



## Original Article

# DNA methylation in transposable elements buffers the connection between three-dimensional chromatin organization and gene transcription upon rice genome duplication



Zhenfei Sun<sup>a,1</sup>, Yunlong Wang<sup>b,1</sup>, Zhaojian Song<sup>c,1</sup>, Hui Zhang<sup>d</sup>, Yuanda Wang<sup>a</sup>, Kunpeng Liu<sup>a</sup>, Min Ma<sup>a</sup>, Pan Wang<sup>a</sup>, Yaping Fang<sup>b</sup>, Detian Cai<sup>c,\*</sup>, Guoliang Li<sup>b,\*</sup>, Yuda Fang<sup>a,\*</sup>

<sup>a</sup>Shanghai Collaborative Innovation Center of Agri-Seeds / School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>b</sup>National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, China

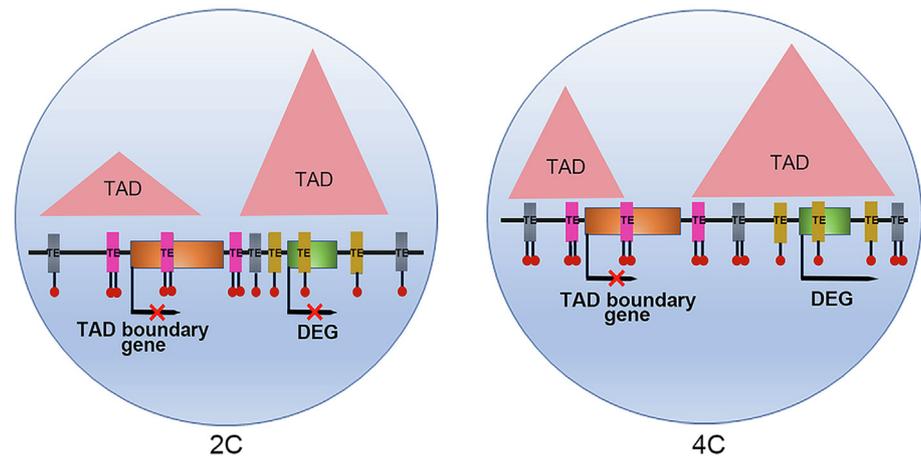
<sup>c</sup>School of Life Sciences, Hubei University, Wuhan 430062, China

<sup>d</sup>National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai 200032, China

## HIGHLIGHTS

- Rice whole-genome duplication leads to alterations of phenotypes, gene transcriptions, 3D chromatin structures and DNA methylation.
- Three-dimensional (3D) chromatin architectures including chromatin compartments and topologically associating domains (TADs) are uncorrelated with gene transcription upon rice genome duplication.
- The first study to reveal the mechanism regarding the disconnection between 3D genome architecture and gene transcription.
- Hypermethylated transposable elements (TEs) suppress the role of 3D chromatin structure in modulating gene transcription upon rice genome duplication.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

## Article history:

Received 28 March 2022

Revised 5 July 2022

Accepted 23 July 2022

Available online 4 August 2022

## Keywords:

Polyploidy rice

## ABSTRACT

**Introduction:** Polyploidy is a major force in plant evolution and the domestication of cultivated crops. **Objectives:** The study aimed to explore the relationship and underlying mechanism between three-dimensional (3D) chromatin organization and gene transcription upon rice genome duplication. **Methods:** The 3D chromatin structures between diploid (2C) and autotetraploid (4C) rice were compared using high-throughput chromosome conformation capture (Hi-C) analysis. The study combined genetics, transcriptomics, whole-genome bisulfite sequencing (WGBS-seq) and 3D genomics approaches to uncover the mechanism for DNA methylation in modulating gene transcription through 3D chromatin architectures upon rice genome duplication.

Peer review under responsibility of Cairo University.

\* Corresponding authors.

E-mail addresses: [dtcai8866@163.com](mailto:dtcai8866@163.com) (D. Cai), [guoliang.li@mail.hzau.edu.cn](mailto:guoliang.li@mail.hzau.edu.cn) (G. Li), [yuda.fang@sjtu.edu.cn](mailto:yuda.fang@sjtu.edu.cn) (Y. Fang).

<sup>1</sup> Contribute equally.

<https://doi.org/10.1016/j.jare.2022.07.007>

2090-1232/© 2022 The Authors. Published by Elsevier B.V. on behalf of Cairo University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Hi-C  
TAD  
Transposable elements  
DNA methylation

**Results:** We found that 4C rice presents weakened intra-chromosomal interactions compared to its 2C progenitor in some chromosomes. In addition, we found that changes of 3D chromatin organizations including chromatin compartments, topologically associating domains (TADs), and loops, are uncorrelated with gene transcription. Moreover, DNA methylations in the regulatory sequences of genes in compartment A/B switched regions and TAD boundaries are unrelated to their expression. Importantly, although there was no significant difference in the methylation levels in transposable elements (TEs) in differentially expressed gene (DEG) and non-DEG promoters between 2C and 4C rice, we found that the hypermethylated TEs across genes in compartment A/B switched regions and TAD boundaries may suppress the expression of these genes.

**Conclusion:** The study proposed that the rice genome doubling might modulate TE methylation to buffer the effects of chromatin architecture on gene transcription in compartment A/B switched regions and TAD boundaries, resulting in the disconnection between 3D chromatin structure alteration and gene transcription upon rice genome duplication.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Polyploidy plays an important role in the formation of new plant species, and plant evolution and breeding [1,2]. A polyploidy plant is often accompanied by powerful biological potentials, improved environmental adaptation, and increased biomass and yield [3,4]. It was known that autotetraploid rice shows larger kernels, higher protein content, better amino acid composition and higher 1000-grain weight than its diploid counterpart [4]. Autotetraploid *Arabidopsis* exhibits obvious phenotypes at both vegetative and reproductive stages, including large leaves, increased whole plant size, late flowering and large seeds [5]. Autotetraploid birch is superior in volume, leaf, breast-height diameter, fruit and stoma, and inferior in height compared to diploid birch [6]. Polyploidization at the cellular and molecular levels often leads to changed chromatin structures and gene expression [5,7].

Chromatin is organized in a highly ordered three-dimensional (3D) architecture instead of a linear nucleotide sequence of the genome [8,9]. The 3D genome is packed in the nucleus in a hierarchical pattern. Chromosome territory (CT) at the several megabase-scale is a higher level of chromatin domain [10,11]. Chromatin in CT is divided into compartments A and B. Compartment A is associated with open chromatin and active transcription, while compartment B is associated with closed chromatin and inactive transcription [9]. Topologically associating domain (TAD) is a predominant structural feature in most organisms [12]. TADs often represent functional domains, as a given TAD contains regulatory elements for the genes inside the same TAD [13]. Therefore, the integrity of the TAD structure is necessary for gene regulation [14]. The location of the TAD boundary is strongly related to the local gene expression, epigenetic landscape, and binding of various insulator proteins [15]. Chromatin loops that appear at 10 kb to 1 Mb [16] function in transcription, replication and recombination [17].

Chromatin structures alterations of are coupled with the changes in gene expressions in some biological processes [18]. For example, in *Arabidopsis*, the switches of compartment A/B lead to the change of transcription during the genome doubling [5]. Higher-order chromatin architecture in rice correlates with transcriptional regulation under heat stress [19]. In cotton, changes in TADs affect the transcriptional activities of abundant genes in tetraploid cotton than in diploid cotton [20]. Other studies reported that 3D structural changes are unrelated to gene expression. For example, the uncoupling relationship between genome topology and gene expression was observed in highly rearranged chromosomes (balancers) spanning ~ 75 % of the *Drosophila* genome [21]. Most TAD disruptions do not result in marked changes in gene expression in human cancers [22]. Recently, chromatin structure and gene expression regulation have been found inde-

pendent during the development of *Drosophila* [23,24]. However, the mechanism of association and unassociation between 3D chromatin structure and gene expression remains unknown.

Polyploidy events trigger many epigenetic and transcriptional changes in the replicated or merged genome [25,26]. In addition to the 3D genomic organization, another important epigenetic factor is DNA methylation at cytosine residues associated with gene transcription by affecting the binding of chromatin proteins including transcription factors to DNA [27]. The precise regulation of DNA methylation is essential for plant and animal development [28,29]. In plants, in addition to the CG context, DNA methylation occurs in sequence contexts of CHG and CHH with which small interfering RNAs (siRNAs) are mainly associated. Most DNA methylation is found in transposable elements (TEs) with CG, CHG, and CHH contexts to suppress the activities of TEs [30]. Substantial methylation is found in the bodies of active genes, which generally occurs in the CG context [31]. DNA methylation in regulatory sequences, such as promoters and enhancers, often leads to gene silencing [32].

In this study, we found that the changes in 3D chromatin structures are not related to transcriptional changes when diploid (2C) rice is duplicated to autotetraploid (4C) rice. In addition, DNA methylation in regulatory regions of genes in compartment A/B switched regions or TAD boundaries is not important for their differential regulation of transcription between 2C and 4C rice. By comparing the methylation of TEs adjacent to genes in compartment A/B switched regions or TAD boundaries and differentially expressed genes (DEGs) between 2C and 4C rice, this study revealed that increased methylation in TEs adjacent to genes in compartment A/B switched regions or TAD boundaries might suppress the transcription of these genes upon rice genome duplication.

## Materials and Methods

### Plant materials

Autotetraploid (4C) rice line was induced from *O. sativa* ssp. *indica* cultivar 9311 (2C) by chromosome doubling according to the method [33–34].

### Characterization of agronomic traits

All plants were cultured in nutrient solution [35] in a growth chamber with 28 °C /25 °C (day/night) and 12 h/12 h (light/dark) cycles. After germination for 15 days, the rice seedlings were transferred to the field. The agronomic traits including plant height, flag leaf length, flag leaf width, tillering number, panicle length, grains per panicle, grain weight, grain length, grain width were scored in

parallel between 2C and 4C rice. The traits were selected and analyzed according to the Descriptors and Data Standard for Rice (*O. sativa* L.) [35].

**Measurement of chlorophylls contents and photosynthetic characteristics** Chlorophylls were measured in fresh leaves collected from 2C and 4C plants. Leaf samples (~50 mg) were immersed in 10 ml extract solution (80 % acetone + 20 % water) for overnight in the dark and examined spectrophotometrically at 663, 646 and 470 nm following the method described by Arnon [36].

Photosynthetic parameters including net carbon assimilation rate (An), intracellular CO<sub>2</sub> concentration (Ci), V<sub>c</sub>max and J<sub>max</sub> were measured with an LI-6400 portable photosynthesis system (LI-COR, Lincoln, USA).

#### Flow cytometry

Ten-day-old 2C and 4C rice seedlings were collected into pre-cooled plate, and then chopped with a new razor blade to release nuclei in the sterile lysis buffer (45 mM MgCl<sub>2</sub>·6H<sub>2</sub>O, 30 mM trisodium citrate, 20 mM MOPS, 1 % Triton X-100, pH 7.0) for 3–5 min until the buffer turns green. Transfer the mixture into strainer. The filtrates were added with the final concentration of 1 ng / ml DAPI solution, and then cultured in dark on ice for 30 min. The ploidy levels were measured by flow cytometry (Beckman Coulter MoFlo XDP, USA).

#### Hi-C library preparation

Hi-C experiments were performed essentially as described [37] with some modifications. Two biological replicates of 2C and 4C rice were performed. In short, 2.5 g of the above ground parts of 10-day-old seedlings were fixed (2 % formaldehyde, 10 % PBS) and ground into powder in liquid nitrogen. The extracted nuclei were digested by incubation with 600 U *Hind*III restriction enzyme at 37 °C overnight, and the digested chromatin at 1 μl 10 mM dATP, dTTP, dGTP and 25 μl 0.4 mM biotin-14-dCTP and 100 U Klenow fragment was placed at 37 °C for 45 min. The ligation reaction was then carried out in 10 × volume of ligation buffer and shaken with 745 μl 10 × ligation buffer, 10 % Triton X-100, 80 μl 10 mg/ml BSA and ATP, 100 Weiss U T4 DNA ligase at 16 °C for 6 h. Then reversely cross-linked with proteinase K at 65 °C overnight. Subsequently, the extracted chromatin was fragmented into an average size of 300 bp by ultrasound (Covaris s220). The Hi-C library was constructed with NEB Next Multiplex Oligos kit and KAPA Hyper Prep Kit. The final library was sequenced on Illumina HiSeq X Ten instrument with 2 × 150 bp reads.

#### Hi-C sequencing data processing

Hic-pro [38] and Bowtie2 [39] were used for Hi-C read mapping. The clean Hi-C reads of 2C and 4C rice were aligned to the genome of *O. sativa* Indica [40] after removing the adapter. Following processing with HiC-Pro and Juicer software [41], valid pairs of 2C and 4C rice were used to create interaction matrixes with bin size 50 kb for further analysis. The reproducibility of two biological replicates was tested with Pearson correlation coefficient from the ICE normalized interaction matrixes [42]. Hicpro2juicebox was used to generate input file for Juicebox. The interaction matrixes were normalized with KR method from Juicer at resolutions 5 kb, 10 kb and 50 kb [38]. After excluding the pericentromeres as reported [37], the first principal component was used to identify compartments with Juicer at 50 kb resolution, and the direction of compartment with high gene expression was defined as A compartment, and the opposite direction as B compartment. We calculated loops and differential loops at 5 kb and 10 kb resolutions, using HICCUPS and HICCUPS Diff in Juicer [40].

We used TADCompare [43] to calculate the TAD structure at a 10 kb resolution.

#### Calculation of chromatin interactions and interaction decay exponents

The normalized interaction matrix of 2C rice was divided by the normalized interaction matrix of 4C rice, and all zeros in the matrix were replaced with the smallest non-zero elements in each matrix to analyze the difference between 2C and 4C rice interaction matrixes. We used log<sub>2</sub> transform and median normalization to standardize the difference matrix. Interaction decay exponents (IDEs) of chromosomes, pericentromeres and telomeres were calculated [37] to study the variation of interaction frequency dependent on the genome distance, the CPM normalization method was used to process the data.

#### Bootstrapping analysis

In the bootstrapping strategies [44], we randomly selected 10,000 groups (n = 10000 times) of the same number of DEGs, and performed the same analysis to determine the percentage of those genes overlapped with A/B switch region genes or TAD changed genes. When the percentile of the test sample was higher than the top five percentiles of the control distribution, it is considered as statistically significant.

#### RNA-seq analysis

Total RNAs were extracted from 10-day-old 2C and 4C rice seedlings using the RNeasy plant mini kit (Qiagen). cDNA library construction and sequencing were carried out by Beijing Genomics Institute (BGI) using BGISEQ-500 platform for 50 bp single-end sequencing as previously described [45]. At least 20 M clean reads of sequencing depth were obtained for each sample. Three independent biological replicates were performed. The clean reads were separately aligned to the genome of *O. sativa* Indica [38] with orientation mode using Tophat software (<http://tophat.cbcb.umd.edu/>). The fragments per kilobase of exon per million mapped reads (FPKM) method was used to calculate the expression level of each transcript. The differential expression analysis was carried out using the classical normalization method of DESeq2 R package [46] with a 0.05p-value, 0.05 false discovery rate, and cutoff of 1 logfold change. The hypergeometric test was performed as previously described [47]. Blast2GO method was used to find homologous genes in japonica rice genome (MSU), and GO functional enrichment analysis was performed by DAVID (<https://david.ncifcrf.gov/>).

#### DNA methylation analysis

For whole genome bisulfite sequencing (WGBS), genomic DNA (gDNA) was extracted from 10-day-old 2C and 4C rice seedlings with the DNeasy plant mini kit (Qiagen) per manufacturer's introduction. Library construction and sequencing were performed by Beijing Genomics Institute (BGI) using Illumina HiSeq-2000 for 100 bp paired-end sequencing. To facilitate the analysis of DNA methylation data, we used Batmeth2 (<https://github.com/GuoliangLi-HZAU/BatMeth2>), an integrated multi-functional software for DNA methylation analysis [48], including sequencing sequence quality filtering, DNA methylation sequence alignment, DNA methylation level calculation and functional annotation.

### Calculation of the distance between a DEG and differentially methylated regions (DMRs)

The whole genome was divided into 1000-bp bins to identify the DMRs in which the absolute value of difference in DNA methylation between 2C and 4C was 0.6 or above, and the adjusted *q* value of Fisher's exact test was 0.05 or less by using BatMeth2 [46]. Finally, the shortest genomic distance of a given DEG and all DMRs in turn was calculated.

### TE annotation and analysis

By running RepeatMasker (v4.0.3, [www.repeatmasker.org](http://www.repeatmasker.org)), the repetitive library of RepBase (v20130422) was used to compare the rice reference genome sequences. To compare the methylation status of TEs between the 2C and 4C rice genomes, we excluded TEs with <40 % of cytosines and coverage of BS reads <3. The remaining TEs were used for further analysis. Using this cut-off value, we obtained a data set of 478,599 TEs for the subsequent analysis.

### Small RNA-seq and data processing

Small RNAs (sRNAs) were isolated from 10-day-old rice seedlings using mirVana™ miRNA Isolation Kit (Ambion, AM1561) and sequenced by Illumina high-throughput sequencing. The small RNA data were processed and analyzed according to the previous description [49,50] with minor modifications. In brief, the raw sequencing reads were trimmed using cutadapt (v1.2.1) to remove adapters, and sRNAs between 16 and 35 nt in lengths were selected and mapped to the rice genome [40,49,50].

sRNAs that matched against the databases including the Rfam database [51] and miRBase [52] were discarded. 24-nt reads that did not match miRNAs, snRNAs, rRNAs, tRNAs, or snoRNAs were filtered and mapped to the genome 1–1,000 times as siRNAs for analyses. The siRNA count was based on the total abundance of genome matched small RNA reads, normalized to reads per million, excluding sRNAs of the above structures, and dividing the number of reads evenly by the number of genome hits. A siRNA cluster was defined as containing at least five different siRNA reading sequences, and adjacent reading sequences <200 bp apart were combined into a cluster.

### Data availability

The Hi-C, WGBS and RNA-seq datasets have been submitted to NCBI (PRJNA725914).

## Results

### Changes in phenotypes and gene expression upon rice whole-genome duplication

2C ( $2 \times 9311$ ) and 4C rice ( $4 \times 9311$ ) were confirmed by flow cytometry (Fig. S1). Compared to 2C rice, 10-day-old 4C rice seedlings exhibited increased chlorophyll content (Fig. S2A–S2B), which effectively improved electron transport rate (ETR) as measured by chlorophyll fluorescence together with parameters calculated from gas exchange [53] (Fig. S2C). The higher non-photochemical quenching (NPQ) [54] (Fig. S2D) indicated that 4C rice had stronger photo-protection capability. In addition, photosynthetic parameters including net carbon assimilation rate (*A*<sub>n</sub>), intracellular CO<sub>2</sub> concentration (*C*<sub>i</sub>), *V*<sub>cmax</sub> and *J*<sub>max</sub> were significantly increased (Table S1). In addition, mature 4C rice seedlings showed no obvious changes of plant height, flag leaf width, and plant weight (Fig. 1A–D) with a decreased tillering number and increased flag

leaf length, 1000-grain weight, number of effective panicles per plant, grain length, and grain width (Fig. 1E–J), similar to the phenotypes of 4C rice *Oryza sativa* ssp. *indica* cv. *Aijiaonante* previously reported [51]. In addition, the nuclei in leaf and root cells of 4C rice are larger than those of 2C rice (Fig. 1K–N).

To evaluate the effect of rice genome doubling on transcription, the transcriptomes of the above ground parts between 10-day-old 2C and 4C seedlings were compared by RNA sequencing. Results indicated that the predominant part of genes was not changed obviously. Among the 698 genes significantly regulated, 510 genes are upregulated and 188 are downregulated (Fig. 1O–P and Table S2) in 4C rice. Gene Ontology (GO) analysis indicated that these DEGs are associated with various biosynthetic and metabolic processes (Fig. S3 and Table S3).

### Intrachromosomal interactions are weakened in autotetraploid rice

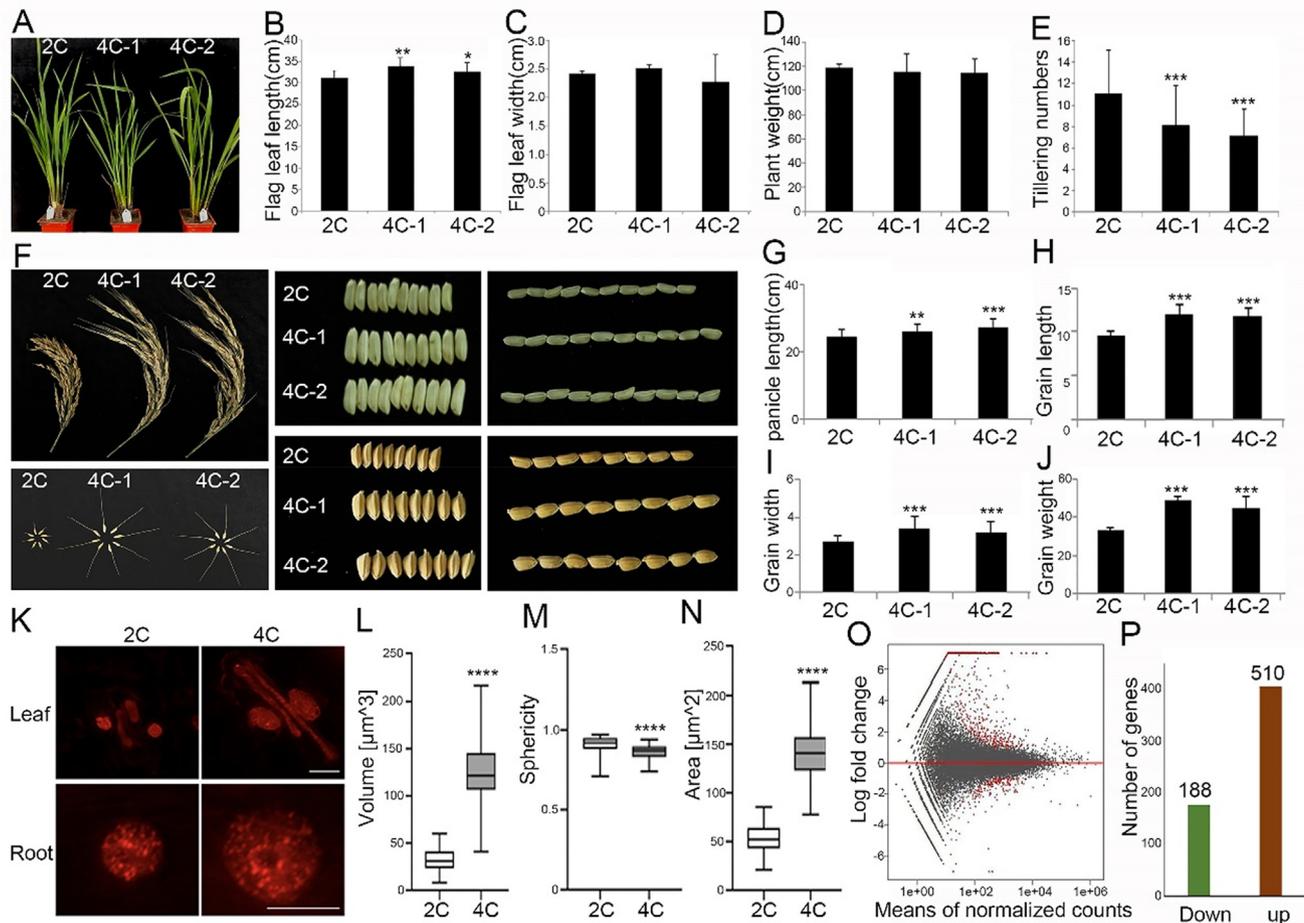
To test whether chromatin organization is rearranged after rice genome duplication, high-throughput chromosome conformation capture (Hi-C) experiments were performed to map chromatin interactions. >843 million and 842 million raw Hi-C reads from 2C and 4C rice were obtained (Table S4), respectively, with a high reproducibility between the two biological replicates of 2C or 4C rice (Fig. S4A and S4B). The relative interaction difference between 2C and 4C rice was calculated. Results showed that 4C rice had slightly increased interchromosomal interactions in Chr2, Chr5, and Chr6; other chromosomes had no significant difference; and dampened intrachromosomal interactions in Chr1, Chr2 and Chr4, and other chromosomes had no significant difference (Fig. 2A–2D). There were also decreased interchromosome arm interactions in most chromosomes in 4C rice than in 2C rice (Fig. 2E and 2F) but no significant difference in intrachromosome arm interactions (Fig. 2H).

To quantitatively assess the chromatin contacts, this study calculated interaction decay exponents (IDEs), that characterize chromatin packing as the slopes of a linear fit of average interaction intensities detected at a given range of genomic distances in a logarithm scale [38]. Results showed that IDEs of intrachromosomes (Fig. 2D; Fig. S5), interchromosome arms (Fig. 2G), and intrachromosome arms (Fig. 2H) were slightly lower in 4C rice than in 2C rice but without significant difference.

### Switches between chromatin compartments A and B upon rice genome doubling are not related to transcriptional regulation

To determine whether the chromatin compartment change is associated with gene expression alteration during rice genome duplication, the first principal component of Pearson's matrix in Hi-C data and gene expression was used to define the active (A) and inactive (B) chromatin compartments in 2C and 4C rice. Compartments A and B were compared using the first principal component at a 50 kb resolution between 2C and 4C rice, and 47.31 % and 50.3 % of the genome showed conserved compartments A and B between 2C and 4C rice, respectively (Fig. 3A). Rice genome doubling induced switches between compartments A and B (Fig. 3A and 3B), with 1 % and 1.39 % of the compartments converted from A to B and from B to A, respectively (Fig. 3A). However, we found that the numbers of expressed genes in either conserved compartments (A to A, or B to B) or switched compartments (B to A, or A to B) had no significant difference between 4C and 2C rice (Fig. 3B).

The study further analyzed the association between chromatin compartment switches and transcription. Only 24 DEGs (24 of 698, ~3.4 %) overlapped with compartment A/B switched regions (Fig. 3C and 3D). To know the confidence of the results, bootstrapping randomized analysis was performed, and 10,000 groups of equal number (698) of genes were randomly selected to determine



**Fig. 1. Comparisons of morphology and transcriptome between 2C ( $2 \times 9311$ ) and 4C rice ( $4 \times 9311$ ).** A, Seedling morphology of 2C and 4C rice. B-E, Flag leaf length (B), flag leaf width (C), plant weight (D), and tillering numbers (E) of 2C and 4C rice seedlings. B-E, rice represent means  $\pm$  SEM (standard error of mean,  $n = 3$  biological replicates). Statistical significance was analyzed by  $t$ -test, \*\*\* $p < 0.001$ . F, Panicle and grain morphology of 2C and 4C rice. G-J, Panicle length (G), grain length (H), grain width (I), and grain weight (J) of 2C and 4C rice plants. Error bars represent means  $\pm$  SEM (standard error of mean,  $n = 3$  biological replicates). Statistical significance was analyzed by  $t$ -test, \*\*\* $p < 0.001$ . K, Nuclei of guard cells and root cells of 2C and 4C rice stained by DAPI. L-N, Volume (L), sphericity (M), and area (N) of 2C and 4C rice nuclei. Error bars represent means  $\pm$  SEM (standard error of mean,  $n = 3$  biological replicates). Statistical significance was analyzed by  $t$ -test, \*\*\* $p < 0.001$ . O, MA plot for statistical significance against gene fold change between 2C and 4C rice. Each gene was marked as a dot. Red dots above 0 represent up-regulated genes in 4C rice, red dots below 0 represent down-regulated genes in 4C rice and black dots represent the other genes in 4C rice. P, Numbers of up- and down-regulated genes ( $|\log_2(\text{fold change})| > 1$ ,  $P_{\text{adj}} < 0.05$ ). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

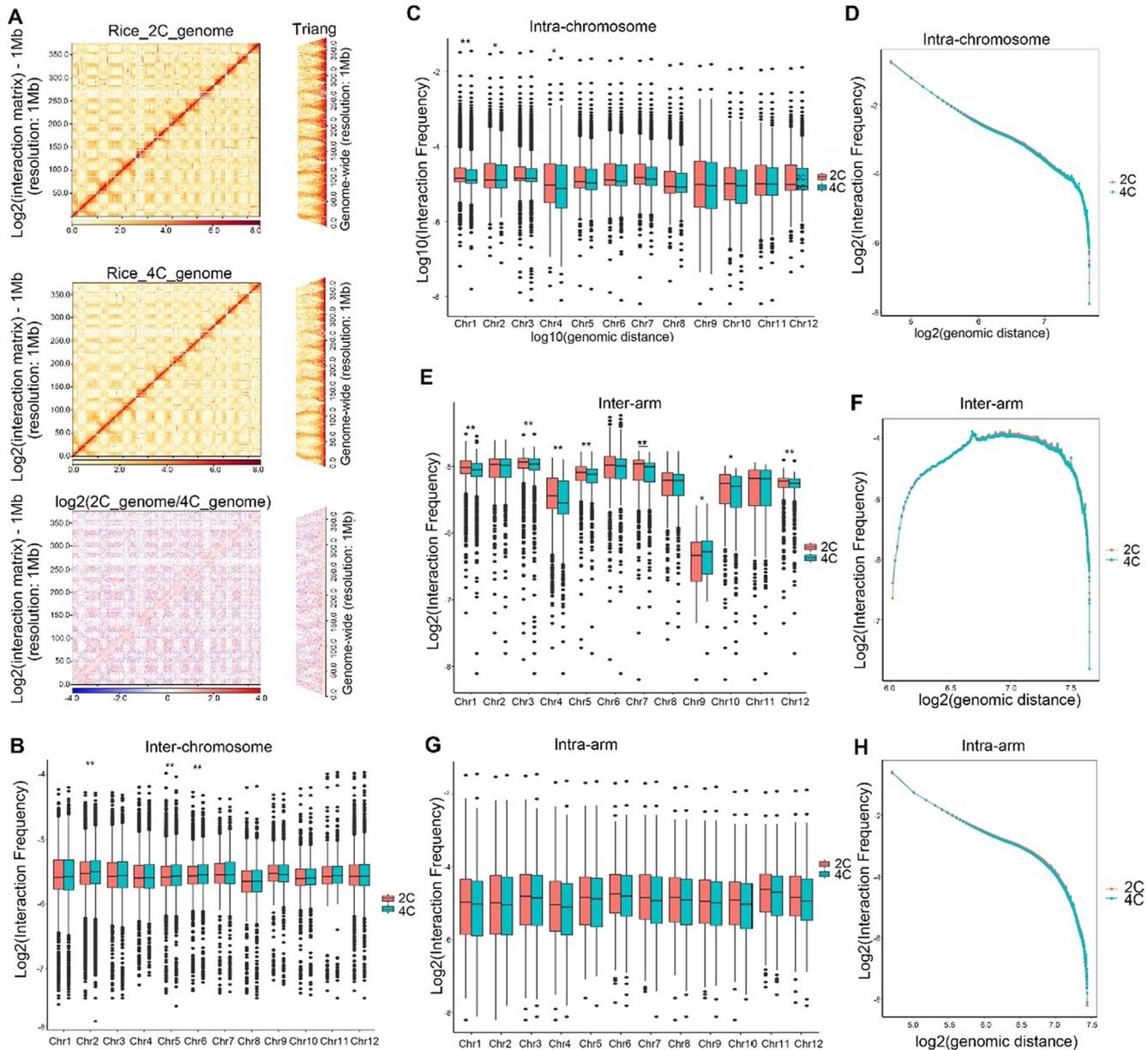
the percentage of genes that overlapped with compartment A/B switched regions. Results showed that 7.24 % of the randomly selected groups had more genes than DEGs overlapped with compartment A/B switched compartments (Fig. 3E), indicating that the overlap of DEGs with compartment A/B switched compartments is not statistically significant. It means that chromatin compartment switches after rice genome duplication is not related to transcription.

#### Differential TAD boundaries and loops are not important for the transcriptional regulation upon rice genome doubling

To know if TAD changes (TADs to non-TAD regions, or non-TADs to TAD regions, are called TAD changes) after rice genome duplication are related to transcription, TADs were first identified by a modified "TADcompare" algorithm [43]. Thereafter, 2688 and 2759 TADs were defined in 2C and 4C rice with a median size of 150 kb, respectively (Fig. S6A). Five group TADs with each group having an equal number of TADs were categorized and created. The five group TADs were arranged according to gene content, with

the gene-poorest bin encompassing  $< 8\%$  of all genes, whereas the gene-richest group carried  $> 50\%$  of all genes (Fig. 4A and 4B).

In addition to TADs, the 5 kb regions adjacent to TADs were defined as TAD boundaries, and the other regions were defined as non-TADs, given that TAD structures were calculated at a 10 kb resolution. In both 2C and 4C rice, the overall protein-coding gene density in TAD regions of 4C rice was similar to that of 2C rice (Fig. S6B), and that in non-TAD regions of 4C rice was slightly higher than that of 2C rice near TAD boundaries (Fig. S6C). The relative locations of non-DEGs were altered in 2C rice than in 4C rice in both TADs and non-TADs (Fig. 4C and 4D). In contrast, the relative locations of DEGs in both TADs and non-TADs in 4C rice were similar to those in 2C rice (Fig. 4E and F), indicating that the changes in TADs during rice genome duplication affect the relative locations of some non-DEGs but did not cause significant differential expression. Among 698 DEGs, 628 and 652 were located in TADs of 2C and 4C rice, respectively. Only a small portion of DEGs (68 of 698,  $\sim 9.76\%$ ) was localized in TAD-changed regions between 2C and 4C rice genomes (Fig. 4G and Table S5). Bootstrapping randomized analysis was then performed, and 10,000 groups of equal number (698) of DEGs were randomly



**Fig. 2. Rice genome doubling weakens the chromatin interactions.** A, Chromatin interaction heatmaps of 2C and 4C rice, and differential chromatin interaction heatmap between 2C and 4C rice at a 1 M resolution. Chromosomes stacked from bottom left to up right were chr1, chr2, chr3, chr4 chr5. B, Boxplots showing inter-chromosome interaction frequencies among all chromosome pairs. C, Boxplots showing the comparison of intra-chromosome interaction frequencies between 2C and 4C rice. D, Interaction decay exponents of intra-chromosome interactions. E, Boxplots showing the comparison of inter-arm interaction frequencies between 2C and 4C rice. Inter-arm interactions are the interactions with both sides inside one chromosome, but from different arms of the same chromosome. F, Interaction decay exponents of inter-arm interactions. G, Boxplots showing the comparison of intra-arm interaction frequencies between 2C and 4C rice. H, Interaction decay exponents of intra-chromosome arm interactions. (\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ , NS  $p > 0.05$ ). The  $p$  values were tested by Wilcoxon–Mann–Whitney test).

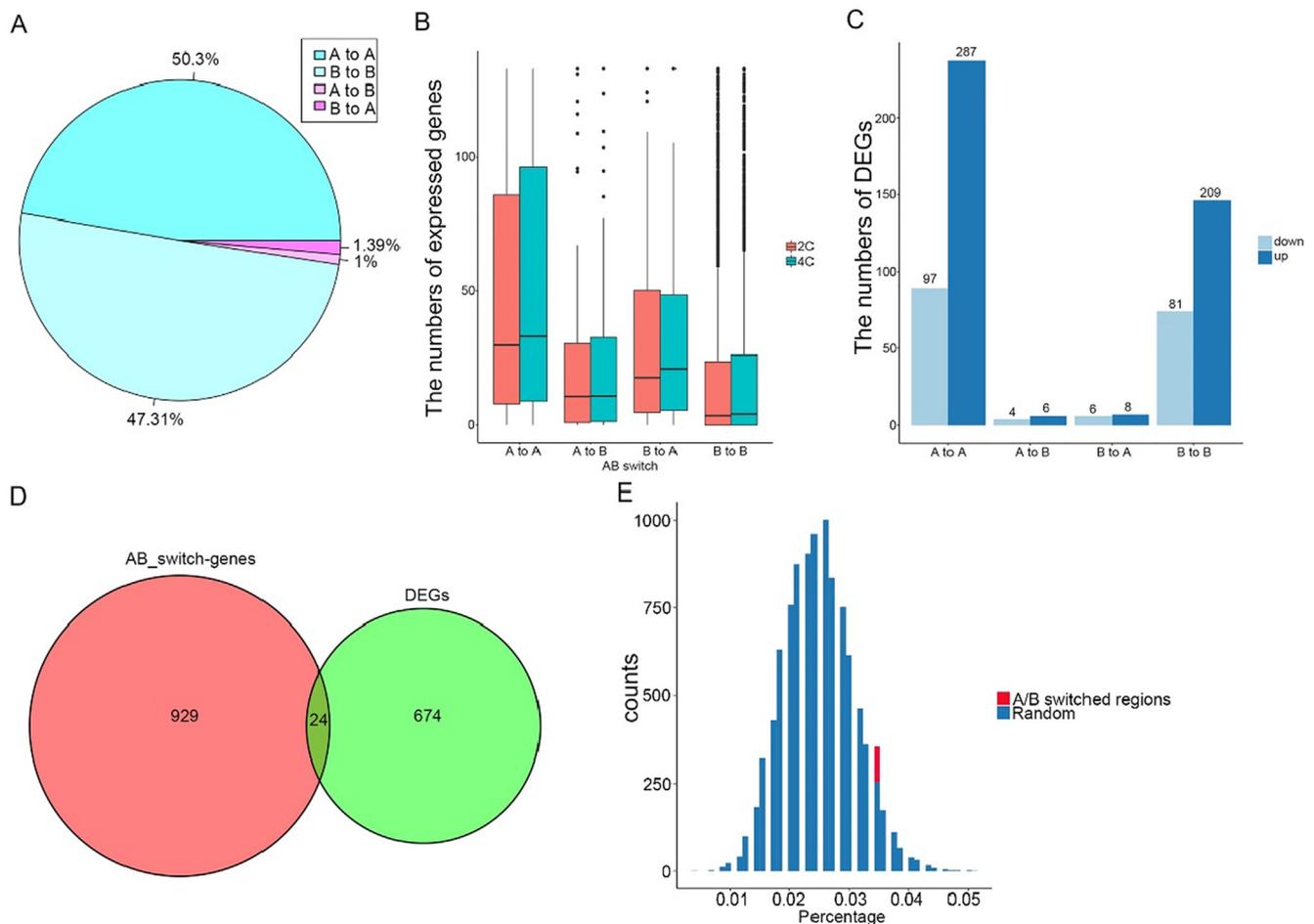
selected to determine the percentage of those genes that overlapped with TAD-changed genes. Results showed that the percentage of DEGs (9 %) was localized in the percentage of randomly selected control genes (~12 %; Fig. 4H), indicating that the changed TAD boundaries after rice genome duplication are not related to transcription.

To know whether loop changes after rice genome duplication are related to transcriptional regulation, 4822 loops in 2C rice and 5365 loops in 4C rice were first annotated, and 79 loops specific for 2C rice and 81 loops specific for 4C rice were then identified (Fig. S6D and E; Table S6). Fragments per kilobase of transcript per million (FPKMs) of genes in loops or nonloops between 2C and 4C rice were of no significant difference (Fig. 4I), and FPKMs of genes

in specific and conservative loops were also of no significant difference in both 2C and 4C rice (Fig. 4J), suggesting that chromatin loops are not important for transcriptional regulation when 2C rice is duplicated to 4C rice.

*DNA methylations in the promoter regions of genes in compartment A/B switched regions and TAD boundaries are not involved in transcriptional regulation upon rice genome duplication*

The global level of DNA methylation increases when 2C rice is duplicated to 4C rice [55]. DNA methylation in different regions of genes and the relationship between DNA methylation and 3D genomic structures in 2C and 4C rice were analyzed. The



**Fig. 3. The changed chromatin compartments are uncorrelated with the gene expression.** A, Pie chart representing the percentages of chromatin compartment switches between 2C and 4C rice. B, Boxplots showing the numbers of protein-coding genes in A/B compartments between 2C and 4C rice. C, Bar graph showing the statistics of DEGs in the switches of compartments A/B. D, Venn diagram showing the numbers of genes in compartment A/B switched regions (pink) and DEGs (green) between 2C and 4C rice. E, Histogram of randomly selected DEGs in AB switch regions ( $n = 10000$ ). The red bar chart shows the true AB switch genes in DEGs, and the blue chart shows randomly selected genes with the same number of DEGs overlapped with switched AB compartments. X-coordinate shows proportion of different genes associated with AB switches. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

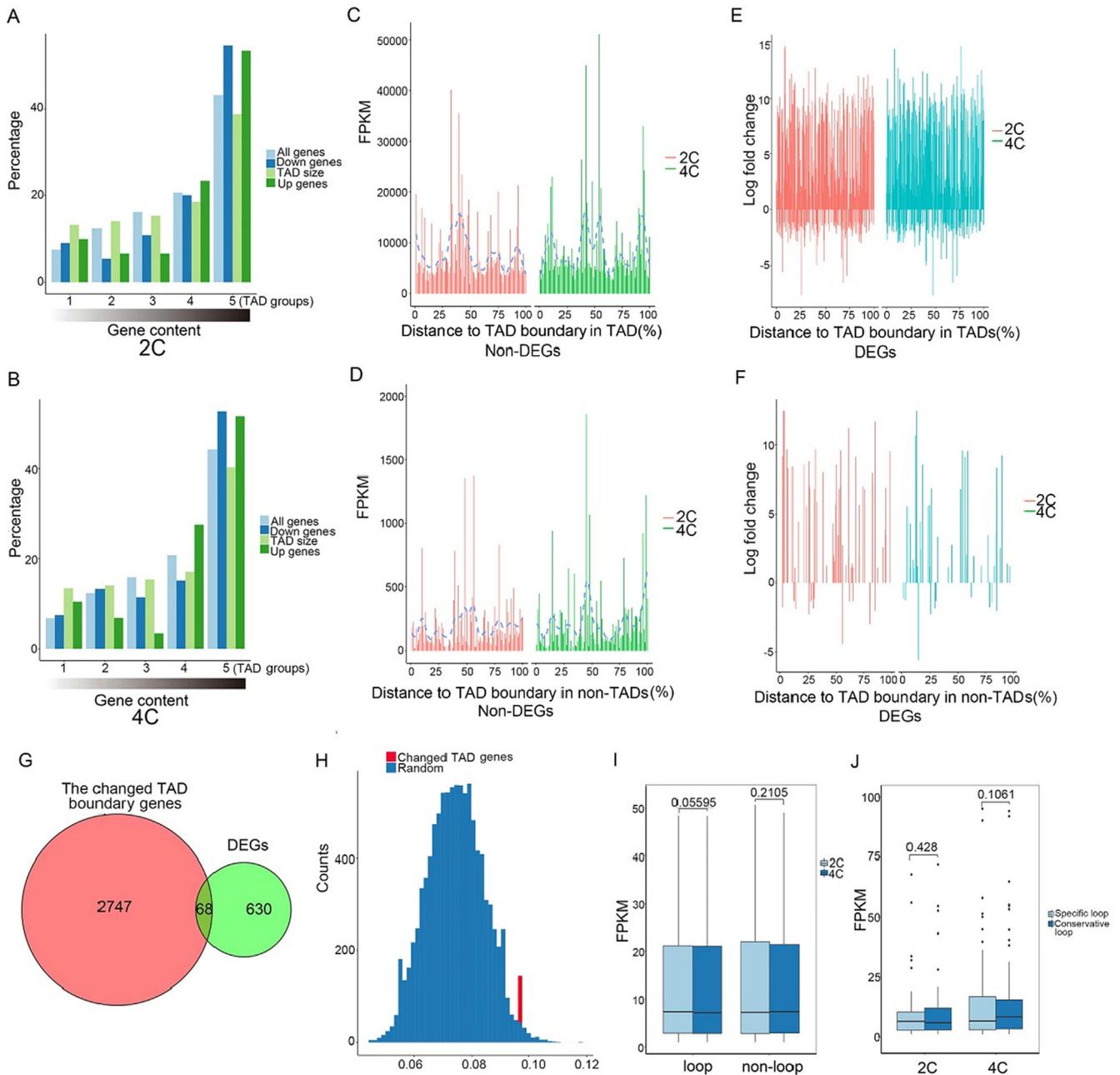
genome-wide DNA methylation revealed by WGBS-seq (Table S7) indicated that the average levels of DNA methylations in CG, CHG and CHH contexts are increased in 4C rice than in 2C rice in genes, upstream and downstream of genes with each gene defined by a continuous exon and intron sequence (Fig. S7A). In 2C and 4C rice, 59 % and 62.5 % of DNA methylation were in CG context, 29 % and 31 % were in CHG context, and 3.5 % and 3.5 % were in CHH context, respectively (Fig. S7B). Compared to 2C rice, 4C rice exhibited increased proportions of methylated cytosines in CG and CHG contexts (Fig. S7B), and CG methylation level was higher than CHG and CHH methylation (Fig. S7B), similar to the previous report [55].

As DNA methylation in the regulatory regions of genes plays an important role in transcriptional regulation [47], DNA methylation in DEG promoters between 2C and 4C rice was compared. All DEGs and non-DEGs were classified according to the log fold change of genes (Fig. S8). Results showed that there was no significant difference in DNA methylation levels in DEG promoters [1000 bp around the transcription start site (TSS)] between 2C and 4C rice, and the CG and CHG methylation levels of non-DEG promoters were lower in 2C rice than in 4C rice (Fig. S8).

This study then analyzed the association between DNA methylation in 4 kb upstream of genes in compartment A/B switched regions and the transcriptional changes in these genes upon rice genome duplication. The CG, CHG, and CHH methylation levels in

4 kb upstream of all genes in compartment A/B switched regions were not significantly changed in 4C rice compared to those in 2C rice (Fig. S9A). Moreover, DNA methylation levels in 4 kb upstream of DEGs in compartment A/B switched regions were similar to those of non-DEGs in 2C and 4C rice (Fig. S9B and C). These results suggested no obvious link between DNA methylation in the upstream of genes in compartment A/B switched regions and the transcriptional regulation of these genes upon rice genome doubling.

To know the role of the methylation of TAD boundary genes in transcriptional regulation, the methylation level of TAD boundary genes was compared to that of DEGs upon rice genome doubling. The CG, CHG, and CHH methylation levels in 4 kb upstream of TAD boundary genes (3247) in 4C rice remained unchanged compared to those (3203) in 2C rice (Fig. 5D; Table S8). In addition, the CG and CHG methylation levels in 4 kb upstream of TAD boundary genes did not significantly change compared to those in 4 kb upstream of DEGs in both 2C and 4C rice (Fig. S9D). Moreover, the CHH methylation level in 4 kb upstream of TAD boundary genes was lower than that of DEGs in both 2C and 4C rice (Fig. S9E and F). In contrast, FPKMs of TAD boundary genes were lower than those of DEGs (Fig. S9G), opposite to the expectation that FPKMs of hypomethylated TAD boundary genes were higher than those of hypermethylated DEGs. Therefore, DNA methylation in upstream

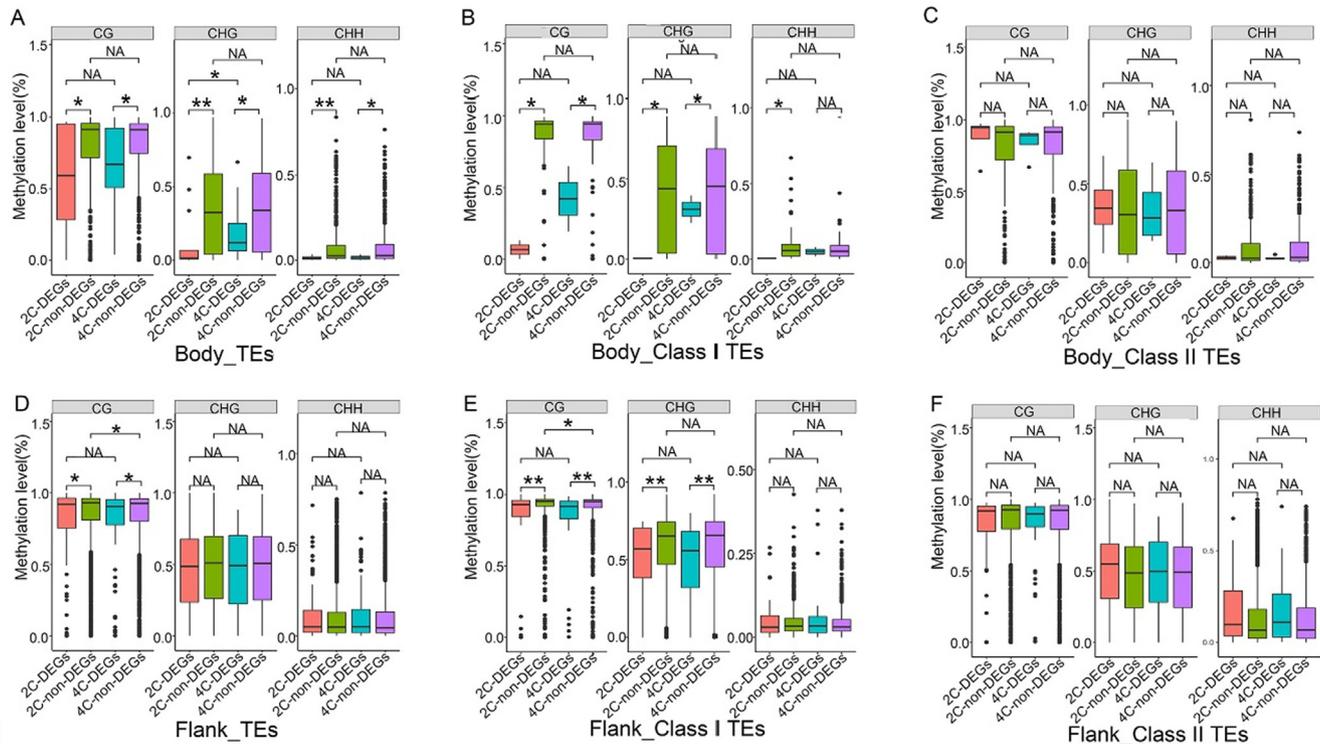


**Fig. 4. TADs and loops are uncorrelated with gene expression.** A and B, Percentages of genes compared with the percentages of DEGs in 2C rice (A) and 4C rice (B) in TADs grouped by the number of overlapping genes (gene content). The ranking on the x-axis is such that the leftmost group contains 20 % TADs with the lowest number of genes and the rightmost group contains 20 % TADs with the highest number of genes. C, Bar plots showing the distance distribution of non-DEGs from the TAD boundary in TADs. X-coordinate shows the percentage from the TSS of genes to the left boundary of TADs, 0 represents the left boundary of TADs, 100 represents the right boundary of TADs, and y-coordinate shows the gene expression. D, Bar plots showing the distance distribution of non-DEGs from the TAD boundary in non-TADs. X-coordinate shows the percentage from the TSS of genes to the TAD boundary, and y-coordinate shows the gene expression. E, Bar plots showing the distance distribution of DEGs from the TAD boundary in TADs. X-coordinate shows the percentage from the TSS of genes to the TAD boundary, and y-coordinate shows the log fold change of DEGs. F, Bar plots showing the distance distribution of DEGs from the TAD boundary in non-TADs. X-coordinate shows the percentage from the TSS of genes to the TAD boundary, and y-coordinate shows the log fold change of DEGs. Non-TAD indicates the regions in the genome except TADs. G, Venn diagrams showing numbers of genes in changed TADs (pink) and DEGs (green) between 2C and 4C rice. H, Histogram of randomly selected DEGs in TAD changed regions (n = 10000). The red bar chart shows the true TAD changed genes in DEGs, and the blue chart shows the TAD changed genes in random number of DEGs. X-coordinate shows proportion of different genes associated with TAD changed. TAD changed means that a gene goes from TAD to non-TAD, or non-TAD to TAD, after chromosome doubling. I, Boxplots showing the normalized RNA-seq FPKMs between genes in loops and non-loops in 2C and 4C rice. J: Boxplots showing the normalized RNA-seq FPKMs of genes in specific loops and conservative loops in 2C and 4C rice. Specific loops indicate loops unique to 2C or 4C, conservative loops indicate loops shared by 2C and 4C. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

regions of TAD boundary genes was not involved in transcriptional regulation.

This study further analyzed the relationships between DEGs and differential methylation regions (DMRs) known to participate in

gene transcription [53]. A total of 1484 CG, 89 CHG, and 1 CHH DMRs were identified in 4C rice than in 2C rice, including 892 CG, 26 CHG, and 1 CHH hypermethylated DMRs and 592 CG, 63 CHG, and 0 CHH hypomethylated DMRs. The genomic distances



**Fig. 5.** TEs in bodies of non-DEGs or regions flanking non-DEGs in compartment A/B switched regions are hypermethylated compared to those of DEGs in 2C and 4C rice. A, Comparison of TE methylation levels in gene bodies between non-DEGs and DEGs in compartments A/B switched regions of 2C and 4C rice. B, Comparison of Class I TE methylation levels in gene bodies between non-DEGs and DEGs in compartments A/B switched regions of 2C and 4C rice. C, Comparison of Class II TE methylation levels in gene bodies between non-DEGs and DEGs in compartments A/B switched regions of 2C and 4C rice. D, Comparison of TE methylation levels in regions flanking genes between non-DEGs and DEGs in compartments A/B switched regions of 2C and 4C rice. E, Comparison of Class I TE methylation levels in regions flanking genes between non-DEGs and DEGs in compartments A/B switched regions of 2C and 4C rice. F, Comparison of Class II TE methylation levels in regions flanking genes between non-DEGs and DEGs in compartments A/B switched regions of 2C and 4C rice. (Class I is retrotransposons including Copia, Gypsy, LTR, LINE and SINE, and class II is transposons including Helitron, Stowaway, DNA, Harbinger, MULE\_MuDR and hAT).

between DEGs and DMRs were then examined [56, 57, 58]. The distance between the TSS locus of each DEG and all DMRs was calculated, and the shortest distance was taken as the distance between a given DEG and DMRs. The genomic distances between DEGs and DMRs were very long (Fig. S10), suggesting no obvious relationships between DEGs and DMRs.

#### *Hypermethylated TEs across non-DEGs in compartment A/B switched regions correlate with inhibited gene transcription upon rice genome doubling*

Given that the number, distance, and methylation level of TEs were known to affect the expression of genes neighboring TEs in *Arabidopsis*, rice, and maize [55,57,59], this study addressed if DNA methylation in TEs was involved in 3D genome-mediated transcriptional regulation. In the rice 9311 genome, 14.57 % of the TEs localized in genes, 38.19 % upstream of genes, and 38.33 % downstream of genes. The methylation levels in CG, CHG, and CHH contexts between 2C and 4C rice were compared for 12 major types of TEs, including class I retrotransposons Copia, Gypsy, LTR, LINE, and SINE and class II transposons Helitron, Stowaway, DNA, Harbinger, MULE\_MuDR, and hAT. Results indicated that the methylation levels of Copia, DNA, Harbinger, LINE, MULE\_MuDR, and SINE in 4C rice differed from those in 2C rice (Fig. S11 and S12). The methylation levels of TEs in gene promoters between 2C and 4C rice were then compared. Results showed no significant difference in methylation levels in TEs in promoters of DEGs and non-DEGs between 2C and 4C rice (Fig. S13).

To analyze the effect of DNA methylation in TEs across genes in compartment A/B switched regions on gene expression, we compared the methylation levels of all TEs, Class I and II TEs across non-DEGs with those across DEGs in compartment A/B switched regions between 2C and 4C rice, with TEs across a gene defined by TEs in the gene body and 4 kb regions flanking the gene. Results indicate that all TEs and class I TEs in non-DEG bodies show hypermethylation in CG, CHG and CHH contexts (Fig. 5A and 5B), and class II TEs in non-DEG bodies show no change in CG, CHG and CHH contexts compared to those in DEG bodies in 2C and 4C rice (Fig. 5C). In addition, results showed that all TEs in 4 kb regions flanking non-DEGs show hypermethylation in the CG context (Fig. 5D), class I TEs show hypermethylation in CG and CHG contexts (Fig. 5E), and class II TEs show no change of methylation in CG, CHG and CHH contexts compared to those in 4 kb regions flanking DEGs in 2C and 4C rice (Fig. 5C). In contrast, FPKMs of non-DEGs in compartment A/B switched regions were lower than those of DEGs in both 2C and 4C rice (Fig. S14). Results showed that the hypermethylation of TEs across non-DEGs in compartment A/B switched regions correlates with the suppressed transcription of these genes.

#### *Hypermethylated CG in TEs across TAD boundary genes correlates with suppressed gene transcription upon rice genome duplication*

As TAD boundary plays an important role in the regulating local transcription and epigenetic landscape in different species [18,60], the effects of the methylation levels of TEs across TAD boundary genes on the role of 3D chromatin structure in transcriptional reg-

ulation were evaluated. As TEs including Copia, DNA, Harbinger, LINE, MULE\_MuDR and SINE exhibit differential DNA methylation between 2C and 4C rice (Fig. S11 and S12), the methylation levels of these TEs across TAD boundary genes in 2C and 4C rice were compared to those across DEGs, including gene bodies and 4 kb regions flanking these genes. Results indicated that Copia, DNA, and MULE-MuDR across TAD boundary genes show hypermethylation in the CG context, no change in the CHG context, and hypomethylation in the CHH context compared to those across DEGs in 2C and 4C rice (Fig. 6A–6D).

To verify CHH hypomethylation in TEs, the clusters of siRNAs were analyzed, as siRNAs often mediate CHG and CHH methylation in plants, and 24 nt siRNAs can enter the RdDM pathway to trigger DNA methylation and transcriptional silencing to suppress TE activities [61]. The relationship between CHH methylation level and siRNA abundance in TEs across TAD boundary genes and DEGs was investigated. siRNA length profiles similar to those previously reported [53] were obtained in 2C and 4C rice (Fig. S15A). A total of 154,968 siRNA clusters in 2C rice and 147,589 in 4C rice were identified. Most siRNA clusters overlapped with class I and II TEs, and the cluster fractions in class I and II TEs, genes, and intergenic regions of 4C rice were similar to those of 2C (Fig. S15B). The lower CHH methylation level in TEs across TAD boundary genes than in TEs across DEGs (Fig. 6A–D) was parallel with the lower siRNA level in TEs across TAD boundary genes than in TEs across DEGs (Fig. S16A–D).

The CG methylation level of TEs across TAD boundary genes was much higher than those of CHG and CHH methylation in 2C and 4C (Fig. S17A and 17B; Fig. 6A–6D). In contrast, FPKMs of TAD boundary genes were lower than FPKMs of DEGs in 2C and 4C rice (Fig. S9G), indicating that CG hypermethylation of TEs across TAD boundary genes correlates with the inhibited transcriptions of these genes.

#### *Hypermethylated TEs adjacent to TAD boundary genes suppress the transcription of TAD boundary genes upon rice genome doubling*

This study confirmed the role of TE methylation in regulating TAD boundary genes by comparing the expression levels between TAD boundary genes and DEGs at different distances to the closest TE in 2C and 4C rice. As > 97 % TEs were clustered in the 1.2 kb regions away from TAD boundary genes or DEGs in both 2C and 4C, the expression of TAD boundary genes or DEGs from 0 to 1.2 kb away from TEs was analyzed (Table S9). In addition, the average TE length was 337 bp; therefore, 400 bp was chosen as the distance interval. Compared to TEs in the body of boundary genes (0 bp in Fig. S18A for 2C and Fig. S18B for 4C), CG, CHG, and CHH methylation levels in TEs at 0 to 400 bp slightly increased (0–400 bp in Fig. S18A for 2C and Fig. S18B for 4C), whereas the expression level of TAD boundary genes slightly decreased (0–400 bp in Fig. S18C for 2C and Fig. S18D for 4C). Compared to TEs at 0 to 400 bp, CG, CHG, and CHH methylation levels of TEs at 400 to 800 bp decreased (400–800 bp in Fig. S18A for 2C and Fig. S18B for 4C), whereas the expression level of TAD boundary genes dramatically increased (400–800 bp in Fig. S18C for 2C and Fig. S18D for 4C). Compared to TEs at 400 to 800 bp, the CG, CHG, and CHH methylation levels of TEs at 800 to 1200 bp increased (400–800 bp in Fig. S18A for 2C and Fig. S18B for 4C), whereas the expression level of TAD boundary genes decreased (400–800 bp in Fig. S18C for 2C and Fig. S18D for 4C). Importantly, the methylation trends in TEs adjacent to TAD boundary genes were opposite to the gene expression trends in 2C and 4C rice (Fig. S18A vs 8C and Fig. S18B vs 8D). The troughs of DNA methylation levels appeared within 400 to 800 bp away from the closest TE (Fig. S18A and S18B). These data suggested that the expression

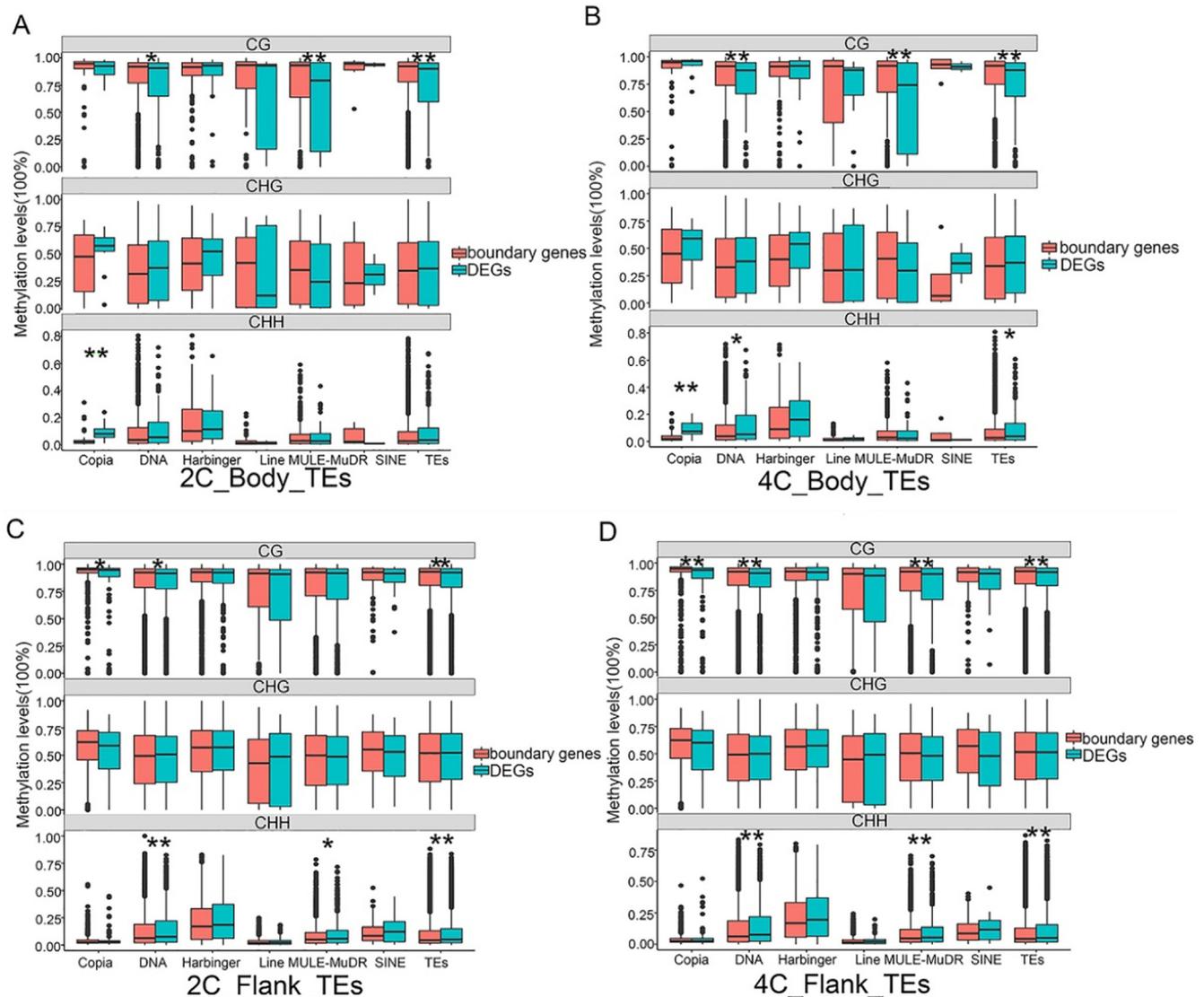
levels of TAD boundary genes are positively correlated with the distance to the closest TE in 2C and 4C rice.

In contrast, the expression levels of DEGs of 2C rice (2C vs 4C) decreased with an increasing distance from TEs. Conversely, the expression levels of DEGs of 4C rice (4C vs 2C) increased with an increasing distance from TEs, and the peak of those expressional levels appeared within 800 to 1200 bp away from the closest TEs (Fig. S18C and S18D). Importantly, the methylation trends in TEs adjacent to DEGs within 1.2 kb were not opposite of gene expression in both 2C and 4C rice (Fig. S18A vs S18C and Fig. S18B vs S18D), indicating that the expression levels of DEGs are uncorrelated with the distance to the closest TE within 1.2 kb in 2C and 4C rice. It was proposed that the hypermethylation in TEs adjacent to TAD boundary genes may buffer the effects of TAD boundaries on gene transcription.

#### Discussion

Polyploidization promotes the evolution of higher plants [65]. Many plants, including *Arabidopsis*, rice, soybean, poplar, sorghum, and maize, might have experienced whole-genome duplication events during their evolution [66]. More attention has been paid to the phenotypic, epigenetic, and gene expression changes in autopolyploidy plants but less to the changes in the 3D genome and their effects on gene expression during autopolyploidization. The study of the 3D chromatin topology of autopolyploid crops is important for understanding the contribution of the spatial organization of the genome to the success of polyploidy species. It was previously reported that altered chromatin interactions in 4C dicotyledonous *Arabidopsis* modulate the transcription compared to its 2C progenitor [5]. To better understand the change of the 3D genome during monocotyledonous genome doubling and its potential effects on gene expression, Hi-C, epigenome, and transcriptome analysis was performed in 4C rice and its progenitor 2C rice. Results showed that rice genome doubling dampens intrachromosomal interactions. Importantly, changes in 3D chromatin structure upon rice genome duplication were uncoupled with gene transcriptions, reminiscent of several reports that 3D chromatin structures are unrelated to gene expression [23,24,60,67]. The phenotypes of tetraploid and 2C *Arabidopsis* were very different, and the changes in 3D genome structure were obvious. There are 68 DEGs localized in TAD-changed regions between 2C and 4C rice genomes (Fig. 4G and Table S5). Among these 68 DEGs, *LOC\_Os12g12514*, which is upregulated in 4C rice (Table S5), encodes a NADP-dependent oxidoreductase responsible for catalyzing the last step of photosynthetic linear electron transfer [63,64]. *LOC\_Os12g12514* is close to a TE with hypomethylated CG, which might be related to the upregulation of *LOC\_Os12g12514* and improvement of electron transport rate (ETR) in 10-day-old 4C rice seedlings compared to 2C rice (Fig. S2C). Therefore, the DNA methylation of TE has important biological significance for maintaining the stability of the duplicated genomes and regulating the phenotypes.

Although DNA methylation variation has been observed [68–70] during genome duplication in plants, TE methylations also affected the expression of nearby genes in *Arabidopsis* [58], rice [55], and maize [59]. The specific role of TE methylation in 3D chromatin structure alterations is unclear. The 3D genome architecture modulated gene transcription by bringing together distant promoters, enhancers, and other *cis*-regulatory elements [71]. Rice genome doubling might be accompanied by locational changes in regulatory elements, such as promoters and enhancers, which, however, do not result in a change in gene expression. Interestingly, the disconnection between transcriptional regulation and A/B switches or TAD boundaries upon genome duplication was not due to DNA methylation alterations in the regulatory sequences of the A/B switch- and TAD boundary-related genes but to the DNA methylation changes in TEs adjacent to these genes.



**Fig. 6.** TEs across non-DEGs in TAD boundaries are hypermethylated compared to those across DEGs in 2C and 4C rice. A, Comparison of TE methylation levels in gene bodies between TAD boundary genes and DEGs in 2C rice. B, Comparison of TE methylation levels in gene bodies between TAD boundary genes and DEGs in 4C rice. C, Comparison of TE methylation levels in regions flanking genes between TAD boundary genes and DEGs in 2C rice. D, Comparison of TE methylation levels in regions flanking genes between TAD boundary genes and DEGs in 4C rice. (“flanking” represents the 4 kb regions flanking genes).

This study suggested that autopolyploidization may stimulate TE modification to reduce the effects of the changes in 3D chromatin structure on gene expression during genome doubling. The underlying mechanism might be that the decrease of intrachromosomal interactions is beneficial to the activities or methylation of TEs by increasing the genomic accessibility to DNA methyltransferases and/or demethylases to antagonize the effects of decreased chromatin interactions on genomic regulation, resulting in the disruption of the association between 3D chromatin structure and gene expression, which might contribute to the success of polyploidy plants during evolution.

In humans and animals, many CTCF binding sites are derived from TEs [72], and the CTCF protein defines TAD boundaries to mediate TAD formation [15]. TEs adjacent to TAD boundary genes can inhibit the expression of these genes in 4C and 2C rice. However, CTCF-like proteins have not been identified in plants. It is possible that there are other elements or proteins in plants that might play a similar role in defining TAD boundaries to that of the CTCF factor in animals.

Increased interchromosomal interactions and decreased intra-chromosome arm interactions were observed in 4C *Arabidopsis* compared to its 2C progenitor [5]. 4C *Arabidopsis* seedlings showed obvious phenotypic changes in the vegetative stage, including serrated leaves, more rosette leaves, and increased whole plant size [5], which are related to 3D chromatin structure alterations, resulting in gene expression changes at this stage. In contrast, no obvious phenotypes of monocotyledons 4C rice were observed at the vegetative stage, as 3D chromatin structural changes were not related to gene expression when 2C rice is duplicated to 4C rice. In the ripening stage, 4C rice acquires many morphologic traits compared to 2C rice [55] (Fig. 1). Therefore, it is interesting to study the relationship between transcriptional regulation and 3D chromatin organization in the reproductive stage upon rice genome duplication. In addition, there were 698 DEGs in the vegetative stage. As their differential expression was not caused by DNA methylation alterations or 3D chromatin rearrangement, other epigenetic mechanisms might be involved in modulating the expression of these 698 genes, such as histone modification, acetylation, or noncoding RNAs.

The relationships of higher-order chromatin structures with epigenetic regulation, including DNA methylation, histone modifications, and noncoding RNAs, are implicated in multiple developmental processes. In human cells, large DNA methylation nadirs can mediate the formation of long loops [73]. In *Arabidopsis*, long noncoding RNA (APOLO) can modulate local chromatin 3D conformation by regulating the conformation of DNA-RNA loops within the nuclei [74]. Histone modifications mediate the effects of genetic variants related to schizophrenia by modulating chromatin higher-order structure [75]. Here, TE methylation diminishes the effects of chromatin higher-order structure alterations on gene expression. It will be worth studying the specific epigenetic networks that include high-order chromatin architecture, DNA and histone modifications, noncoding RNAs, and other epigenetic factors during plant genome duplication.

## Conclusions

This study artificially synthesized a 4C rice line from *O. sativa* ssp. *indica* cultivar 9311 and observed the changes in agriculture traits and enlarged nuclei in 4C rice. Rice genome doubling led to dampened intrachromosomal interactions using Hi-C analysis. Changes in 3D chromatin structures, including chromatin compartments and TADs, were uncorrelated with gene transcription upon rice genome duplication. DNA methylation in the regulatory sequences of genes in compartment A/B switched regions and TAD boundaries did not contribute to transcriptional regulation in 2C and 4C rice. Importantly, hypermethylated TEs across genes in compartment A/B switched regions and TAD boundaries buffered the effects of chromatin architecture changes on gene transcription upon genome doubling.

## Compliance with ethics requirements ethical statement

This article does not contain any studies with human or animal subjects.

## Data availability

The Hi-C, WGBS and RNA-seq datasets have been submitted to NCBI (PRJNA725914).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement and funding sources

This work was supported by the National Key Research and Development Program of China (2016YFD0100902) and National Science Foundation of China (31871230 and 32170585 to Yuda Fang; 31971334 to Zhen-fei Sun).

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jare.2022.07.007>.

## References

- Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen KS, et al. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 2014;345(6194).
- Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. Polyploidy and genome evolution in plants. *Curr Opin Genet Dev* 2015;35:119–25.
- Chao DY, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, et al. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* 2013;341:658–9.
- Wu J, Shahid MQ, Guo H, Yin W, Chen Z, Wang L, et al. Comparative cytological and transcriptomic analysis of pollen development in autotetraploid and diploid rice. *Plant Reprod* 2014;27(4):181–96.
- Zhang H, Zheng R, Wang Y, Zhang Y, Hong P, Fang Y, et al. The effects of *Arabidopsis* genome duplication on the chromatin organization and transcriptional regulation. *Nucleic Acids Res* 2019; 47(15): 7857–7869.
- Mu HZ, Liu ZJ, Lin L, Li HY, Jiang J, Liu GF. Transcriptomic analysis of phenotypic changes in birch (*Betula platyphylla*) autotetraploids. *Int J Mol Sci* 2012;13(10):13012–29.
- Concia L, Veluchamy A, Ramirez-Prado JS, Martin-Ramirez A, Huang Y, Perez M, et al. Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol* 2020;21(1).
- Meaburn KJ, Misteli T. Cell biology: chromosome territories. *Nature* 2007;445(7126):379–781.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009;326(5950):289–93.
- Cremer T, Cremer M, Dietzel S, Muller S, Solovei I, Fakan S. Chromosome territories—a functional nuclear landscape. *Curr Opin Cell Biol* 2006;18(3):307–16.
- Gibcus JH, Dekker J. The hierarchy of the 3D genome. *Mol Cell* 2013;49(5):773–82.
- Wang C, Liu C, Roqueiro D, Grimm D, Schwab R, Becker C, et al. Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res* 2015;25(2):246–56.
- de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 2013;502(7472):499–506.
- Lupiáñez DaríoG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015;161(5):1012–25.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012;485(7398):376–80.
- Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell* 2009;137(7):1194–211.
- Mukherjee A, Mukherjee RN. Kinetic regulation of hexokinase activity in a heterogeneous branched bienzyme system. *Biochim Biophys Acta* 1988;954(1):126–36.
- Ouyang W, Xiong D, Li G, Li X. Unraveling the 3D genome architecture in plants: present and future. *Mol Plant* 2020;13(12):1676–93.
- Liang Z, Zhang Q, Ji C, Hu G, Zhang P, Wang Y, et al. Reorganization of the 3D chromatin architecture of rice genomes during heat stress. *BMC Biol* 2021;19(1).
- Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, et al. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat Plants* 2018;4(2):90–7.
- Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong E. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* 2019;51(8):1272–82.
- Akdemir KC, Le VT, Chandran S, Li Y, Verhaak RG, Beroukhim R, et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet* 2020;52(3):294–305.
- Espinola SM, Götz M, Bellec M, Messina O, Fiche J-B, Houbron C, et al. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early *Drosophila* development. *Nat Genet* 2021;53(4):477–86.
- Ing-Simmons E, Vaid R, Bing XY, Levine M, Mannervik M, Vaquerizas JM. Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning. *Nat Genet* 2021;53(4):487–99.
- Seoighe C, Gehring C. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet* 2004;20(10):461–4.
- Diez CM, Roessler K, Gaut BS. Epigenetics and plant genome evolution. *Curr Opin Plant Biol* 2014;18:1–8.
- Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology* 2013;38(1):23–38.
- Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet* 2013;14(3):204–20.
- Kawashima T, Berger F. Epigenetic reprogramming in plant sexual reproduction. *Nat Rev Genet* 2014;15(9):613–24.
- Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 2010;328(5980):916–9.
- Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 2010;11(3):204–20.
- Schubeler D. Function and information content of DNA methylation. *Nature* 2015;517(7534):321–6.
- Cai DT, Chen DL, Chen JG, Liu YQ. A method of inducing polyploid rice with high frequency by combing tissue culture and chemical agent. *China Patent* 2004; ZL01133529.7.
- Cai DeTian, Chen JianGuo, Chen DongLing, Dai BingCheng, Zhang W, Song ZhaoJian, et al. The breeding of two polyploid rice lines with the characteristic of polyploid meiosis stability. *Sci China, Ser C Life Sci* 2007;50(3):356–66.

- [35] Han LZ WX. Descriptors and Data Standard for Rice (*Oryza sativa* L.). 2006. (China Agricultural Press, Beijing).
- [36] Arnon DJ. Copper enzymes in isolated chloroplasts. Polyphenoloxidase in *Beta vulgaris*. *Plant Physiol* 1949;24:1–15.
- [37] Grob S, Schmid MW, Grossniklaus U. Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol Cell* 2014;55(5):678–93.
- [38] Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 2015;16(1).
- [39] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357–9.
- [40] Yu J, Hu S, Wang J, Wong G-S, Li S, Liu B, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 2002;296(5565):79–92.
- [41] Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* 2016;3(1):95–8.
- [42] Lin Da, Hong P, Zhang S, Xu W, Jamal M, Yan K, et al. Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nat Genet* 2018;50(5):754–63.
- [43] Cresswell KG, Dozmorov MG. TADCompare: an R package for differential and temporal analysis of topologically associated domains. *Front Genet* 2020;11:158.
- [44] Buonaccorsi JP, Romeo G, Thoresen M. Model-based bootstrapping when correcting for measurement error with application to logistic regression. *Biometrics* 2018;74(1):135–44.
- [45] Huang Y, Xu Z, Xiong S, Qin G, Sun F, Yang J, et al. Dual extra-retinal origins of microglia in the model of retinal microglia repopulation. *CellDiscov* 2018;4(1).
- [46] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15(12):550.
- [47] Wollmann H, Stroud H, Yelagandula R, Tarutani Y, Jiang D, Jing Li, et al. The histone H3 variant H3.3 regulates gene body DNA methylation in *Arabidopsis thaliana*. *Genome Biol* 2017;18(1).
- [48] Zhou Q, Lim JQ, Sung WK, Li G. An integrated package for bisulfite DNA methylation data analysis with Indel-sensitive mapping. *BMC Bioinf* 2019;20(1):47.
- [49] McCormick KP, Willmann MR, Meyers BC. Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence* 2011;2(1):2.
- [50] Zhao W, Wang J, He X, Huang X, Jiao Y, Dai M, et al. BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Res* 2004;32:D377–82.
- [51] Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, et al. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* 2013;41:D226–32.
- [52] Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011;39:D152–7.
- [53] Osei-Bonsu I, McClain AM, Walker BJ, Sharkey TD, Kramer DM. The roles of photorespiration and alternative electron acceptors in the responses of photosynthesis to elevated temperatures in cowpea. *Plant Cell Environ* 2021;44(7):2290–307.
- [54] Chen J, Wang P, Hi Mi, Chen GY, Xu DQ. Reversible association of ribulose-1, 5-bisphosphate carboxylase/oxygenase activase with the thylakoid membrane depends upon the ATP level and pH in rice without heat stress. *J Exp Bot* 2010;61:2939–50.
- [55] Zhang J, Liu Y, Xia EH, Yao QY, Liu XD, Gao LZ. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proc Natl Acad Sci USA* 2015;112(50):E7022–9.
- [57] Becker C, Hagemann J, Müller J, Koenig D, Stegle O, Borgwardt K, et al. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 2011;480(7376):245–9.
- [58] Wang X, Weigel D, Smith LM. Transposon variants and their effects on gene expression in *Arabidopsis*. *Plos Genet* 2013;9(2):e1003255.
- [59] Forestan C, Farinati S, Aiese CR, Lunardon A, Sanseverino W, Varotto S. Maize RNA PolIV affects the expression of genes with nearby TE insertions and has a genome-wide repressive impact on transcription. *BMC Plant Bio* 2017;17(1):161.
- [60] Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* 2017;171(3):557–572.e24.
- [61] Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* 2014;15(6):394–408.
- [63] Kidokoro S, Kim J-S, Ishikawa T, Suzuki T, Shinozaki K, Yamaguchi-Shinozaki K. DREB1A/CBF3 is repressed by transgene-induced DNA methylation in the *Arabidopsis* *ice1-1* mutant. *Plant Cell* 2020;32(4):1035–48.
- [64] Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Ulrich MA, Libiger O, et al. Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 2011;334(6054):369–73.
- [65] Otto SP. The evolutionary consequences of polyploidy. *Cell* 2007;131(3):452–62.
- [66] Jiang WK, Liu YL, Xia EH, Gao LZ. Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants. *Plant Physiol* 2013;161(4):1844–61.
- [67] Dong P, Tu X, Li H, Zhang J, Grierson D, Li P, et al. Tissue-specific Hi-C analyses of rice, foxtail millet and maize suggest non-canonical function of plant chromatin domains. *J Inter Plant Biol* 2020;62(2):201–17.
- [68] Lee HS, Chen ZJ. Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc Natl Acad Sci USA* 2001;98(12):6753–8.
- [69] Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiol* 2002;129(2):733–46.
- [70] Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiol* 2006;140(1):336–48.
- [71] Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 2012;13(9):613–26.
- [72] Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, et al. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 2012;148(1–2):335–48.
- [73] Zhang X, Jeong M, Huang X, Wang XQ, Wang X, Zhou W, et al. Large DNA methylation nadirs anchor chromatin loops maintaining hematopoietic stem cell identity. *Mol Cell* 2020;78(3):506–521.e6.
- [74] Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, Veluchamy A, et al. R-loop mediated trans action of the APOLO long noncoding RNA. *Mol Cell* 2020;77(5):1055–1065.e4.
- [75] Punzi G, Bharadwaj R, Ursini G. Neuroepigenetics of Schizophrenia. *Prog Mol Biol Transl Sci* 2018;158:195–226.