Article

Characterization of Transposon-Derived Accessible Chromatin Regions in Rice (*Oryza Sativa*)

Aicen Zhang ⬦ and Wenli Zhang *⬦

State Key Laboratory for Crop Genetics and Germplasm Enhancement, Collaborative Innovation Center for Modern Crop Production Co-sponsored by Province and Ministry (CIC-MCP), Nanjing Agricultural University, No.1 Weigang, Nanjing 210095, China
* Correspondence: wzhang25@njau.edu.cn

Abstract: Growing evidence indicates that transposons or transposable elements (TEs)-derived accessible chromatin regions (ACRs) play essential roles in multiple biological processes by interacting with *trans*-acting factors. However, the function of TE-derived ACRs in the regulation of gene expression in the rice genome has not been well characterized. In this study, we examined the chromatin dynamics in six types of rice tissues and found that ~8% of ACRs were derived from TEs and exhibited distinct levels of accessibility and conservation as compared to those without TEs. TEs exhibited a TE subtype-dependent impact on ACR formation, which can be mediated by changes in the underlying DNA methylation levels. Moreover, we found that tissue-specific TE-derived ACRs might function in the tissue development through the modulation of nearby gene expression. Interestingly, many genes in domestication sweeps were found to overlap with TE-derived ACRs, suggesting their potential functions in the rice domestication. In addition, we found that the expression divergence of 1070 duplicate gene pairs were associated with TE-derived ACRs and had distinct distributions of TEs and ACRs around the transcription start sites (TSSs), which may experience different selection pressures. Thus, our study provides some insights into the biological implications of TE-derived ACRs in the rice genome. Our results imply that these ACRs are likely involved in the regulation of tissue development, rice domestication and functional divergence of duplicated genes.

Keywords: transposable elements; accessible chromatin regions; regulation of gene transcription; *Oryza sativa*

1. Introduction

Transposable elements (TEs) have been the subject of much debate since they were first discovered and described as "controlling elements" in maize (*Zea mays*) by Barbara McClintock [1]. They represent a major type of repetitive DNA sequences and account for large proportions of many plant genomes, such as 40% in rice, 60% in cotton, and more than 80% in wheat and maize [2–5]. Initially, TEs were considered as foreign DNA invading the host genome, and their genome-wide propagation, or so-called transposition, frequently occurs during evolution or in response to biotic and abiotic stresses [6–10]. According to their transposition mechanisms, TEs are usually divided into two distinct classes [11]. Class I TEs transpose via a "copy and paste" mechanism with a reversely transcribed RNA as an intermediate. They can be further classified into several subtypes with distinct DNA sequence features, including long terminal repeat (LTR) elements such as the *Gypsy* superfamily and the *Copia* superfamily, long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs). The Class II TEs, referred to as DNA transposons, transpose via a "cut and paste" mechanism with a DNA as an intermediate, they can be further grouped into several superfamilies based on their transposase proteins and sequence structures, such as *CACTA*, *Mutator*, *Tc1/Mariner*, et al.

In addition, a new type of DNA transposon, *Helitron*, can transpose via the "rolling circle replication" model [11]; another type of complex Class II DNA transposons, called *Polintons*, transposing in a self-synthesizing manner has been found in several eukaryotic species, like protists, fungi, and animals [12].

The jumping of TEs frequently disrupts genome structures, resulting in genome instability, and induces some deleterious mutations in the host genome [13]. The host genomes have evolved multiple mechanisms to regulate TE transposition, including enzymatic systems for elimination of TEs and epigenetic mechanisms for silencing TE activities, such as DNA methylation, histone modifications, and RNA interference (RNAi) [14–17]. However, accumulating evidence has shown that a subset of TEs can be activated under certain conditions, such as in a specific tissue or developmental stage or in responses to external stresses, and they can play vital roles in the regulation of gene transcription, including generation of novel or abnormal transcripts, formation of novel *cis*-regulatory elements (CREs) such as alternative promoters and enhancers essential for the regulation of gene transcription, and alteration of chromatin features of nearby genes, thereby being involved in various biological processes [18,19]. These findings indicate that TE transcription leads to complex outcomes in the host genome.

Open chromatin, also referred to as accessible chromatin, represents unique chromatin regions with less nucleosome occupancy or nucleosome depletion (nucleosome-free regions) [20]. It usually harbors intricate CREs, such as promoters, enhancers, silencers, and insulators, which are essential for the modulation of gene transcription through interactions with various *trans*-acting factors during growth and development and in response to environmental cues in mammals and plants [21,22]. Open chromatin has been extensively investigated in eukaryotic genomes, including yeast, mammals, humans, and plants [23–27]. For instance, chromatin openness exhibits a positive correlation with the expression of corresponding genes; 25% of open chromatin sites are located in the promoters of rice genes [23]. Moreover, distal regulatory elements (DREs) in open chromatin regions have been found to be dynamic across different tissues and associated with tissue-specific gene expression in rice [28].

Accumulating evidence has shown that TEs can potentially contribute to the formation of open chromatin owing to partially/fully overlapping open chromatin in mammalian and plant genomes [26,29–31]. Forty-four percent of open chromatin regions are associated with TEs in human cells [29]. Over 20% of TEs overlapping open chromatin are potentially related to tissue development in mouse [31]. Similarly, TE-derived open chromatin has been reported to regulate gene transcription [30], and three TE families overlapping DNase I hypersensitive sites (DHSs) can function as enhancer candidates in maize [26]. Some TEs located in the distal regions of genes contain most of the TF binding sites (TFBSs) in wheat, and their expansion is associated with the responses of wheat-specific genes to environmental stimuli [32].

However, relationships between TEs and accessible chromatin regions (ACRs) haven't been well characterized in rice as compared with other plants like maize. To this end, we analyzed the dynamics of chromatin accessibility in six rice tissues, associated accessible chromatin regions (ACRs) with different TE superfamilies, and characterized the genomic features of corresponding TEs and ACRs in rice. We also investigated the role of TE-derived ACRs in the regulation of tissue-specific and duplicated gene expression. Our study provides some insights into the contribution of TEs in ACRs formation, thereby advancing our understanding of the biological implications of TE-derived ACRs in the rice genome.

2. Results

2.1. Identification of TE-Related Accessible Chromatin Regions in Rice

To identify TE-related accessible chromatin regions in rice, we used the published ATAC-seq (Assay for Transposase-Accessible Chromatin with high throughput sequencing) data derived from six tissues [33], including young leaf (YL), flag leaf (FL), root (RT),

stamen & pistil (SP), lemma & palea (LP), and young panicle (YP). The ATAC-seq is a robust technique widely applied for global profiling of chromatin accessibility through sequencing DNA fragments preferentially tagged by Tn5 transposase dimer. To minimize the impact of sequencing depth on peak calling, we down-sampled ATAC reads to the same amount for each tissue, and identified 43,988, 40,899, 52,601, 55,363, 51,395, and 44,884 ACRs corresponding to YL, FL, RT, SP, LP, and YP, respectively (Figure S1). All ACRs were further divided into three subgroups, high, middle, and low accessibility, according to the density of normalized read counts (RPKM value, reads per kilobase per million mapped reads), reflecting the degree of chromatin accessibility. We then counted TE number across ±2 kb of ACRs with different accessibility levels, and found that the normalized TE number was negatively associated with the level of chromatin accessibility in each tissue, even though the overall chromatin openness around TEs was dramatically lower than random (Figures 1A and S2). After a closer examination, we found that the percentage of ACRs with at least 1 bp overlapping TEs ranged from 26.2% in SP to 33.4% in YL. Surprisingly, we found that nearly 90% of the ACRs were located within 1 kb of the nearest TEs (Figure 1B) and exhibited less variations in the chromatin accessibility levels. In contrast, the chromatin accessibility of ACRs located more than 1 kb away from the nearby TEs increased with increasing distance to nearby TEs (Figure 1C), indicating that TE presence does not favor chromatin openness. After looking into TE types, we found that the majority of the TEs overlapping with ACRs were DNA transposons (24.4~27.9%) and the rest (3.7~5.5%) were retrotransposons (Figure 1B).

Next, we calculated the coverage ratio of ACRs by TE sequences and found that only a small proportion of ACRs were fully covered by TEs (Figure S3). For instance, DNA transposons *hAT* and *Helitron* covered ca. 25% of the full ACR length (the length of ATAC peak) on average, while *PIF_Harbinger* and *Tcl_Mariner* exhibited the least coverage on ACRs. This was possibly caused by the short length of their DNA sequences. Most *Gypsy* can fully cover the overlapping ACRs. These results showed that most ACRs covered the junctional regions of TE/non-TE DNA, which is similar to the findings in mouse [34]. A similar trend was observed for the coverage of the ATAC peak summits by the overlapping TEs (Figure 1D). There were 4288–5380 ACRs (accounting for 8.3~12.0% of all ACRs) with their summits overlapping with TEs (Figure 1D), and *Mutator*, *hAT*, and *Gypsy* exhibited more coverage of ACR summits than others (Figure S4).

Taken together, these results indicate that some TE sequences, in combination with their flanking non-TE DNA sequences, are able to contribute to the formation of ACRs, implying potential regulatory roles of TEs in rice.

### 2.2. Profiling of TEs in Proximal and Distal ACRs

It was reported that ACRs harbor *cis*-regulatory elements (CREs), which can be derived from TE sequences [21]. To explore whether TEs are also associated with functional ACRs in the rice genome, we merged ACRs from six tissues and obtained a total of 83,087 ACRs (Table S1). We then divided them into four subtypes according to their genomic loci, including 38,676 ACRs (46.5%) located within 2 kb upstream of the TSSs of genes (hereafter designated as promoter ACRs, pACRs); 14,622 ACRs (17.6%) located >2 kb from their nearest genes (hereafter designated as distal ACRs, dACRs); 16,814 ACRs (20.3%) overlapping with 2 kb downstream of the TTSs of the genes (hereafter designated as downstream ACRs, dnACRs); and the remaining 12,975 (15.6%) ACRs overlapping gene body regions (hereafter designated as genic ACRs, gACRs) (Figure S5). We then calculated the TE number across ±2 kb around the center of the pACRs and dACRs (Figure 2A). As shown in Figure 2A, the majority of TEs tended to be absent in the center of ACRs, and the dACRs exhibited a higher normalized TE number than the pACRs. Unlike other TE superfamilies, the short elements *PIF_Harbinger* and *Tcl_Mariner* were significantly enriched in the ACR boundary. And the number of *Mutator* and *hAT* was higher in the center region of pACRs compared with dACRs. (Figure 2A). These analyses showed that TEs exhibited a TE-type dependent distribution in or near the ACRs.
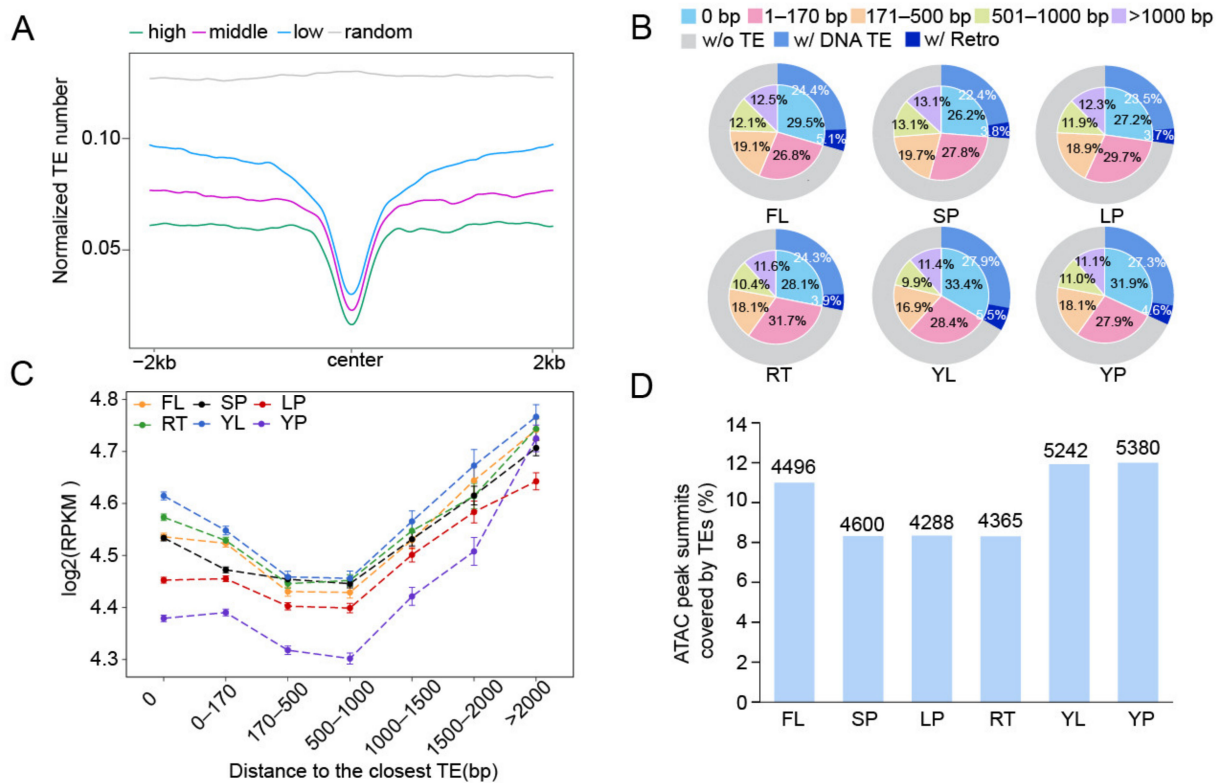
Figure 1. The TE-overlapping ACRs (accessible chromatin regions) in rice. (**A**) Curve plot showing normalized TE number across ±2 kb of ACR center (in flag leaf). All ACRs were divided into three subgroups (high, mid, low) according to their levels of accessibility. The same number of randomly selected regions were used as control. (**B**) Pie plot showing the ratio of ACRs with different distances to the nearest TEs (0 bp, representing ACRs with at least 1 bp overlap with TEs, 1~170 bp, 171~500 bp, 501~1000 bp, >1000 bp) in six tissues (FL, flag leaf; SP, stamen & pistil; LP, lemma & palea; RT, root; YL, young leaf; YP, young panicle), 26.2% (SP) to 33.4% (YL) ACRs had at least 1 bp overlap with TEs (*w/o* represents "without", *w/* represents "with"). (**C**) The accessibility of ACRs with different distances to the nearest TEs in six tissues, the reads of ATAC-seq within ACRs were normalized by reads per kilobase per million mapped reads (RPKM). (**D**) Bar plots illustrating the ratio and number of ACRs with their summits covered by TEs.

It is well known that TEs are hypermethylated, while open chromatin is hypomethylated [35], driving us to assess the DNA methylation levels of TE-derived ACRs. To this end, we analyzed public BS-seq data derived from young rice leaves [28]. After calculating the DNA methylation levels in each cytosine context in ±2 kb flank regions and bodies of related TEs, we observed that the TE body was hypermethylated relative to the flanking regions in CG, CHG and CHH context (Figure 2B), with the methylation level of ACR-related TEs showing a significant decrease than that of TEs without ACRs in TE bodies (Figure 2B). In addition, we also found that TEs with ACRs were significantly enriched with active histone mark (H3K4m3 and H3K27ac) in their bodies than that of TEs without ACRs, while the repressive mark H3K27me3 was more enriched in the flanking regions of TEs without ACRs, this further confirmed the relatively open state of TEs with ACRs (Figure S6). In contrast, the TE-derived ACRs were hypomethylated in their centers, and the CHH context was enriched in the region next to the center (Figure 2C), and the TE-derived dACRs showed a higher methylation level than pACRs. Corresponding to the methylation variations, we found the TE-derived pACRs were more open than that of dACRs (Figure S7). Thus, TEs exhibited an overall negative association between DNA methylation and accessible regions, TE-derived ACRs were prone to form in the hypomethylated regions within TEs. For instance, a *Copia* retrotransposon-derived ACR and a *Helitron* DNA transposon-derived

ACR showed a hypomethylation level as compared to surrounding areas (Figure 2D,E). These results indicated that DNA methylation levels might be key determinants of the formation of TE-derived ACRs.
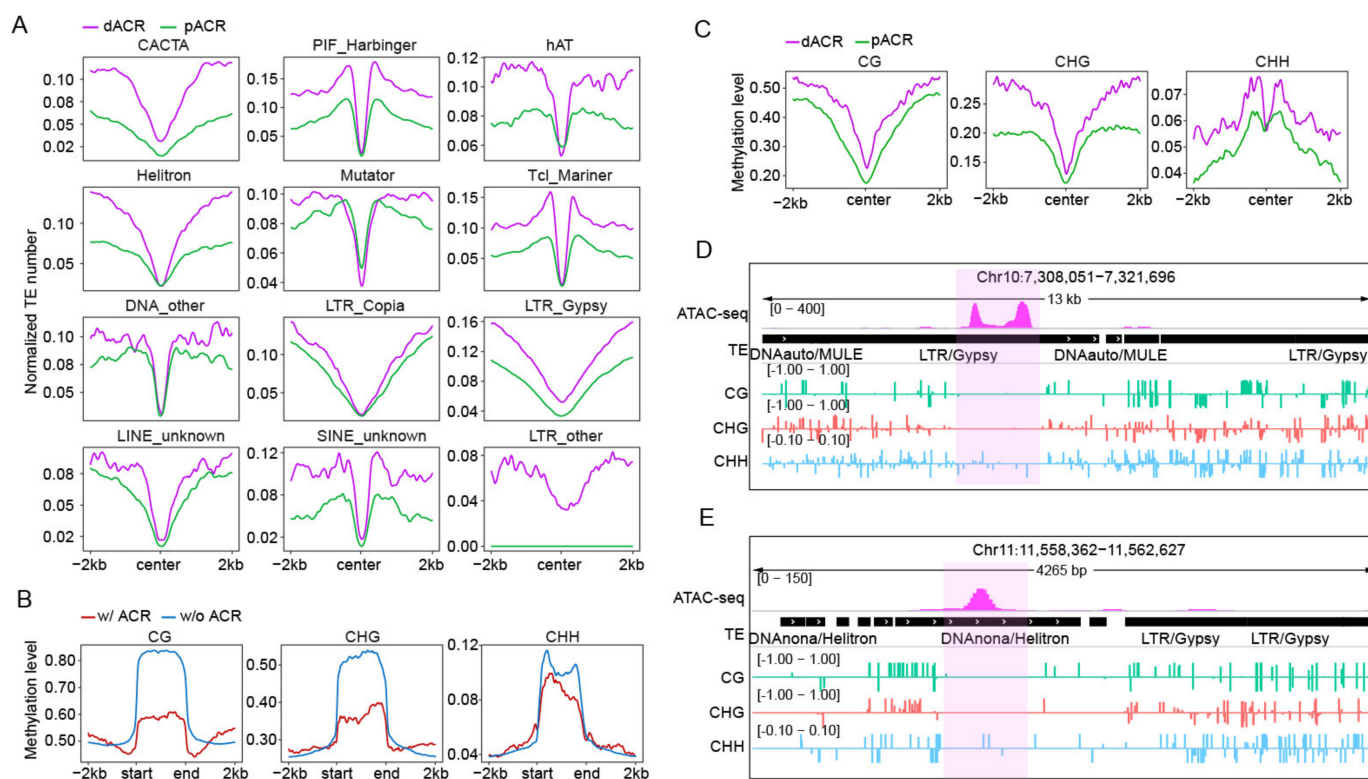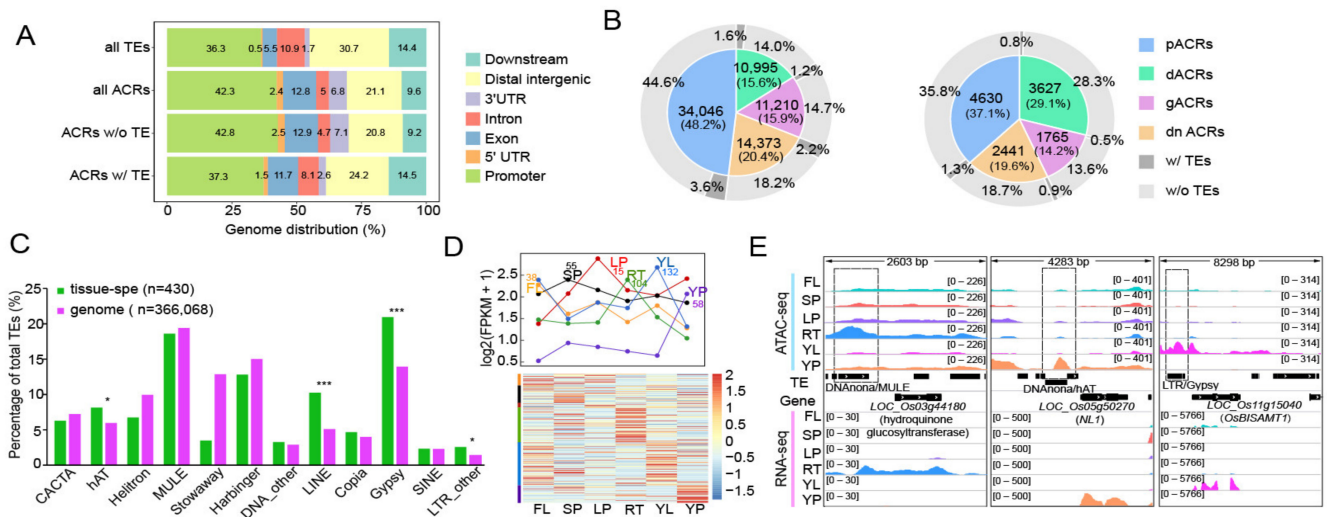

Figure 2. Distributions of TEs around ACRs and the methylation levels of corresponding ACRs and TEs. (**A**) Curve plots showing normalized TE number across ±2 kb of the center of pACRs (promoter ACRs) and dACRs (distal ACRs). (**B**) Average methylation level (CG, CHG, CHH) distribution over TEs with (*w/*) ACRs and without (*w/o*) ACRs. (**C**) The methylation level (CG, CHG, CHH) across ±2 kb of the center of TE-derived pACRs and dACRs. (**D**,**E**) IGV snapshots showing the methylation level around a *Copia* retrotransposon-derived ACR (**D**) and a *Helitron* DNA transposon-derived ACR (**E**), the ACRs were hypomethylated while the other regions located in TEs were hypermethylated.

Taken together, these results demonstrate that the differential distributions of TE types around dACRs and pACRs may be caused by methylation changes in the underlying DNA sequences.

2.3. Involvement of TE-Derived ACRs in the Regulation of Tissue-Specific Gene Expression

ACRs have been reported to play vital roles in the regulation of gene expression in eukaryotic genomes by interacting with *trans*-acting factors [36,37]. However, the role of TE-derived ACRs in the regulation of rice gene transcription remains unclear. To address this, we calculated the genomic distributions of TEs and ACRs and found that all ACRs and non-TE-derived ACRs tended to be located in genic regions, especially in promoters. In contrast, TEs and TE-derived ACRs were distributed more in the distal intergenic regions and downstream of genes (Figure 3A), which is consistent with previous findings in mammals that TE-derived ACRs can act as potential enhancers [34]. We then divided all ACRs as either common (*n* = 70,624) or tissue-specific (*n* = 12,463) ACRs, according to the entropy values (Figure S8 and Table S2). Compared with common ACRs, tissue-specific ACRs were distributed more in the distal regions (Figure 3B). Moreover, we found that less tissue-specific ACRs were TE-derived (3.5%), including 1.3% of pACRs, 0.8% of dACRs, 0.5% of gACRs, and 0.9% of dnACRs (Figure 3B), while more TE-derived ACRs were

found in the common ones (8.6%), these TE-derived tissue-specific ACRs were mainly enriched in *hAT*, *LINE*, *Gypsy* and other *LTR* retrotransposons, as compared to the whole genome background (Figure 3C). After calculating the fold-change in accessibility levels for tissue-specific ACRs in six tissue types, we found a similar variation between TE and non-TE-derived tissue-specific ACRs (Figure S9), suggesting that TE-derived ACRs can also be involved in tissue development like regular ACRs. To confirm this, we calculated the expression levels of TE-derived ACRs nearby genes in six tissues, and found that a subset of genes exhibited a similar trend to the changes in chromatin accessibility levels (Figure 3D). As shown in Figure 3E, some TE-derived tissue-specific ACRs were located at the upstream of subsets of functional tissue-specific genes. For example, a *MULE*-derived ACR was located in the promoter of a hydroquinone glucosyltransferase encoding gene (*LOC_Os03g44180*), preferentially expressed in root (RT); a *hAT* transposon related ACR was located near the TSS of *NL1* with higher expression levels in young panicle (YP), which is a GATA type transcription factor related to the panicle development [38]; a *Gypsy* retrotransposon-derived ACR was located in the upstream region of *OsBISAMT1*, a gene encoding S-adenosyl-L-methionine related to defense responses, which was specially expressed in young leaf (YL) [39].


Figure 3. The association between TE-derived tissue-specific ACRs and tissue-specific gene expression.(**A**) Sub-genomic distributions of all TEs, all ACRs, ACRs without (*w/o*) TE and ACRs with (*w/*) TEs, including promoters (upstream 2 kb), exons, introns, downstream (2 kb) and distal intergenic regions. (**B**) Pie plot showing the ratio of different types of common (**left**) and tissue-specific ACRs (**right**) based on their distances to the nearest genes, 8.6% and 3.5% of these ACRs were TE-derived ACRs (>50% ACRs length covered by TE sequences), respectively. (**C**) Bar plot illustrating the percentage of TE superfamilies associated with tissue-specific ACRs and the percentage of TE super-families in the whole genome, significance test was determined using hypergeometric test, * $p < 0.05$, *** $p < 0.001$. (**D**) The expression pattern of the nearest genes of TE-derived tissue-specific ACRs, the lineplot (upper) showing the mean expression levels of genes in each tissue, each row in the heatmap (bottom) represents a gene expression variation in six tissue types, the color represents gene expression levels. The number represents the amount of genes in each tissue. (**E**) IGV snapshots showing expressed tissue-specific genes with tissue-specific TE-derived ACRs.

Collectively, our results show that TE-derived ACRs may play vital roles in the regulation of tissue-specific gene expression, thereby being involved in the tissue development in rice.

2.4. Impacts of TEs on Chromatin Accessibility and Its Conservation

To investigate how the presence of TEs affects chromatin accessibility, we plotted normalized ATAC-seq read counts across ±2 kb of the center of ACRs with and without TEs (Figures 4A and S10), and found that non-TE-derived ACRs were more open than TE-derived ones. At the sequence level, we found that TEs containing ACRs had a longer length (Figure 4B). A similar trend was also observed in TEs superfamilies (Figure S11), and TEs containing ACRs had higher GC content, which facilitates nucleosome formation [40], as compared to inaccessible TEs (Figure 4C).
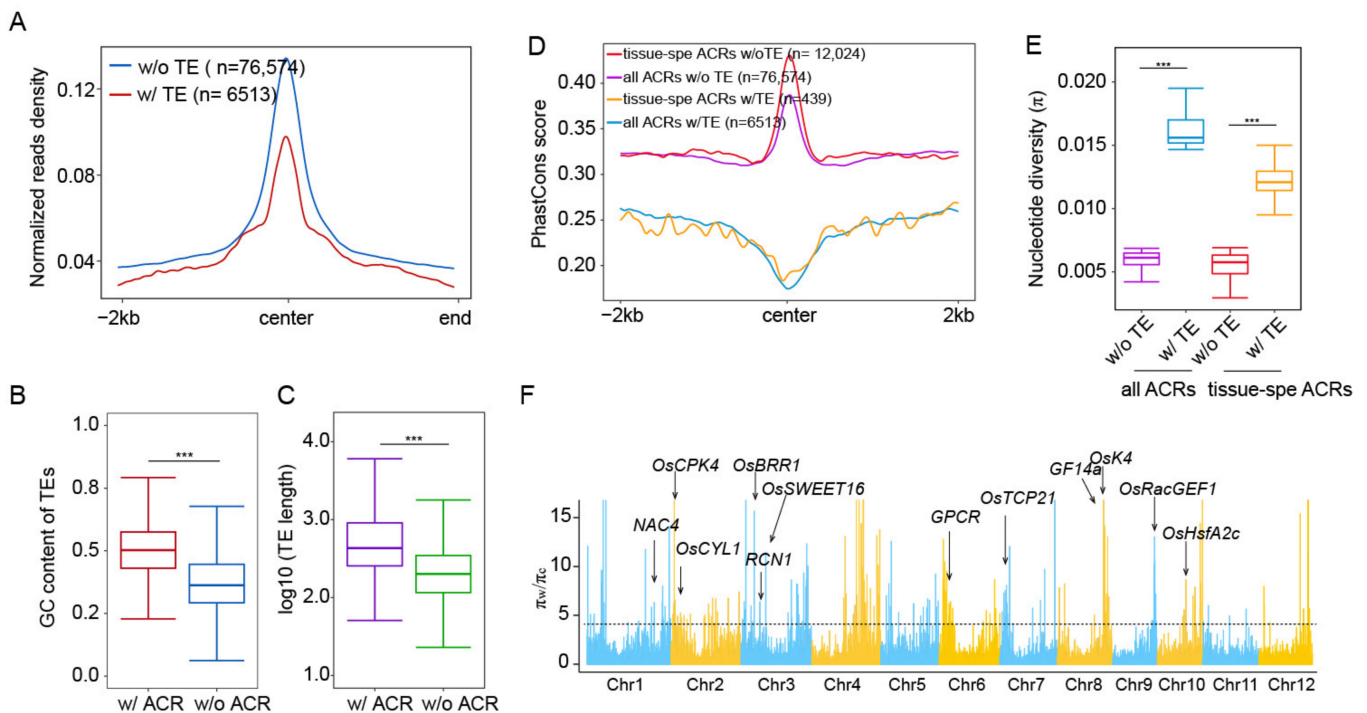
Figure 4. Genomic features and conservation of TE-derived ACRs. (**A**) Curve plot demonstrating normalized ATAC-seq read density across ±2 kb of the center of ACRs with ($w/$)/without (w/o) TEs in flag leaf (FL). (**B**) Boxplot showing the difference of GC content of TE (with/without ACRs). Significance test was determined using Wilcoxon rank sum test, *** $p < 0.001$. (**C**) Boxplot showing the difference of TE (with/without ACRs) length. Significance test was determined using Wilcoxon rank sum test, *** $p < 0.001$. (**D**) Averaged phastCons score across 4 kb regions centered by all ACRs and tissue-specific ACRs, with or without TEs. (**E**) Boxplot demonstrating mean π values around 2 kb region of ACRs with/without TEs. Significance test was determined using Wilcoxon rank sum test, *** $p < 0.001$. (**F**) Selection signals in the rice genome. The horizontal black dashed lines showing the genome-wide threshold for domestication sweeps ($\pi_w/\pi_c > 4.1$). Functional genes overlapped with TE-derived ACRs under domestication selection were labeled in corresponding chromosomes.

It is well known that the transposition activity of TEs potentially leads to various genomic mutations during evolution [41]. To explore the conservation of TE-derived ACRs during rice evolution, we compared the PhastCons score (representing conservation landscapes among 63 different plant species) for different types of ACRs [42], including the tissue-specific ACRs shown in Figure 3B. We found that TE-derived ACRs had much lower PhastCons scores than non-TE-derived ACRs, especially in the center regions of ACRs, and tissue-specific ACRs showed significantly higher PhastCons scores than all ACRs (Figure 4D). These results indicated that most of TE-derived ACRs were less conserved at the sequence level than the non-TE-derived ones and tended to be species-specific during plant evolution; in contrast, tissue-specific ACRs, exhibited a high conservation between different plant species, which may be caused by directional selection. We then calculated the π values, which can reflect the nucleotide diversity within different rice

varieties, for the ACRs shown in Figure 4D, and found that TE-derived ACRs had higher π values, while tissue-specific ACRs had much lower π values than all ACRs (Figure 4E), indicating that conservation changes for different types of ACRs are highly associated with the presence of TE during rice domestication. Domestication usually leads to adaptive changes in rice. To determine whether TE-derived ACRs are also involved in domestication, we identified genomic regions with notable decreases in nucleotide diversity by comparing the π values between the wild (π_w) and cultivated rice accessions (π_c). According to π_w/π_c values, we identified 187 putative domestication sweeps, containing 270 TE-derived ACRs (~4%) (Figure 4F). Indeed, we found that in these domestication regions, some genes essential for the rice development were associated with TE-derived ACRs, including flower development gene *OsK4* [43], hormone-related gene *OsCYL1* [44], and several disease/stress resistance genes (*OsCPK4, OsBRR1, RCN1, OsSWEET16, GPCR, OsTCP21, GF14a, OsRacGEF1, OsHsfA2c*) [45–53].

Taken together, these results indicate that TE-derived ACRs may have coevolved with nearby genes during artificial and natural selection, thereby functioning in the formation of favorable agronomic traits in rice.

2.5. Relationship between TE-Derived ACR around TSS and Duplicated Genes

Gene duplication is one of the major driving forces for genome complexity and gene subfunctionalization or neofunctionalization during plant polyploidization and evolution [54–56]. Growing evidence suggests that epigenetic mechanisms are involved in the regulation of duplicated gene transcription [57–61]. TEs have been found to affect the epigenetic status of duplicate genes [62], however, how TE-derived ACRs function in the regulation of duplicated genes in rice remains unclear. To answer this question, we chose ATAC-seq data from the stamen and pistil (SP) for analyses because it has the most ACRs among the six tissues. We plotted normalized ATAC-seq read counts across ±1 kb for all duplicated genes (Figure 5A and Table S3). We observed that differential read enrichment levels occurred among genes in different duplicate modes and a significant read abundance occurred around the TSSs (Figure 5A). Given that epigenetic features around the TSS usually play an important role in the regulation of gene expression, we specifically counted duplicated genes ($n = 561$) with TE-derived ACRs at ±500 bp of the TSS (Table S4) and observed a similar trend in reads enrichment between these duplicated genes and all genes shown in Figure 5A. Chromatin was more accessible for the WGD genes, whereas it was much less accessible for the proximal and tandem duplicated genes (Figure 5B), and the variation between different duplicate modes was smaller in genes related to TE-derived ACRs, which may cause by the fewer number of related genes and lower accessibility of TE-derived ACRs. Since these 561 duplicated genes corresponded to 1070 duplicated gene pairs in the rice genome (Table S5), we compared the expression level and chromatin accessibility of these gene pairs and found that the change in chromatin accessibility exhibited a positive correlation with the expression change of gene pairs (Figure 5C), indicating that the chromatin dynamics of TE-derived ACRs are involved in the regulation of gene pair expression.

TEs can act as potential novel regulatory sequences in the host genome, and TE inserted near genes can be selected during evolution [2]. To identify TE-related regulatory sequences of duplicated genes, we divided all duplicated gene pairs with TE-derived ACRs around the TSSs into three subgroups: subgroup 1, duplicated gene pairs with both copies of TE-related ACRs around the TSSs ($n = 29$); subgroup 2, duplicated gene pairs with TE-related ACRs around the TSSs for one copy and regular ACR for the other copy ($n = 532$); subgroup 3, duplicated gene pairs with only one copy having TE-related ACRs around the TSSs ($n = 509$) (Figure 5D). After conducting novel motif enrichment assays, we found that the majority of motifs overrepresented in ACRs were "G"- or "A"-rich sequences. In particular, we found that the motif for the binding of the TCP transcription factor was overrepresented in TE-related ACRs instead of regular ACRs (Figure S12), indicating that TE insertion can introduce new regulatory DNA sequences, which may be involved

in the modulation of gene pair expression. Furthermore, we examined the differential distributions of *Ks* (synonymous substitutions per synonymous site) among duplicated gene pairs and observed the descending order of *Ks* distribution changes among duplicated genes in three subgroups: subgroup 1 < subgroup 3 < subgroup 2 (Figure 5E), indicating a distinct TE-derived ACRs distribution between younger (lower *Ks* value) and older gene duplicates (higher *Ks* value). To assess whether genes with the absence/presence of TE-related ACRs around TSSs were subjected to different intensities of selective constraints, we calculated *Ka* (nonsynonymous substitutions per nonsynonymous site)/*Ks* values for gene pairs, and found that most of the *Ka/Ks* were less than 1 (Figure 5F). For subgroups of gene pairs, we found that subgroups 1 and 2 had the highest and lowest mean levels of *Ka/Ks*, respectively (Figure 5F).

These results suggest that TE-derived ACRs related gene pairs are primarily subjected to purifying selection, and a combination of TEs and ACRs is associated with the strength of the selection pressure.
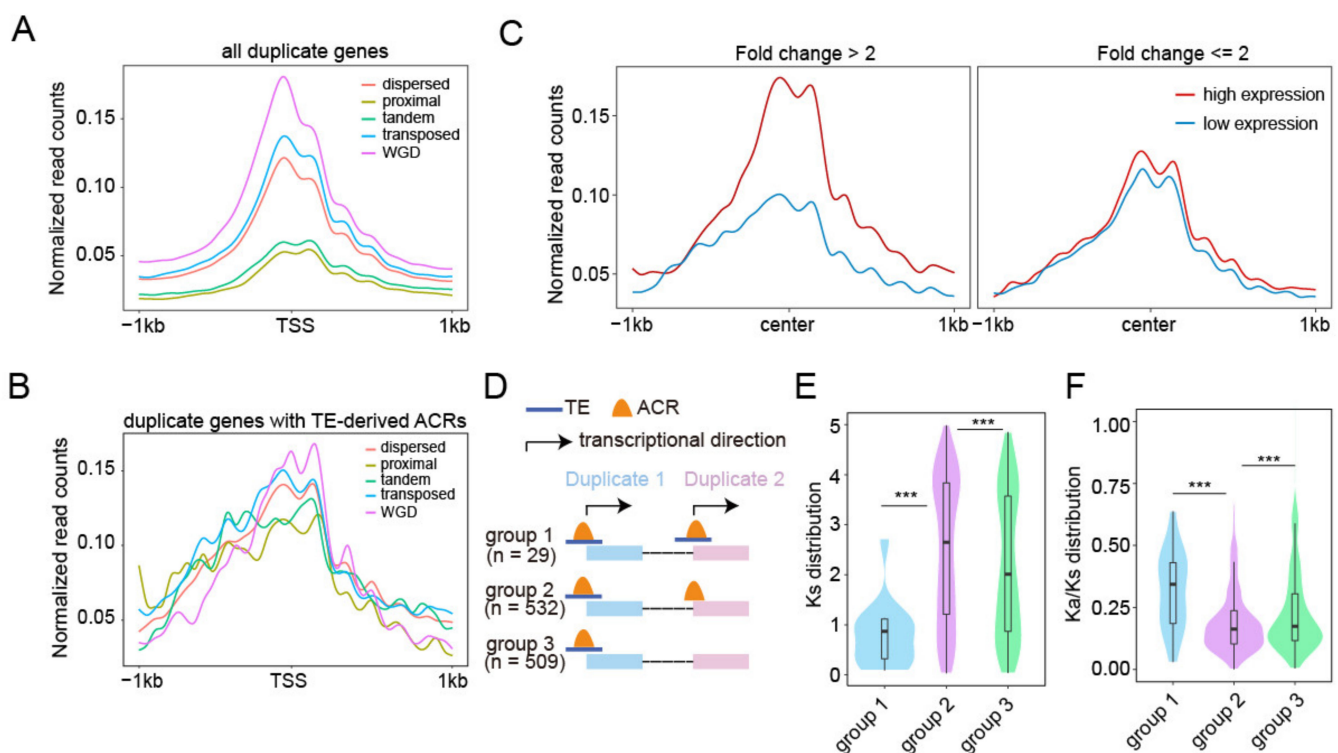
Figure 5. The chromatin accessibility for duplicated gene pairs. (**A**) Genome-wide normalized ATAC-seq read density of WGD, transposed, tandem, proximal, dispersed duplicated genes. (**B**) Normalized ATAC-seq read density of duplicated genes associated with TE-derive ACRs. (**C**) Correlation between chromatin accessibility changes and gene expression changes of duplicated gene pairs corresponding to TE-derive ACRs. All gene pairs were divided into two groups according to their expression differences: fold change > 2 and fold change <= 2, the chromatin accessibility variations showing greater differences in the fold change > 2 group. (**D**) Gene pairs with TE-derived ACRs around TSSs were divide into three subgroups. Only 29 pairs of duplicated genes had TE-derived ACRs around TSSs of both copies. (**E,F**) Comparisons of distributions of *Ks* (**E**) and *Ka/Ks* (**F**) between duplicated gene pairs in three groups are illustrated in (**D**). Significance test was determined using Wilcoxon rank sum test, *** $p < 0.001$.

3. Discussion

TEs can induce a wide range of genomic alterations, including chromosome rearrangement, epigenetic regulation, and insertion mutagenesis [63], resulting in genetic diversity and genome evolution. Various epigenetic mechanisms have been found to regulate TE activity in plants [64]. The footprint of TE could regulate gene expression by changes in

epigenetic modifications [2,65]. The function of TEs in the regulation of gene transcription can be mediated by the formation of open chromatin in maize [66] and mouse [31]. TEs account for up to 40% of the rice genome and its superfamilies have already been well characterized [63]. However, the mechanism by which TEs regulate gene transcription in rice remains understudied. Our study revealed several discoveries by associating TEs with open chromatin and epigenetic modifications.

Global identification of ACRs helps to mine functional CREs like promoters, enhancers, et al. Characterization of TE-derived ACRs helps to identify CREs associated with TEs, thereby advancing understanding of roles of TEs in the regulation of gene expression. Accumulating evidence shows variations in the formation of TE-derived CREs among different organisms. For instance, nearly 30% of TEs containing active elements, with 45% and 20% of them belonging to LTR retrotransposons and DNA transposons, respectively, were found in the classic model organism *D. melanogaster* [67]; 25% of TE-derived CREs were found in humans, and of which the LTRs contain almost all of the TFBSs [68,69]; About 20% of open chromatin regions associated with TEs were found in mouse, which mainly consist of LTR and SINE retrotransposons [31]. Similarly, TE-derived ACRs have also been reported in plants. Over 57% and 26.7% TE-derived DHSs were found to be associated with LTRs, and DNA transposons in maize, respectively [30]. Our study showed that ~8% ACRs had more than 50% sequence covered by TEs in rice. It is worth noting that our analyses cannot distinguish if TE-derived ACRs are from TEs carrying ACRs or insertion of TEs into pre-existing ACRs within the genome. After a closer examination, we found that 45.5% of these ACRs were fully covered by TEs, which are most likely associated with TE-derived ACRs; the remaining ones were partially (50–99%) covered by TEs, which could be further divided into two groups, the length of TEs is shorter than that of ACRs, and the TEs overlapping with one edge of ACRs. Technically, it is not easy (nearly impossible) to distinguish the two distinct cases of TE insertion. In our opinion, our analyses indicate that as least parts of TEs if not all have the potentials to form ACRs, which has already been reported in other species [34]. Comprehensive comparisons of TEs and ATAC-seq data between different genomes help to identify polymorphic TE- and shared TE-derived ACRs between species. Our study showed that DNA transposons contributed to most of the TE-derived ACRs in rice, which is different from the findings reported in aforementioned organisms. This is possibly caused by characteristics of rice genome. The copy number of DNA transposons far exceed that of retrotransposons, while most retrotransposons are longer and prone to locate in distal regions of genes. The insertion of some short DNA transposons could influence the activities of nearby genes. We found the majority of retrotransposons contribute to ACRs in rice were *LTR_Copia*, it was reported some solo-LTRs tend to keep TF-specific CREs [70], indicating that functions of these LTRs in the regulation of gene expression can be mediated through distal interactions between CREs and TFs or *trans*-acting factors. Variation of TE-derived CREs could reflect differentiation of TE activities among different organisms during genome evolution.

The degree of chromatin openness correlated negatively with the density. This is possibly caused by the overall hypermethylation of TEs in plant genomes [71–73], while hypomethylation favors the formation of open chromatin [35]. Consistent with findings about TEs sequestered far from genes derived by purifying selection and genetic drift during evolution [74], a much higher normalized TE number was observed around dACRs than around pACRs in rice (Figure 2A). This provided evidence showing TE subtype-dependent effects on the formation of ACRs, which is possibly mediated by changes in local DNA methylation. A similar finding was reported in mouse [34], indicating that the differential functions of TE superfamilies in ACRs formation are conserved between plants and mammals.

CREs embedded in open chromatin regions can be activated in a tissue-specific manner [27,31,75,76]. Our study provided evidence that the connections between TEs and tissue development can be mediated by the formation of TE-derived ACRs that regulate tissue-specific gene expression. For example, the known functional gene *NL1* and

OsBISAMT1 were found to have tissue-specific TE-derived ACRs in their promoters (Figure 3E). Similar findings have been reported in humans [77–79]. This implies that the chromatin status in TE-derived ACRs could be dynamic in order to temporally regulate nearby gene expression during tissue development or in response to external environmental cues. In addition, if possible, analyses of TEs among different tissues of a single plant or between individual plants help to identify tissue or individual plant related polymorphic TE insertions, which could help to address variations of TE functions at the population level or a tissue wide.

Chromatin accessibility is positively correlated with the conservation of accessible chromatin regions [80]. Our study showed that TE-derived ACRs were involved in rice domestication, as evidenced by the following: TE-derived ACRs were more accessible than ACRs without TEs; TE-derived ACRs exhibited low conservation and high nucleotide diversity; and more importantly, domestication sweep regions of rice had some agronomic trait-related genes with TE-derived ACRs (Figure 4), indicating that these ACRs-related TE insertions in rice may favor co-selection and domestication of key agronomic traits as opposed to having detrimental impacts on the genome. The selection is also related to the epigenetic status of the TEs. For instance, hypermethylated TEs that insert nearby genes in *Arabidopsis* are strongly selected against, ultimately resulting in elimination of related TEs [2].

Expression divergence of duplicated genes results in the neo- or subfunctionalization of duplicated genes [81], which can be regulated by changes in epigenetic modifications, such as DNA methylation and histone modifications [57,58,60]. It has been reported that TEs can affect the evolutionary divergence of duplicated genes through changes in epigenetic status [62]. Our study indicated that TE-derived ACRs were involved in the divergence of duplicated gene pairs through the regulation of their expression. This was evidenced as follows: divergence of TE-derived ACRs was positively correlated with the expression changes of duplicated genes, which could be mediated by divergence of the underlying regulatory sequences [82]; gene pairs with both copies having TE-derived ACRs tended to be younger than others; and duplicated gene pairs having a TE-derived ACR around the TSSs for one copy and a regular ACR for the other copy (subgroup 2) experienced stronger purifying selection than those with both copies having TE-derived ACRs around the TSSs (subgroup1) (Figure 5). Young gene pairs have similar TE environments [62]. The TE insertions affect the selection strength of nearby genes [83,84]. TE-free genes have stronger purifying selection than TE-rich genes [84], indicating that the presence of TEs has a potential impact on the selection of subsets of genes.

4. Methods

4.1. Processing of Public RNA-Seq, ATAC-Seq and ChIP-Seq Data

Public RNA-seq and ATAC-seq data with high quality from six rice tissues [33], and ChIP-seq data from young leaves [28] of rice variety Zhenshan97 were used in this study. Fastp (v0.21.0) [85] was used to filter the raw reads. Any reads with low-quality values (Q < 25) and short lengths (<50 bp) were excluded from further analyses. Clean reads of both data were mapped to the rice MSU7.0 reference genome as the previous study did [33] using Hisat (v2.1.0) [86] and Bowtie2 (v2.2.5) aligner [87], respectively. For RNA-seq data, featureCounts [88] was used to count the reads mapped to the rice genes. A custom R script was used to calculate the fragments per kilobase million (FPKM) values, representing the expression level of each gene.

For the ATAC-seq and ChIP-seq data, PCR duplicates were removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard/, accessed on 22 March 2022). Samtools (v1.9) [89] was used to remove aligned reads with a mapping quality (MapQ) less than 30, followed by the conversion of bam files to bigwig files using deeptools (v3.1.3) [90] based on the fragments per kilobase per million reads mapped (RPKM) normalization method. The read distributions of both datasets across the rice genome were visualized using IGV (v2.4.13) [91].

4.2. BS-Seq Data Analyses

Public BS-seq data generated from young leaves of Zhenshan97 were used for the DNA methylation assay [28]. The raw paired-end reads were filtered using fastp (v0.21.0) [85] with the following parameters: lengthrequired = 50 –q 15. Bismark (v0.23.0) [92] was used for clean read mapping with the default parameters. After removing PCR bias using the deduplicate_bismark module, all uniquely mapped reads were retained for further analyses. The extent of methylation of each cytosine site was extracted using the bismark_methylation_extractor module. Only cytosine sites covered by at least five reads were retained for the methylation assay. CGmaptools (v0.1.2) [93] was used to convert cytosine methylation reports to CGmap formats. A custom Python script was used to calculate the methylation levels for the designated genomic regions.

4.3. Association Analyses of ACRs with Other Genomic Loci

MACS2 software (v2.1.4) [94] was used to define ACRs with the following parameters: $-g$ 3.73134×10^8 –no model –shift -50 –extsize 100 –q 0.01. Bedtools (v2.29.0) [95] was used to correlate the ACRs with other genomic loci. ACR-related genes represent genes (from 2000 bp upstream of the TSS to 2000 bp downstream of the TTS) overlapping ACRs. The TE annotation file was obtained following the published pipeline [96] with rice repeat library (rice7.0.0.liban) downloading from https://github.com/oushujun/riceTELib (accessed on 20 May 2022). ACRs with more than 50% of their length covered by a TE sequence were defined as TE-derived ACRs, and the corresponding TEs were also considered as ACR-related TEs. A custom Python script was used to calculate the normalized ATAC read counts within the designated genomic regions and genes.

4.4. Identification of Tissue Specific ACRs

A total of 83,087 ACRs were obtained by merging all ACRs in each of the six rice tissues according to their genomic coordinates, using the merge function from bedtools (v2.29.0) [95]. The ATAC reads from six tissues located within these ACRs were counted and normalized to reads per million mapped reads (RPM). The RPM value of each ACR across the six tissues was used to identify the tissue specificity. By using the Shannon entropy method [97,98], entropy values of all ACRs were arranged in an ascending order and the top 15% of ACRs were considered as tissue-specific ACRs.

4.5. Estimation of the PhastCons Scores and π Values for ACRs

The evolutionary conservation of ACRs was measured using phastCons conservation scores downloaded from PlantRegMap (http://plantregmap.gao-lab.org/cis-map.php, accessed on 8 June 2022), which were calculated by multiple genome alignments of 63 different plants [42]. Vcftools (version 0.1.16) [97] was used to calculate π values for the cultivated and wild rice groups. The relative vcf format files and resequencing data were downloaded from the Rice SNP-Seek database (https://snp-seek.irri.org/_download.zul, accessed on 8 June 2022). For the wild rice, the resequencing data of 5 *O. rufipogon* plants were obtained from the public literatures [99]. For identification of domestication sweeps, π values were calculated using 100 kb windows, and regions with the top 5% π ratio (π_w/π_c) were set as thresholds for putative domestication sweeps, as described in the previous study [100].

4.6. Identification of Rice Duplicated Genes and Calculation of Ka/Ks Values

The different modes of duplicated genes were identified using the *DupGen_finder* pipeline (https://github.com/qiao-xin/DupGen finder, accessed on 10 June 2022) [101]. Briefly, the protein sequences annotated by the Rice Genome Annotation Project (http://rice.uga.edu, accessed on 10 June 2022) were blasted against themselves with an E-value $< 1 \times 10^{-10}$. The collinear blocks and syntenic gene pairs were identified using MCScanX [102] with default parameters. The "duplicate_ gene_ classifier" tools in MCScanX were used to classify genes into different duplicate modes. The Perl script

DupGen_finder.pl in *DupGen_finder* was used to match duplicated gene pairs, and the transposed duplicated gene pairs were detected using the *Zea mays* genome as their outgroups. *Ka*, *Ks* and *Ka/Ks* values of each duplicated gene pair were calculated following the public pipeline on GitHub (https://github.com/qiao-xin/Scripts_for_GB, accessed on 10 June 2022). *Ks* values > 5.0 were excluded from further analyses because of the saturated substitutions at synonymous sites [103]. Gene pairs with *p* value > 0.05 were also removed.

4.7. Motif Enrichment Analyses

The sequences of designated ACRs were used to identify motifs using the MEME-ChIP module of the MEME Suite Programs (https://meme-suite.org/meme/tools/meme-chip, accessed on 12 June 2022) [104,105], with DAP motifs [106] as known motif datasets. Motifs with E values less than 0.01 were defined as significantly enriched motifs.

References

1. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 344–355. [CrossRef] [PubMed]
2. Bennetzen, J.L.; Wang, H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **2014**, *65*, 505–530. [CrossRef] [PubMed]
3. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A.; et al. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **2009**, *337*, 1040. [CrossRef] [PubMed]
4. Wicker, T.; Gundlach, H.; Spannagl, M.; Uauy, C.; Borrill, P.; Ramirez-Gonzalez, R.H.; de Oliveira, R.; International Wheat Genome Sequencing Consortium; Mayer, K.F.X.; Paux, E.; et al. Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biol.* **2018**, *19*, 103. [CrossRef] [PubMed]
5. Cheng, H.; Sun, G.; He, S.; Gong, W.; Peng, Z.; Wang, R.; Lin, Z.; Du, X. Comparative effect of allopolyploidy on transposable element composition and gene expression between *Gossypium hirsutum* and its two diploid progenitors. *J. Integr. Plant Biol.* **2019**, *61*, 45–59. [CrossRef]
6. Naito, K.; Zhang, F.; Tsukiyama, T.; Saito, H.; Hancock, C.N.; Richardson, A.O.; Okumoto, Y.; Tanisaka, T.; Wessler, S.R. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **2009**, *461*, 1130–1134. [CrossRef]
7. Buchmann, R.C.; Asad, S.; Wolf, J.N.; Mohannath, G.; Bisaro, D.M. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J. Virol.* **2009**, *83*, 5005–5013. [CrossRef]
8. Ito, H.; Gaubert, H.; Bucher, E.; Mirouze, M.; Vaillant, I.; Paszkowski, J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **2011**, *472*, 115–119. [CrossRef]

9. Butelli, E.; Licciardello, C.; Zhang, Y.; Liu, J.; Mackay, S.; Bailey, P.; Reforgiato-Recupero, G.; Martin, C. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* **2012**, *24*, 1242–1255. [CrossRef]

10. Agren, J.A.; Clark, A.G. Selfish genetic elements. *PLoS Genet.* **2018**, *14*, e1007700. [CrossRef]

11. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [CrossRef] [PubMed]

12. Kapitonov, V.V.; Jurka, J. Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4540–4545. [CrossRef] [PubMed]

13. Hedges, D.J.; Deininger, P.L. Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat. Res.* **2007**, *616*, 46–59. [CrossRef] [PubMed]

14. Bucher, E.; Reinders, J.; Mirouze, M. Epigenetic control of transposon transcription and mobility in *Arabidopsis*. *Curr. Opin. Plant Biol.* **2012**, *15*, 503–510. [CrossRef]

15. Cui, X.; Cao, X. Epigenetic regulation and functional exaptation of transposable elements in higher plants. *Curr. Opin. Plant Biol.* **2014**, *21*, 83–88. [CrossRef]

16. Fultz, D.; Choudury, S.G.; Slotkin, R.K. Silencing of active transposable elements in plants. *Curr. Opin. Plant Biol.* **2015**, *27*, 67–76. [CrossRef]

17. Bousios, A.; Gaut, B.S. Mechanistic and evolutionary questions about epigenetic conflicts between transposable elements and their plant hosts. *Curr. Opin. Plant Biol.* **2016**, *30*, 123–133. [CrossRef]

18. Hirsch, C.D.; Springer, N.M. Transposable element influences on gene expression in plants. *Biochim. Biophys. Acta Gene Regul. Mech.* **2017**, *1860*, 157–165. [CrossRef]

19. Sahebi, M.; Hanafi, M.M.; van Wijnen, A.J.; Rice, D.; Rafii, M.Y.; Azizi, P.; Osman, M.; Taheri, S.; Bakar, M.F.A.; Isa, M.N.M.; et al. Contribution of transposable elements in the plant's genome. *Gene* **2018**, *665*, 155–166. [CrossRef]

20. Zhang, W.; Zhang, T.; Wu, Y.; Jiang, J. Open chromatin in plant genomes. *Cytogenet. Genome Res.* **2014**, *143*, 18–27. [CrossRef]

21. Klemm, S.L.; Shipony, Z.; Greenleaf, W.J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **2019**, *20*, 207–220. [CrossRef] [PubMed]

22. Wong, D.C.J.; Lopez Gutierrez, R.; Gambetta, G.A.; Castellarin, S.D. Genome-wide analysis of *cis*-regulatory element structure and discovery of motif-driven gene co-expression networks in grapevine. *DNA Res.* **2017**, *24*, 311–326. [CrossRef] [PubMed]

23. Zhang, W.; Wu, Y.; Schnable, J.C.; Zeng, Z.; Freeling, M.; Crawford, G.E.; Jiang, J. High-resolution mapping of open chromatin in the rice genome. *Genome Res.* **2012**, *22*, 151–162. [CrossRef] [PubMed]

24. Zhu, B.; Zhang, W.; Zhang, T.; Liu, B.; Jiang, J. Genome-wide prediction and validation of intergenic enhancers in *Arabidopsis* using open chromatin signatures. *Plant Cell* **2015**, *27*, 2415–2426. [CrossRef]

25. Zhang, T.; Marand, A.P.; Jiang, J. PlantDHS: A database for DNase I hypersensitive sites in plants. *Nucleic Acids Res.* **2016**, *44*, D1148–D1153. [CrossRef] [PubMed]

26. Oka, R.; Zicola, J.; Weber, B.; Anderson, S.N.; Hodgman, C.; Gent, J.I.; Wesselink, J.J.; Springer, N.M.; Hoefsloot, H.C.J.; Turck, F.; et al. Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* **2017**, *18*, 137. [CrossRef]

27. Sun, Y.; Dong, L.; Zhang, Y.; Lin, D.; Xu, W.; Ke, C.; Han, L.; Deng, L.; Li, G.; Jackson, D.; et al. 3D genome architecture coordinates trans and *cis* regulation of differentially expressed ear and tassel genes in maize. *Genome Biol.* **2020**, *21*, 143. [CrossRef]

28. Zhao, L.; Xie, L.; Zhang, Q.; Ouyang, W.; Deng, L.; Guan, P.; Ma, M.; Li, Y.; Zhang, Y.; Xiao, Q.; et al. Integrative analysis of reference epigenomes in 20 rice varieties. *Nat. Commun.* **2020**, *11*, 2658. [CrossRef]

29. Jacques, P.E.; Jeyakani, J.; Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **2013**, *9*, e1003504. [CrossRef]

30. Zhao, H.; Zhang, W.; Chen, L.; Wang, L.; Marand, A.P.; Wu, Y.; Jiang, J. Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol.* **2018**, *176*, 2789–2803. [CrossRef]

31. Miao, B.; Fu, S.; Lyu, C.; Gontarz, P.; Wang, T.; Zhang, B. Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biol.* **2020**, *21*, 255. [CrossRef] [PubMed]

32. Zhang, Y.; Li, Z.; Zhang, Y.; Lin, K.; Peng, Y.; Ye, L.; Zhuang, Y.; Wang, M.; Xie, Y.; Guo, J.; et al. Evolutionary rewiring of the wheat transcriptional regulatory network by lineage-specific transposable elements. *Genome Res.* **2021**, *31*, 2276–2289. [CrossRef] [PubMed]

33. Zhao, H.; Li, J.; Yang, L.; Qin, G.; Xia, C.; Xu, X.; Su, Y.; Liu, Y.; Ming, L.; Chen, L.L.; et al. An inferred functional impact map of genetic variants in rice. *Mol. Plant* **2021**, *14*, 1584–1599. [CrossRef]

34. Ye, M.; Goudot, C.; Hoyler, T.; Lemoine, B.; Amigorena, S.; Zueva, E. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 7905–7916. [CrossRef] [PubMed]

35. Crisp, P.A.; Marand, A.P.; Noshay, J.M.; Zhou, P.; Lu, Z.; Schmitz, R.J.; Springer, N.M. Stable unmethylated DNA demarcates expressed genes and their *cis*-regulatory space in plant genomes. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 23991–24000. [CrossRef] [PubMed]

36. Huang, M.K.; Zhang, L.; Zhou, L.M.; Yung, W.S.; Li, M.W.; Lam, H.M. Genomic features of open chromatin regions (OCRs) in wild soybean and their effects on gene expressions. *Genes* **2021**, *12*, 640. [CrossRef] [PubMed]

37. Schwope, R.; Magris, G.; Miculan, M.; Paparelli, E.; Celii, M.; Tocci, A.; Marroni, F.; Fornasiero, A.; de Paoli, E.; Morgante, M. Open chromatin in grapevine marks candidate CREs and with other chromatin features correlates with gene expression. *Plant J.* **2021**, *107*, 1631–1647. [CrossRef]

38. Wang, L.; Ming, L.; Liao, K.; Xia, C.; Sun, S.; Chang, Y.; Wang, H.; Fu, D.; Xu, C.; Wang, Z.; et al. Bract suppression regulated by the miR156/529-SPLs-NL1-PLA1 module is required for the transition from vegetative to reproductive branching in rice. *Mol. Plant* **2021**, *14*, 1168–1184. [CrossRef]

39. Xu, R.; Song, F.; Zheng, Z. *OsBISAMT1*, a gene encoding S-adenosyl-L-methionine: Salicylic acid carboxyl methyltransferase, is differentially expressed in rice defense responses. *Mol. Biol. Rep.* **2006**, *33*, 223–231. [CrossRef]

40. Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thastrom, A.; Field, Y.; Moore, I.K.; Wang, J.P.; Widom, J. A genomic code for nucleosome positioning. *Nature* **2006**, *442*, 772–778. [CrossRef]

41. Zhao, D.; Ferguson, A.A.; Jiang, N. What makes up plant genomes: The vanishing line between transposable elements and genes. *Biochim. Biophys. Acta* **2016**, *1859*, 366–380. [CrossRef] [PubMed]

42. Tian, F.; Yang, D.C.; Meng, Y.Q.; Jin, J.; Gao, G. PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Res.* **2020**, *48*, D1104–D1113. [CrossRef] [PubMed]

43. Sun, X.; Zhang, Z.; Wu, J.; Cui, X.; Feng, D.; Wang, K.; Xu, M.; Zhou, L.; Han, X.; Gu, X.; et al. The *Oryza sativa* regulator *HDR1* associates with the kinase *OsK4* to control photoperiodic flowering. *PLoS Genet.* **2016**, *12*, e1005927. [CrossRef] [PubMed]

44. Qin, Y.; Shen, X.; Wang, N.; Ding, X. Characterization of a novel cyclase-like gene family involved in controlling stress tolerance in rice. *J. Plant Physiol.* **2015**, *181*, 30–41. [CrossRef]

45. Campo, S.; Baldrich, P.; Messeguer, J.; Lalanne, E.; Coca, M.; San Segundo, B. Overexpression of a calcium-dependent protein kinase confers salt and drought tolerance in rice by preventing membrane lipid peroxidation. *Plant Physiol.* **2014**, *165*, 688–704. [CrossRef]

46. Peng, H.; Zhang, Q.; Li, Y.; Lei, C.; Zhai, Y.; Sun, X.; Sun, D.; Sun, Y.; Lu, T. A putative leucine-rich repeat receptor kinase, *OsBRR1*, is involved in rice blast resistance. *Planta* **2009**, *230*, 377–385. [CrossRef]

47. Matsuda, S.; Takano, S.; Sato, M.; Furukawa, K.; Nagasawa, H.; Yoshikawa, S.; Kasuga, J.; Tokuji, Y.; Yazaki, K.; Nakazono, M.; et al. Rice stomatal closure requires guard cell plasma membrane ATP-binding cassette transporter *RCN1/OsABCG5*. *Mol. Plant* **2016**, *9*, 417–427. [CrossRef]

48. Streubel, J.; Pesce, C.; Hutin, M.; Koebnik, R.; Boch, J.; Szurek, B. Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* **2013**, *200*, 808–819. [CrossRef]

49. Yadav, D.K.; Tuteja, N. Rice G-protein coupled receptor (GPCR): In silico analysis and transcription regulation under abiotic stress. *Plant Signal Behav.* **2011**, *6*, 1079–1086. [CrossRef]

50. Zhang, W.; Tan, L.; Sun, H.; Zhao, X.; Liu, F.; Cai, H.; Fu, Y.; Sun, X.; Gu, P.; Zhu, Z.; et al. Natural variations at *TIG1* encoding a TCP transcription factor contribute to plant architecture domestication in rice. *Mol. Plant* **2019**, *12*, 1075–1089. [CrossRef]

51. Chen, F.; Li, Q.; Sun, L.; He, Z. The rice 14-3-3 gene family and its involvement in responses to biotic and abiotic stress. *DNA Res.* **2006**, *13*, 53–63. [CrossRef] [PubMed]

52. Akamatsu, A.; Wong, H.L.; Fujiwara, M.; Okuda, J.; Nishide, K.; Uno, K.; Imai, K.; Umemura, K.; Kawasaki, T.; Kawano, Y.; et al. An OsCEBiP/OsCERK1-OsRacGEF1-OsRac1 module is an essential early component of chitin-induced rice immunity. *Cell Host Microbe* **2013**, *13*, 465–476. [CrossRef] [PubMed]

53. Mittal, D.; Chakrabarti, S.; Sarkar, A.; Singh, A.; Grover, A. Heat shock factor gene family in rice: Genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. *Plant Physiol. Biochem.* **2009**, *47*, 785–795. [CrossRef] [PubMed]

54. Freeling, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **2009**, *60*, 433–453. [CrossRef] [PubMed]

55. Innan, H.; Kondrashov, F. The evolution of gene duplications: Classifying and distinguishing between models. *Nat. Rev. Genet.* **2010**, *11*, 97–108. [CrossRef]

56. Panchy, N.; Lehti-Shiu, M.; Shiu, S.H. Evolution of gene duplication in plants. *Plant Physiol.* **2016**, *171*, 2294–2316. [CrossRef]

57. Wang, Y.; Wang, X.; Lee, T.H.; Mansoor, S.; Paterson, A.H. Gene body methylation shows distinct patterns associated with different gene origins and duplication modes and has a heterogeneous relationship with gene expression in *Oryza sativa* (rice). *New Phytol.* **2013**, *198*, 274–283. [CrossRef]

58. Xu, C.; Nadon, B.D.; Kim, K.D.; Jackson, S.A. Genetic and epigenetic divergence of duplicate genes in two legume species. *Plant Cell Environ.* **2018**, *41*, 2033–2044. [CrossRef]

59. Miao, Z.; Zhang, T.; Qi, Y.; Song, J.; Han, Z.; Ma, C. Evolution of the RNA N (6)-methyladenosine methylome mediated by genomic duplication. *Plant Physiol.* **2020**, *182*, 345–360. [CrossRef]

60. Wang, J.; Orlov, Y.L.; Li, X.; Zhou, Y.; Liu, Y.; Yuan, C.; Chen, M. In situ dissecting the evolution of gene duplication with different histone modification patterns based on high-throughput data analysis in *Arabidopsis thaliana*. *PeerJ* **2021**, *9*, e10426. [CrossRef]

61. Shi, T.; Rahmani, R.S.; Gugger, P.F.; Wang, M.; Li, H.; Zhang, Y.; Li, Z.; Wang, Q.; van de Peer, Y.; Marchal, K.; et al. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. *Mol. Biol. Evol.* **2020**, *37*, 2394–2413. [CrossRef] [PubMed]

62. Lannes, R.; Rizzon, C.; Lerat, E. Does the presence of transposable elements impact the epigenetic environment of human duplicated genes? *Genes* **2019**, *10*, 249. [CrossRef] [PubMed]

63. Liu, Z.; Zhao, H.; Yan, Y.; Wei, M.X.; Zheng, Y.C.; Yue, E.K.; Alam, M.S.; Smartt, K.O.; Duan, M.H.; Xu, J.H. Extensively current activity of transposable elements in natural rice accessions revealed by singleton insertions. *Front. Plant Sci.* **2021**, *12*, 745526. [CrossRef] [PubMed]

64. Lisch, D.; Bennetzen, J.L. Transposable element origins of epigenetic gene regulation. *Curr. Opin. Plant Biol.* **2011**, *14*, 156–161. [CrossRef] [PubMed]

65. Kaasik, K.; Lee, C.C. Reciprocal regulation of haem biosynthesis and the circadian clock in mammals. *Nature* **2004**, *430*, 467–471. [CrossRef]

66. Noshay, J.M.; Marand, A.P.; Anderson, S.N.; Zhou, P.; Mejia Guerra, M.K.; Lu, Z.; O'Connor, C.H.; Crisp, P.A.; Hirsch, C.N.; Schmitz, R.J.; et al. Assessing the regulatory potential of transposable elements using chromatin accessibility profiles of maize transposons. *Genetics* **2021**, *217*, 1–13. [CrossRef]

67. Moschetti, R.; Palazzo, A.; Lorusso, P.; Viggiano, L.; Marsano, R.M. "What You Need, Baby, I Got It": Transposable elements as suppliers of *cis*-operating sequences in *Drosophila*. *Biology* **2020**, *9*, 25. [CrossRef]

68. Kunarso, G.; Chia, N.Y.; Jeyakani, J.; Hwang, C.; Lu, X.; Chan, Y.S.; Ng, H.H.; Bourque, G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **2010**, *42*, 631–634. [CrossRef]

69. Thornburg, B.G.; Gotea, V.; Makalowski, W. Transposable elements as a significant source of transcription regulating signals. *Gene* **2006**, *365*, 104–110. [CrossRef]

70. Nicolau, M.; Picault, N.; Moissiard, G. The evolutionary volte-face of transposable elements: From harmful jumping genes to major drivers of genetic innovation. *Cells* **2021**, *10*, 2952. [CrossRef]

71. Zhang, J.; Liu, Y.; Xia, E.H.; Yao, Q.Y.; Liu, X.D.; Gao, L.Z. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E7022–E7029. [CrossRef] [PubMed]

72. Choi, J.Y.; Purugganan, M.D. Evolutionary epigenomics of retrotransposon-mediated methylation spreading in rice. *Mol. Biol. Evol.* **2018**, *35*, 365–382. [CrossRef] [PubMed]

73. Deniz, O.; Frost, J.M.; Branco, M.R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **2019**, *20*, 417–431. [CrossRef] [PubMed]

74. Hollister, J.D.; Gaut, B.S. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **2009**, *19*, 1419–1428. [CrossRef]

75. Zhang, B.; Zhou, Y.; Lin, N.; Lowdon, R.F.; Hong, C.; Nagarajan, R.P.; Cheng, J.B.; Li, D.; Stevens, M.; Lee, H.J.; et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res.* **2013**, *23*, 1522–1540.

76. Todd, C.D.; Deniz, O.; Taylor, D.; Branco, M.R. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *elife* **2019**, *8*, e44344. [CrossRef]

77. Jin, P.; Qin, S.; Chen, X.; Song, Y.; Li-Ling, J.; Xu, X.; Ma, F. Evolutionary rate of human tissue-specific genes are related with transposable element insertions. *Genetica* **2012**, *140*, 513–523. [CrossRef]

78. Chishima, T.; Iwakiri, J.; Hamada, M. Identification of transposable elements contributing to tissue-specific expression of long non-coding RNAs. *Genes (Basel)* **2018**, *9*, 23. [CrossRef]

79. Trizzino, M.; Kapusta, A.; Brown, C.D. Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics.* **2018**, *19*, 468. [CrossRef]

80. Zhou, X.; Zhu, T.; Fang, W.; Yu, R.; He, Z.; Chen, D. Systematic annotation of conservation states provides insights into regulatory regions in rice. *J. Genet. Genomics* 2022, *in press*. [CrossRef]

81. Qian, W.; Liao, B.Y.; Chang, A.Y.; Zhang, J. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet.* **2010**, *26*, 425–430. [CrossRef]

82. Wang, X.; Zhang, Z.; Fu, T.; Hu, L.; Xu, C.; Gong, L.; Wendel, J.F.; Liu, B. Gene-body CG methylation and divergent expression of duplicate genes in rice. *Sci. Rep.* **2017**, *7*, 2675. [CrossRef] [PubMed]

83. Mortada, H.; Vieira, C.; Lerat, E. Genes devoid of full-length transposable element insertions are involved in development and in the regulation of transcription in human and closely related species. *J. Mol. Evol.* **2010**, *71*, 180–191. [CrossRef]

84. Nellaker, C.; Keane, T.M.; Yalcin, B.; Wong, K.; Agam, A.; Belgard, T.G.; Flint, J.; Adams, D.J.; Frankel, W.N.; Ponting, C.P. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* **2012**, *13*, R45. [CrossRef]

85. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]

86. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef]

87. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]

88. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]

89. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]

90. Ramirez, F.; Ryan, D.P.; Gruning, B.; Bhardwaj, V.; Kilpert, F.; Richter, A.S.; Heyne, S.; Dundar, F.; Manke, T. deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **2016**, *44*, W160–W165. [CrossRef] [PubMed]

91. Thorvaldsdottir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform.* **2013**, *14*, 178–192. [CrossRef]

92. Krueger, F.; Andrews, S.R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **2011**, *27*, 1571–1572. [CrossRef] [PubMed]

93. Guo, W.; Zhu, P.; Pellegrini, M.; Zhang, M.Q.; Wang, X.; Ni, Z. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **2018**, *34*, 381–387. [CrossRef] [PubMed]

94. Zhang, Y.; Liu, T.; Meyer, C.A.; Eeckhoute, J.; Johnson, D.S.; Bernstein, B.E.; Nusbaum, C.; Myers, R.M.; Brown, M.; Li, W.; et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **2008**, *9*, R137. [CrossRef] [PubMed]

95. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef]

96. Ou, S.; Su, W.; Liao, Y.; Chougule, K.; Agda, J.R.A.; Hellinga, A.J.; Lugo, C.S.B.; Elliott, T.A.; Ware, D.; Peterson, T.; et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **2019**, *20*, 275. [CrossRef]

97. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]

98. Schug, J.; Schuller, W.P.; Kappen, C.; Salbaum, J.M.; Bucan, M.; Stoeckert, C.J., Jr. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **2005**, *6*, R33. [CrossRef]

99. Pang, H.; Chen, Q.; Li, Y.; Wang, Z.; Wu, L.; Yang, Q.; Zheng, X. Comparative analysis of the transcriptomes of two rice subspecies during domestication. *Sci. Rep.* **2021**, *11*, 3660. [CrossRef] [PubMed]

100. Wang, M.; Tu, L.; Lin, M.; Lin, Z.; Wang, P.; Yang, Q.; Ye, Z.; Shen, C.; Li, J.; Zhang, L.; et al. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. *Nat. Genet.* **2017**, *49*, 579–587. [CrossRef]

101. Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* **2019**, *20*, 38. [CrossRef] [PubMed]

102. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef]

103. Li, Z.; Defoort, J.; Tasdighian, S.; Maere, S.; van de Peer, Y.; de Smet, R. Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* **2016**, *28*, 326–344. [CrossRef] [PubMed]

104. Machanick, P.; Bailey, T.L. MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **2011**, *27*, 1696–1697. [CrossRef] [PubMed]

105. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [CrossRef]

106. O'Malley, R.C.; Huang, S.C.; Song, L.; Lewsey, M.G.; Bartlett, A.; Nery, J.R.; Galli, M.; Gallavotti, A.; Ecker, J.R. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **2016**, *165*, 1280–1292. [CrossRef]