis *H*-test based on the Python library

SciPy. The $-\log(P\text{ value})$ of the *H*-test was assigned to each mutation, which represents the putative effect on the corresponding phenotype.

Finally, we panned all five features to the promoter of the target gene using BEDTools⁶⁸. Each promoter was first split into 10-bp bins and then scores of each feature within the bins were averaged using the following formula:

$$AS_i = \frac{\omega_1 \times O_i + \omega_2 \times M_i + \omega_3 \times C_i + \omega_4 \times H_i + \omega_5 \times G_i}{\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5}$$

where AS means an aggregate score that combines all scores from the five selected features—chromatin accessibility (O), TF binding motifs (M), sequence conservation (C), histone modification (H) and variation effects (G)—and i represents each bin in a promoter. Weight coefficients ω_1 – ω_5 (1, 0.8, 1.2, 0.4 and 0.2) for the five features are determined empirically and evaluated using the non-negative linear regression model from scikit-learn. Among the five features, chromatin accessibility and sequence conservation are the most important features, followed by TF binding motifs. Chromatin accessibility is used to discover the dynamic regulatory elements. Sequence conservation is used to find all potential noncoding regulatory sites regardless of the tissue or conditions. TF binding motifs make the detection of regulatory sites more refined. H3K27ac histone modification facilitates the identification of active regulatory regions. Variation effects are used to further detect the key sites that may affect gene expression. These two features have relatively lower weight coefficients. Aggregate scores in a promoter are divided by the maximum aggregate score to generate normalized scores between zero and one. We also define KRs where bins with aggregate scores are greater than the average aggregate score in the promoter. These regions may be more relevant to gene expression. Based on the model, we performed the analysis of all of the gene promoters in the rice genome. The instructions of feature data processing and the complete pipeline of CAPE for calculating aggregate scores have been deposited to GitHub (<https://github.com/zhangtaolab/CAPE>). An informative genome browser with defined KRs and pre-designed guide RNAs^{69–71} was also provided on the basis of JBrowse2 (ref. 72) (<https://zhangtaolab.org/cape>).

Validating the strength of the estimating model

Three rice genes, two maize genes and one tomato gene were used to validate the estimating model. The rice editing lines were generated in this study. Maize data were retrieved from ref. 7 and tomato data were retrieved from ref. 42. For the maize and tomato data analyses, BWA v0.7.17 (ref. 73) was used to map all assay for transposase-accessible chromatin using sequencing and ChIP-seq reads to the respective reference genomes^{74,75}. The other procedures were the same as the pipeline in rice. The genetic variation feature was dropped when calculating aggregate scores.

In all of the promoter editing mutants derived from a gene, we first summed the aggregate scores overlapping the editing sites in each mutant and then calculated a cutoff to classify the mutants into two groups (low editing and high editing) based on the average aggregate score among all mutants. Next, we counted the phenotype differences between the mutant and WT plants in the two groups. Finally, a significance analysis was performed using a one-tailed Mann–Whitney *U*-test. In addition, Spearman's rank correlation coefficient was performed to examine the relationship between the predicted aggregate scores and observed phenotype changes.

Vector construction

For the Cas9 editing vectors, the sgRNA was cloned into the expression backbone vector pGEL026 according to the previously established method^{76,77}. The individual annealed sgRNA oligonucleotide pair (Oligo1 and Oligo2) was cloned into the region between the *OsU6* promoter and the sgRNA scaffold⁷⁸. The sgRNA oligonucleotide pairs are listed in Supplementary Table 5. A CRISPR–Cas12a toolkit was

developed by Gibson and Golden Gate assembly for dicotyledon and monocotyledon plants. Those vectors were derived from pGEL589 (refs. 29,36). To clone one or two crRNAs into the Cas12a vector, two synthesised oligonucleotides were annealed to generate a crRNA(s) fragment. The annealed mixture was diluted 100× with autoclaved double-distilled H₂O and then cloned into the Cas12a vector via Golden Gate assembly. To clone more than two crRNAs into the Cas12a vector, the crRNA array sequence was split into fragments at the crRNA sequences using the NEBridge SplitSet tool; fragments were generated by annealing two synthesised oligonucleotides. The adaptor contains a BsaI recognition sequence. The subclone sequence was then cloned into the expression backbone vector pGEL589 according to Golden Gate reaction³⁴. The vector pGEL589 contains two expression cassettes: one cassette is the *LbCas12a* gene under a *maize ubiquitin 1* promoter (pZmUbi1) and the other cassette is the crRNA array under a *rice ubiquitin 1* promoter (pOsUbi1). The crRNA expression cassette was inserted between HH-DR and HDV sites with BsaI to replace a *ccdB* selection marker. All crRNAs and subclone sequences used in this study are listed in Supplementary Table 5. All of the T-DNA constructs have been submitted to Addgene, under the IDs 195595–195606.

Stable transformation of rice

Stable transformation of rice was carried out according to a previously published protocol^{9,45}. Briefly, dehulled rice seeds were sterilized and then cultured on N6D solid medium. The precultured calli were transformed by inoculation with *Agrobacterium* EHA105 carrying the recombinant vector. The inoculated calli were co-cultured with *Agrobacterium* for 3 d, washed and moved to selection medium containing 50 mg l⁻¹ hygromycin and 400 mg l⁻¹ carbenicillin. After 3 weeks actively growing calli were moved to regeneration medium I to induce shoot growth. Small shoots were further transferred to regeneration medium II to induce root growth. The resulting transgenic seedlings were then transplanted into soil.

Targeted mutation detection

Genomic DNA was extracted from transgenic plants using a modified CTAB method⁷⁸. Genomic regions of targeted sites were amplified with specific primers (Supplementary Table 5) and the PCR products were analysed by single-strand conformation polymorphism analysis^{76,79}. Briefly, the targeted genomic regions of T0 lines were amplified, and the PCR products amplicons were denatured for 5 min at 95 °C and immediately put on ice to minimize self-annealing. The denatured PCR amplicons were electrophoresed on 15% non-denaturing polyacrylamide gels at 45 mA, 120–200 V. After about 6 h, the polyacrylamide gels were stained using argentation for photographing. The T1 homozygous mutant lines were further genotyped through Sanger sequencing.

Characterization of promoter editing mutant alleles

To clearly display the alleles of rice mutants, the allele visualization was done using custom R and Python scripts. The indel, inversion of each allele and crRNA target sites were plotted together. Aggregate scores along with the other five features scores were displayed to show the contributions of each feature on the prediction.

mRNA transcriptome sequencing and analysis

The T2 stable mutants without T-DNA and WT plants were chosen for mRNA transcriptome sequencing or mRNA-seq⁸⁰. Whole plants (80 d old) grown in the experimental paddy field were collected. Three independent plants for each mutant were chosen for library construction, sequencing and analysis. An Illumina HiSeq 2500 platform was used for mRNA transcriptome sequencing at Biomarker Technologies Co. Ltd. (China). Data processing and analysis were applied using the BMKCloud service (<http://www.biocloud.net/>). The mRNA-seq data of WT plants and other mutants were first aligned to the reference genome using HISAT2 v2.0.4 (ref. 81). The abundance of each gene was then calculated

using StringTie v2.1.7 (ref. 82). Differential expression analysis was performed using the edgeR package⁸³. Differentially expressed genes in hormone pathways were searched in the KEGG database (<http://www.genome.jp/kegg/>). The R package pheatmap was used to plot the heat map of hormone pathway-related gene expression in WT plants, null mutants and promoter editing plants.

RNA extraction and RT-qPCR

For *OsGBSS1* and *OsGS3*, the panicles of promoter editing plants at 7 d post flowering were collected. For *OsD2*, *OsD11* and *OsD18*, aboveground tissues of 80-d-old promoter editing plants were collected. For *OsD18* promoter editing lines with single sgRNA or crRNA, the resistant callus was collected. RNA extraction and RT-qPCR were carried out as described previously^{84,85}. Briefly, total RNA was extracted using TRIzol universal reagent (Tiangen), treated with DNase I and then used for complementary DNA synthesis. Reverse transcription was carried out using HiScript III RT SuperMix for qPCR (Vazyme) and RT-qPCR was performed with ChamQuniversal SYBR qPCR master mix (Vazyme) according to the manufacturer's instructions. Actin mRNA was used as an internal control. The relative levels of gene expression were calculated using the 2^{-ΔΔCt} method. Three biological replicates (three independent mutant seedlings) were examined to ensure reproducibility. The experiments were performed three times independently with similar results.

Grain amylose content and gel consistency measurements

After plant maturation, the seeds were harvested, dried and stored at room temperature. To determine the amylose content and gel consistency of the flours and starches of milled rice, all tests were performed in triplicate. The total starch, amylose content and gel consistency were analysed by Shanghai SANSHU Biotechnologies Co. Ltd, which specializes in plant testing services⁸⁶. For statistical analysis, data were subjected to a Duncan test analysis using the IBM SPSS Statistics 20.0 software program. Results with a probability value of *P* < 0.05 were considered statistically significant.

Iodine staining of endosperm

The hull of rice seeds was removed to observe the external appearance of the grains. The grains were cut through the centre to expose the endosperm. Iodine reagent (0.2%) was dropped on the endosperm surface and photographs were taken after 3–5 min (ref. 87).

Gibberellin content measurement

Whole 80-d-old T2 mutants and WT plants (aboveground tissues) grown in the experimental paddy field were collected. Three independent plants from each condition were chosen for gibberellin measurements. The gibberellin content was determined by high-resolution liquid chromatography and high-resolution mass spectrometry at Shanghai SANSHU Biotechnologies Co. Ltd.

Reporting summary

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

Data availability

Sequence conservation data were downloaded from PlantRegMap. The TF binding motif datasets were downloaded from JASPAR (2020 Core Plants collections). Genetic variations and phenotype data of rice were downloaded from MBKbase. The rice reference genome (MSU/Tigr7) was downloaded from the Rice Genome Annotation Project. The tomato reference genome (SL4.0) was downloaded from the Solanaceae Genomics Network. The maize reference genome (AGPv4) was downloaded from MaizeGDB. Public open chromatin and histone modification datasets were retrieved from the NCBI under the accession numbers GSE26610 and GSM2084219 for rice, GSE164297 and PRJNA381300 for tomato, and PRJNA599454 and PRJNA417726 for maize.

Code availability

The instructions of feature data processing and the complete pipeline of CAPE for calculating aggregate scores and detecting KRs have been deposited to GitHub (<https://github.com/zhangtaolab/CAPE>).

References

1. Hickey, L. T. et al. Breeding crops to feed 10 billion. *Nat. Biotechnol.* **37**, 744–754 (2019).
2. Rodriguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E. & Lippman, Z. B. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **171**, 470–480 (2017).
3. Cui, Y. et al. Production of novel beneficial alleles of a rice yield-related QTL by CRISPR/Cas9. *Plant Biotechnol. J.* **18**, 1987–1989 (2020).
4. Zeng, D. et al. Quantitative regulation of Waxy expression by CRISPR/Cas9-based promoter and 5'UTR-intron editing improves grain quality in rice. *Plant Biotechnol. J.* **18**, 2385–2387 (2020).
5. Huang, L. et al. Creating novel *Wx* alleles with fine-tuned amylose levels and improved grain quality in rice by promoter editing using CRISPR/Cas9 system. *Plant Biotechnol. J.* **18**, 2164–2166 (2020).
6. Song, X. et al. Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. *Nat. Biotechnol.* **40**, 1403–1411 (2022).
7. Liu, L. et al. Enhancing grain-yield-related traits by CRISPR-Cas9 promoter editing of maize *CLE* genes. *Nat. Plants* **7**, 287–294 (2021).
8. Eshed, Y. & Lippman, Z. B. Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science* **366**, eaax0025 (2019).
9. Ren, Q. et al. PAM-less plant genome editing using a CRISPR-SpRY toolbox. *Nat. Plants* **7**, 25–33 (2021).
10. Liu, S. et al. Hypercompact CRISPR-Cas12j2 (CasPhi) enables genome editing, gene activation, and epigenome editing in plants. *Plant Commun.* **3**, 100453 (2022).
11. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* **21**, 71–87 (2020).
12. Hu, G. et al. The auxin-responsive transcription factor SIDOF9 regulates inflorescence and flower development in tomato. *Nat. Plants* **8**, 419–433 (2022).
13. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
14. Zhao, H. et al. Genome-wide MNase hypersensitivity assay unveils distinct classes of open chromatin associated with H3K27me3 and DNA methylation in *Arabidopsis thaliana*. *Genome Biol.* **21**, 24 (2020).
15. Chen, A. F. et al. NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat. Methods* **19**, 547–553 (2022).
16. Lu, Z. et al. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* **5**, 1250–1259 (2019).
17. Deplancke, B., Alpern, D. & Gardeux, V. The genetics of transcription factor DNA binding variation. *Cell* **166**, 538–554 (2016).
18. Xiao, Q. et al. The landscape of promoter-centred RNA-DNA interactions in rice. *Nat. Plants* **8**, 157–170 (2022).
19. Song, B. et al. Conserved noncoding sequences provide insights into regulatory sequence and loss of gene expression in maize. *Genome Res.* **31**, 1245–1257 (2021).
20. Sackton, T. B. et al. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
21. Parvathaneni, R. K. et al. The regulatory landscape of early maize inflorescence development. *Genome Biol.* **21**, 165 (2020).
22. Kremling, K. A. G. et al. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**, 520–523 (2018).
23. Bozhilov, Y. K. et al. A gain-of-function single nucleotide variant creates a new promoter which acts as an orientation-dependent enhancer-blocker. *Nat. Commun.* **12**, 3806 (2021).
24. Zhu, C. et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat. Struct. Mol. Biol.* **26**, 1063–1070 (2019).
25. Zhang, W., Zhang, T., Wu, Y. & Jiang, J. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* **24**, 2719–2731 (2012).
26. Tao, S., Lin, K., Zhu, Q. & Zhang, W. MH-seq for functional characterization of open chromatin in plants. *Trends Plant Sci.* **25**, 618–619 (2020).
27. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR–Cas9. *Science* **346**, 1258096 (2014).
28. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
29. Tang, X. et al. A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* **3**, 17018 (2017).
30. Zhong, Z. et al. Plant genome editing using FnCpf1 and LbCpf1 nucleases at redefined and altered PAM sites. *Mol. Plant* **11**, 999–1002 (2018).
31. Tang, X. et al. A single transcript CRISPR–Cas9 system for efficient genome editing in plants. *Mol. Plant* **9**, 1088–1091 (2016).
32. Zetsche, B. et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
33. Wang, M., Mao, Y., Lu, Y., Tao, X. & Zhu, J. K. Multiplex gene editing in rice using the CRISPR–Cpf1 system. *Mol. Plant* **10**, 1011–1013 (2017).
34. Zhang, Y. et al. Expanding the scope of plant genome engineering with Cas12a orthologs and highly multiplexable editing systems. *Nat. Commun.* **12**, 1944 (2021).
35. Tang, X. et al. Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. *Plant Biotechnol. J.* **17**, 1431–1445 (2019).
36. Malzahn, A. A. et al. Application of CRISPR–Cas12a temperature sensitivity for improved genome editing in rice, maize, and *Arabidopsis*. *BMC Biol.* **17**, 9 (2019).
37. Zhong, Z. et al. Intron-based single transcript unit CRISPR systems for plant genome editing. *Rice* **13**, 8 (2020).
38. Huang, L., Sreenivasulu, N. & Liu, Q. Waxy editing: old meets new. *Trends Plant Sci.* **25**, 963–966 (2020).
39. Fan, C. et al. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–1171 (2006).
40. Sasaki, A. et al. Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* **416**, 701–702 (2002).
41. Spielmeyer, W., Ellis, M. H. & Chandler, P. M. Semidwarf (sd-1), “Green Revolution” rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl Acad. Sci. USA* **99**, 9043–9048 (2002).
42. Hendelman, A. et al. Conserved pleiotropy of an ancient plant homeobox gene uncovered by *cis*-regulatory dissection. *Cell* **184**, 1724–1739 (2021).

43. Tian, Z. et al. Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc. Natl Acad. Sci. USA* **106**, 21760–21765 (2009).
44. Mao, H. et al. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl Acad. Sci. USA* **107**, 19579–19584 (2010).
45. Zhou, J. et al. Multiplex QTL editing of grain-related genes improves yield in elite rice varieties. *Plant Cell Rep.* **38**, 475–485 (2019).
46. Takano-Kai, N. et al. Evolutionary history of GS3, a gene conferring grain length in rice. *Genetics* **182**, 1323–1334 (2009).
47. Peng, J. R. et al. ‘Green revolution’ genes encode mutant gibberellin response modulators. *Nature* **400**, 256–261 (1999).
48. Wang, Y. et al. A strigolactone biosynthesis gene contributed to the Green Revolution in rice. *Mol. Plant* **13**, 923–932 (2020).
49. Fang, N. et al. SMALL GRAIN 11 controls grain size, grain number and grain yield in rice. *Rice* **9**, 64 (2016).
50. Wu, Y. et al. CLUSTERED PRIMARY BRANCH 1, a new allele of DWARF11, controls panicle architecture and seed size in rice. *Plant Biotechnol. J.* **14**, 377–386 (2016).
51. Arite, T. et al. DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud outgrowth in rice. *Plant J.* **51**, 1019–1029 (2007).
52. Hu, S. et al. Xiaowei, a new rice germplasm for large-scale indoor research. *Mol. Plant* **11**, 1418–1420 (2018).
53. Nagai, K. et al. Antagonistic regulation of the gibberellic acid response during stem growth in rice. *Nature* **584**, 109–114 (2020).
54. Spielmann, M., Lupianez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
55. Meng, F. L. et al. Genomic editing of intronic enhancers unveils their role in fine-tuning tissue-specific gene expression in *Arabidopsis thaliana*. *Plant Cell* **33**, 1997–2014 (2021).
56. Wang, X. et al. Dissecting *cis*-regulatory control of quantitative trait variation in a plant stem cell circuit. *Nat. Plants* **7**, 419–427 (2021).
57. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
58. Tian, F., Yang, D. C., Meng, Y. Q., Jin, J. & Gao, G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* **48**, D1104–D1113 (2020).
59. Peng, H. et al. MBKbase for rice: an integrated omics knowledgebase for molecular breeding in rice. *Nucleic Acids Res.* **48**, D1085–D1092 (2020).
60. Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4 (2013).
61. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 11–12 (2011).
62. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
63. Zhang, T., Marand, A. P. & Jiang, J. PlantDHS: a database for DNase I hypersensitive sites in plants. *Nucleic Acids Res.* **44**, D1148–D1153 (2016).
64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
65. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
66. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
67. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
68. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
69. Wu, Y. et al. CRISPR-BETS: a base-editing design tool for generating stop codons. *Plant Biotechnol. J.* **20**, 499–510 (2022).
70. 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).
71. Wu, Y. et al. Genome-wide analyses of PAM-relaxed Cas9 genome editors reveal substantial off-target effects by ABE8e in rice. *Plant Biotechnol. J.* **20**, 1670–1682 (2022).
72. Buels, R. et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).
73. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
74. Hosmani, P. S. et al. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. Preprint at *bioRxiv* <https://doi.org/10.1101/767764> (2019).
75. Jiao, Y. et al. Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).
76. Zhou, J. et al. CRISPR–Cas9 based genome editing reveals new insights into microRNA function and regulation in rice. *Front. Plant Sci.* **8**, 1598 (2017).
77. Zhong, Z. et al. Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. *Mol. Plant* **12**, 1027–1036 (2019).
78. Ren, Q. et al. Improved plant cytosine base editors with high editing activity, purity, and specificity. *Plant Biotechnol. J.* **19**, 2052–2068 (2021).
79. Zheng, X. et al. Effective screen of CRISPR/Cas9-induced mutants in rice by single-strand conformation polymorphism. *Plant Cell Rep.* **35**, 1545–1554 (2016).
80. Wang, B. et al. Targeted mutagenesis of NAC transcription factor gene, OsNAC041, leading to salt sensitivity in rice. *Rice Sci.* **26**, 98–108 (2019).
81. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
82. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
83. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
84. Zhou, J. et al. CRISPR–Cas9 mediated OsMIR168a knockout reveals its pleiotropy in rice. *Plant Biotechnol. J.* **20**, 310–322 (2022).
85. Zhou, J. et al. Efficient deletion of multiple circle RNA loci by CRISPR–Cas9 reveals Os06circ02797 as a putative sponge for OsMIR408 in rice. *Plant Biotechnol. J.* **19**, 1240–1252 (2021).
86. Zheng, X. et al. Loss-function mutants of OsCKX gene family based on CRISPR–Cas systems revealed their diversified roles in rice. *Plant Genome* **20**, e20283 (2023).
87. Zheng, X. et al. MIGS as a simple and efficient method for gene silencing in rice. *Front. Plant Sci.* **9**, 662 (2018).

Acknowledgements

We thank Q. Qian (China National Rice Research Institute, Chinese Academy of Agricultural Sciences) for providing xiaowei, a *OsD18*-null mutant in a slightly different Nipponbare background. This research was supported by the National Natural Science Foundation of China (award nos. 32270433, 32101205, 32072045 and 31960423) to Y. Zhang, X.Z. and X.T.; and the National Agricultural Major Science and

Technology Program of China (award no. NK2022010204) to Y. Zhang and J.Z.; the Technology Innovation and Application Development Program of Chongqing (award no. CSTC2021JSCX-CYLHX0001) to X.T., J.Z. and Y. Zhang; the Sichuan Science and Technology Program (award nos. 2021JDRC0032, 2021YFH0084 and 2021YFYZ0016) to J.Z. and Y. Zhang; and the Open Foundation of Jiangsu Key Laboratory of Crop Genetics and Physiology (award no. YCSL202009) to J.Z., Y. Zhang and T.Z. It was also supported by the NSF Plant Genome Research Program (grant nos. IOS-2029889 and IOS-2132693) and the MAES Competitive Grants Program (CGP) to Y.Q.

Author contributions

Y. Zhang proposed the project. Y. Zhang, T.Z., Y.Q. and J.Z. conceived and designed the experiments. G.L. and T.Z. developed the CAPE estimating system. J.Z., X.T. and Y. Zhao designed guide RNA and generated all the constructs for CAPE. G.L., Y. Han and T.Z. built the genome browser for CAPE. Y. Zhao, R.Z., L.L., X.J., Y.G., Y. He, H.Y. and X.Z. performed the rice stable transformation and plant material assays. J.Z., R.Z., Y. Zhao, Q.H. and X.J. performed the mRNA-seq and qPCR experiments. Y.W., Y. Han and Y.B. assisted with data analysis. Y. Zhang, Y.Q., T.Z., J.Z. and G.L. analysed the data and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41477-023-01384-2>.

Correspondence and requests for materials should be addressed to Yiping Qi, Tao Zhang or Yong Zhang.

Peer review information *Nature Plants* thanks Pengcheng Wei and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) 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

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- 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)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- 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	No software were used for data collection. Public datasets were collected from NCBI, PlantRegMap, MBKBase and JASPAR (2020 Core Plants collections) database. Detailed information are available in Methods.
Data analysis	Data analysis was performed using the following software and packages: cutadapt, v3.5; Bowtie, v1.2; Bowtie2, v2.3.2; BWA, version 0.7.17; HISAT2, v2.0.4; StringTie, v2.1.7; Popera, version 1.0.3; deepTools, v3.5.1; FIMO, v4.11.2; bedGraphToBigWig, v377; BEDTools, v2.30.0; Python, v3.8.5; SciPy, v1.6.2; Scikit-learn, v0.24.1; Biopython, v1.76; R, v4.0.5; Patchwork, v1.1.0.9000; ggsignif, v0.6.3; Pheatmap, v1.0.12; edgeR, v3.8.6; IBM SPSS Statistics, 20.0. Genome browser visualization was constructed using JBrowse2, v2.0.0. The source code of CAPE pipeline has been deposited on GitHub: https://github.com/zhangtaolab/CAPE .

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The authors declare that the data supporting the findings are available within the manuscript or Supplementary information.

Human research participants

Policy information about [studies involving human research participants](#) and [Sex and Gender in Research](#).

Reporting on sex and gender

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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

Life sciences study design

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

Sample size

The sample size for our current study was determined based on the specific requirements of each experiment to have sufficient statistical power, as well as accepted sample sizes in relevant published reports within the field. The actual sample size for each experiment was fully described in either the figure legends or the Methods section.

Data exclusions

No data was excluded.

Replication

To ensure reliable and reproducible results, all experimental procedures in this study, including genotype identification, editing efficiency calculation, qPCR and RNA-seq analysis for gene expression levels, and agronomic trait evaluations, fully comply with the widely accepted standards in the field of plant molecular biology, which require a minimum of three independent biological replicates for each experiment. Specific details regarding the replication design of each experiment can be found in the methods section and figure legends of the paper.

Randomization

To guarantee unbiased sampling in this study, all samples, including hygromycin-resistant callus, WT plants, T0 plants, T1 plants, and T2 plants, were randomly distributed across all replicates.

Blinding

To clearly demonstrate the objectivity of the experimental design, for each experiment, the treatments were compared to a control treatment without any prior knowledge of whether the experimental variables being altered would have a positive or negative impact on the results. Moreover, no additional blinding was implemented.

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	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

Methods

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