

RESEARCH

Open Access



# Genome-wide analysis and functional annotation of chromatin-enriched noncoding RNAs in rice during somatic cell regeneration

Yu-Chan Zhang<sup>1,2\*</sup>, Yan-Fei Zhou<sup>1†</sup>, Yu Cheng<sup>1†</sup>, Jia-Hui Huang<sup>1</sup>, Jian-Ping Lian<sup>1</sup>, Lu Yang<sup>1</sup>, Rui-Rui He<sup>1</sup>, Meng-Qi Lei<sup>1</sup>, Yu-Wei Liu<sup>1</sup>, Chao Yuan<sup>1</sup>, Wen-Long Zhao<sup>1</sup>, Shi Xiao<sup>1</sup> and Yue-Qin Chen<sup>1,2\*</sup>

\* Correspondence: [zhyuchan@mail.sysu.edu.cn](mailto:zhyuchan@mail.sysu.edu.cn); [lsscyq@mail.sysu.edu.cn](mailto:lsscyq@mail.sysu.edu.cn)

<sup>†</sup>Yu-Chan Zhang, Yan-Fei Zhou and Yu Cheng contributed equally to this work.

<sup>1</sup>Guangdong Provincial Key Laboratory of Plant Resources, State Key Laboratory for Biocontrol, School of Life Science, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China  
Full list of author information is available at the end of the article

## Abstract

**Background:** Plants have the remarkable ability to generate callus, a pluripotent cell mass that acquires competence for subsequent tissue regeneration. Global chromatin remodeling is required for this cell fate transition, but how the process is regulated is not fully understood. Chromatin-enriched noncoding RNAs (cheRNAs) are thought to play important roles in maintaining chromatin state. However, whether cheRNAs participate in somatic cell regeneration in plants has not yet been clarified.

**Results:** To uncover the characteristics and functions of cheRNAs during somatic cell reprogramming in plants, we systematically investigate cheRNAs during callus induction, proliferation and regeneration in rice. We identify 2284 cheRNAs, most of which are novel long non-coding RNAs or small nucleolar RNAs. These cheRNAs, which are highly conserved across plant species, shuttle between chromatin and the nucleoplasm during somatic cell regeneration. They positively regulate the expression of neighboring genes via specific RNA motifs, which may interact with DNA motifs around cheRNA loci. Large-scale mutant analysis shows that cheRNAs are associated with plant size and seed morphology. Further detailed functional investigation of two che-lncRNAs demonstrates that their loss of function impairs cell dedifferentiation and plant regeneration, highlighting the functions of cheRNAs in regulating the expression of neighboring genes via specific motifs. These findings support *cis*- regulatory roles of cheRNAs in influencing a variety of rice traits.

**Conclusions:** cheRNAs are a distinct subclass of regulatory non-coding RNAs that are required for somatic cell regeneration and regulate rice traits. Targeting cheRNAs has great potential for crop trait improvement and breeding in future.

**Keywords:** Chromatin-enriched noncoding RNAs, Somatic cell regeneration, Rice traits



## Background

Plant development is driven by specific patterns of gene expression that are tightly regulated in a spatio-temporal manner. Chromatin remodeling plays a central role in establishing transcriptional programs required for organ initiation and differentiation [1–3]. Whereas epigenetic states in animals are established early during embryonic development, epigenetic mechanisms in plants also operate during post-embryonic developmental transitions, such as organogenesis and flowering [4, 5]. The chromatin remodeling activities in plants provide a higher degree of flexibility that likely underlies their developmental plasticity. Specifically, multiple detached plant tissues are capable of forming a pluripotent cell mass called callus, which in turn can regenerate into different organs and form a new plant, a process known as somatic embryogenesis [6, 7]. Various genetic and physiological factors trigger somatic embryogenesis in different types of somatic cells. Chromatin remodeling is believed to play a central role during somatic cell reprogramming and pluripotent cell differentiation [6–8].

Recent genomic research has revealed that the genomes of different organisms, including plants, are more prevalently transcribed than previously thought [9, 10]. Mammalian and plant genomes express not only protein-coding mRNAs but also a large repertoire of non-coding RNAs (ncRNAs) with regulatory roles in different layers of gene expression [9, 10]. In mammals, many ncRNAs appear to act directly on chromatin, as exemplified by various long non-coding RNAs (lncRNAs). Some lncRNAs mediate genomic interactions predominantly *in cis*, whereas others are capable of acting extensively *in trans* [11–14]. These findings point to a role for specific RNA–chromatin interactions in regulating gene expression. lncRNA-directed processes also function in dosage compensation in *Drosophila*, where the localization of the histone acetyltransferase MOF to the male X chromosome is dependent on roX ncRNAs [15]. Moreover, recent findings suggest that ncRNAs are integral components of chromatin [12, 13] that play an important role in the higher-order chromatin structure of pericentric heterochromatin by organizing heterochromatic components [16, 17]. In plants, lncRNAs have been reported to interact with chromatin remodelers [18]. These observations underline the importance of ncRNAs as cofactors in modifying chromatin via the recruitment of chromatin-remodeling complexes. However, the identities of chromatin-enriched ncRNAs (cheRNAs) in plants have not yet been addressed on a global scale.

In this study, we asked whether specific ncRNA–chromatin interactions participate in regulating gene expression during somatic cell regeneration in plants and, if so, what the identities of these chromatin-interacting ncRNAs are. Specifically, during somatic cell reprogramming and pluripotent cell differentiation, what is the landscape of chromatin-associated ncRNAs? Understanding the composition, characteristics, and functions of cheRNAs during cellular reprogramming would provide insight into the molecular network regulating cell pluripotency, thereby facilitating crop breeding.

To address these issues, we used *in vitro*–cultured embryogenic rice callus as a model to investigate chromatin-interacting ncRNAs associated with embryogenesis and post-embryonic development. Embryogenic calli derived from mature embryos contain a set of homogeneous pluripotent cells that are thought to represent proliferating meristematic tissues. When cultured in the appropriate medium, these embryogenic calli undergo somatic embryogenesis. The cell division, cytodifferentiation, and embryogenesis of embryogenic calli are consistent with these biological processes *in vivo*. We

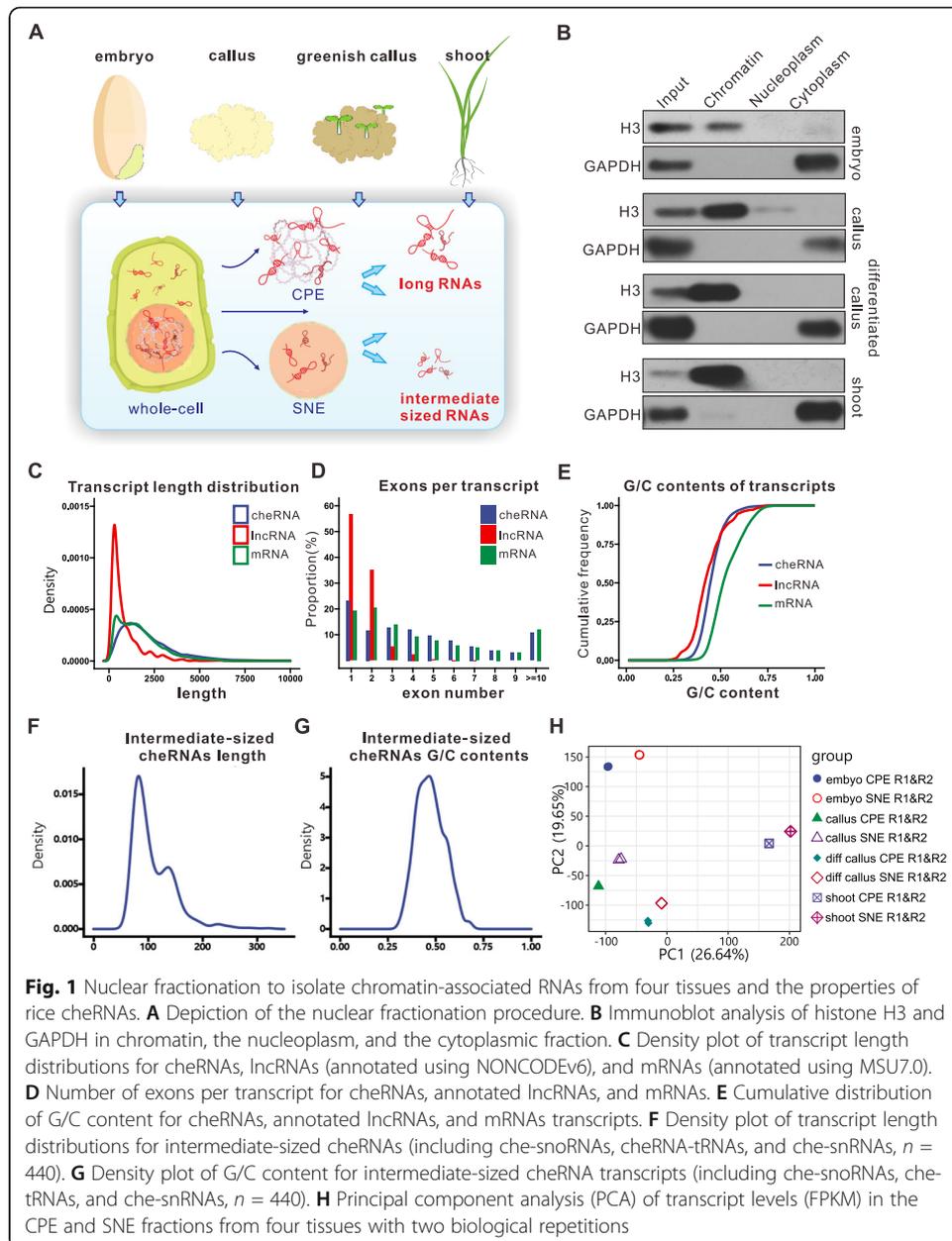
identified 2284 cheRNAs, including lncRNAs, small nucleolar RNA (snoRNAs), and tRNAs. These cheRNAs are highly conserved and represent subclasses of rice ncRNAs with developmental-stage-specific enrichment patterns. During somatic cell regeneration, cheRNAs shuttle between chromatin and the nucleoplasm, where they regulate the expression of specific protein-coding genes via specific RNA motifs, which might interact with DNA motifs around cheRNA loci. Large-scale analysis of mutant and transgenic rice plants indicated that cheRNAs regulate yield-related traits in rice. Thus, cheRNAs have great potential as targets for trait improvement and crop breeding in the future.

## Results

### Global view of RNA-chromatin interactions during somatic cell regeneration and differentiation in rice

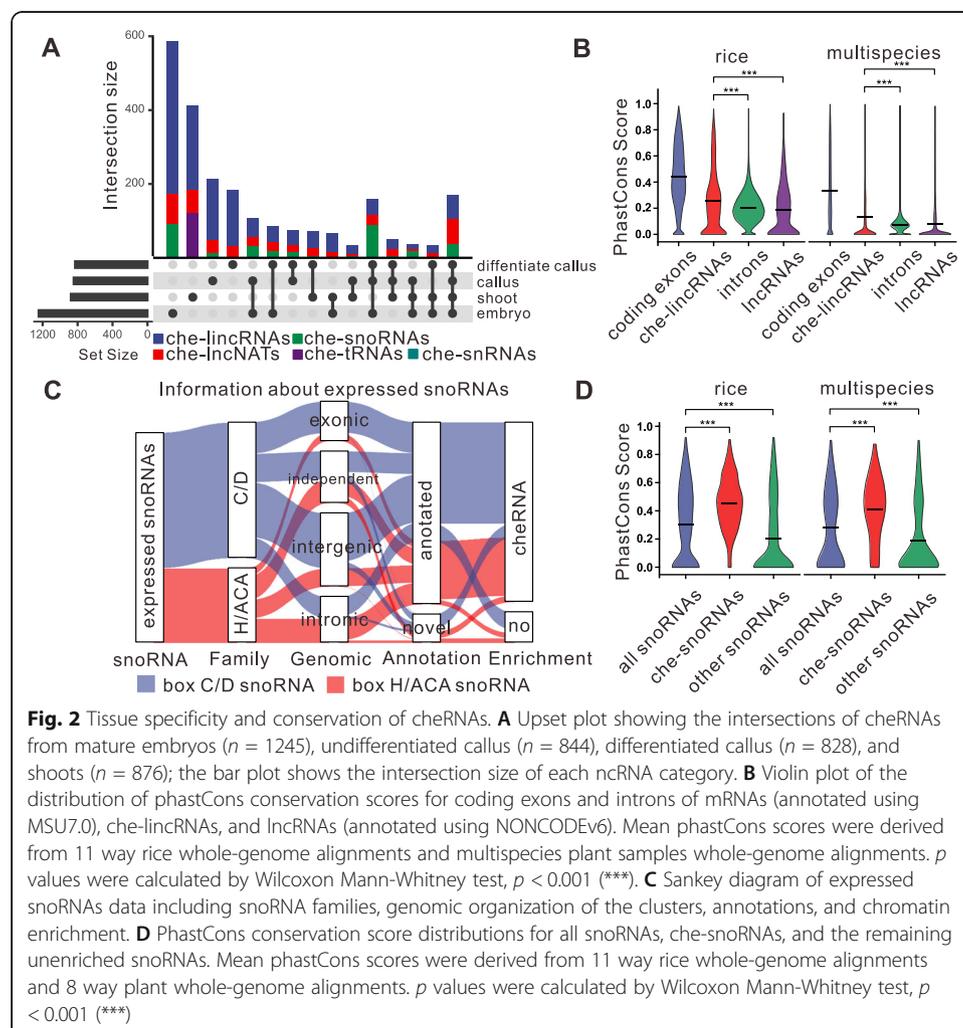
To identify regulators of chromatin reprogramming that underlie cell fate changes, we characterized the landscape of chromatin-associated RNAs during callus induction, proliferation, and regeneration. Four different rice tissues were collected for the fractionation of nuclei (Fig. 1A), including (1) mature embryos, (2) undifferentiated embryogenic callus, (3) greenish, partially regenerated (differentiated) callus after over 8 weeks of subculture (every 2 weeks on the same medium), and (4) shoots. We developed a method to separate both nucleoplasmic RNAs and chromatin-associated RNAs. In brief, different rice samples were ground and subjected to cell lysis, and the nucleus fraction was collected. This fraction was further divided into the nucleoplasmic and chromatin fractions, and RNAs were isolated individually from these fractions for sequencing (Additional file 1: Fig. S1A). Stripping highly abundant mRNA from the chromatin pellet with urea was critical for identifying chromatin pellet extract (CPE) transcripts because it effectively magnified the coverage depth of low-abundance RNA species. The fractionation was validated by confirming robust chromatin enrichment of histone H3 and cytoplasm enrichment of GAPDH (Fig. 1B).

We then performed transcriptome sequencing (RNA-seq) and deep sequencing of intermediate-sized RNAs (50 to 300 nt, respectively) of the resulting chromatin pellet extract (CPE), soluble-nuclear extract (SNE), and the input sample, yielding > 2750 million mapped reads (Additional file 1: Fig. S1B; Additional file 2: Table S1). De novo-assembled transcripts from both the CPE and SNE fractions were mapped to 65,703 distinct loci in the annotated rice genome from the Rice Genome Annotation Project (MSU7.0). Each transcript was scored for its abundance in the CPE vs. SNE fraction; transcripts with a relative abundance in CPE versus SNE > 1.2 (adjusted  $p$  value < 0.05) were defined as chromatin-enriched transcripts. The detected chromatin enriched RNAs include 81.5% mRNAs and 18.5% ncRNAs. The chromatin-enriched ncRNAs have a significantly higher chromatin enrichment ratio than those of total mRNAs and the annotated lncRNAs, and these chromatin-enriched ncRNAs were named cheRNAs and used for the following studies. For better recognizing each XLOC transcript in the genome, these cheRNAs are named according to the type of the cheRNAs and their number coordinating in chromosome (Additional file 1: Fig. S2A; Additional file 3: Table S2).



The coding potential of cheRNAs was much lower than that of protein-coding genes but slightly higher than that of lncRNAs (Additional file 1: Fig. S2B). We statistically characterized the cheRNAs based on their lengths and GC contents. The average exon size and transcript length of long non-coding cheRNAs (che-lncRNAs;  $\geq 200$  nt) were higher than those of annotated lncRNAs. The number of exons in these cheRNAs was greater than for the annotated lncRNAs and similar to that of mRNAs (Fig. 1C, D; Additional file 1: Fig. S2C). The average GC content of these long non-coding cheRNAs was similar to that of annotated lncRNAs and lower than that of known protein-coding genes (Fig. 1E). For the intermediate-sized non-coding transcripts ( $\geq 50$  nt,  $\leq 300$  nt), the length distribution map showed two peaks at  $\sim 85$  nt and  $\sim 140$  nt, respectively (Fig. 1F). The GC content ranged from 25 to 75%, with a peak at 50% (Fig. 1G).

We then compared the chromatin-enriched patterns of RNAs during callus induction from embryos, callus differentiation, and plant regeneration. The biological replicates exhibited high reproducibility, and the samples were well separated from each other, suggesting developmental specificity in a substantial fraction of the captured interactions (Fig. 1H). We have also performed H3 ChRIP (Chromatin RNA Immunoprecipitation) to validate the chromatin enrichment of 8 cheRNAs. The results showed that the examined cheRNAs which are chromatin enrichment in callus (*OsCHELIN1575*, *OsCHELIN2168*, *OsCHENAT0124*, and *OsCHENAT2171*) or in shoot (*OsCHENAT0592*, *OsCHELIN0038*, *OsCHELIN0123*, and *OsCHELIN0456*) were associated to H3, whereas the lncRNAs identified in nucleoplasm were not associated to H3, which further confirmed their chromatin-bound (Additional file 1: Fig. S2D, E). More cheRNAs were detected in embryos than in the other samples (Fig. 2A). Of the 2284 cheRNAs detected in our experiment, 170 (7.4%) were enriched on chromatin at all developmental stages, while 61.1% of the cheRNAs showed stage-specific enrichment patterns (Fig. 2A). Moreover, 31.5% of these cheRNAs shuttled from the CPE to SNE fragments during somatic cell regeneration and differentiation. Of these shuttled cheRNAs, 773 shuttled from CPE to SNE during callus induction, and 391 shuttled from



CPE to SNE during callus differentiation. These results suggest that cheRNAs are under controlled during somatic cell reprogramming.

### Chromatin-interacting ncRNAs are a distinct subclass of ncRNAs

We classified the cheRNAs based on rRNA and tRNA gene annotations from RAP-DB (<https://rapdb.dna.affrc.go.jp/>), and snRNA and snoRNA gene annotations from the Rfamv14.5 database (<http://rfam.xfam.org/>). Novel snoRNAs were predicted by snoSeekerNGS and filtered based on the presence of conserved box motifs and expression levels. We identified the che-lncRNAs by comparing each genomic coordinate and strand with MSU7.0 (<http://rice.uga.edu/>)-annotated mRNA transcripts and estimating the coding potential using the CPC2 program [19]. The chromatin-enriched ncRNAs included lncRNAs (80.74%), snoRNAs (13.75%), tRNAs (5.3%), and snRNAs (0.22%) (Fig. 2A; Additional file 3: Table S2).

Of these cheRNAs, 1527 lacked ncRNA annotations, which was greater than the number of unannotated transcripts (901) enriched in the SNE fraction. Of the lncRNAs, which include both long intergenic noncoding RNAs (lincRNAs) and long noncoding natural antisense transcripts (lncNATs), 77.49% (1429) were not annotated in the PLncDB [20], RNACentral [21], EVLncRNAs [22], and NONCODE (<http://www.noncode.org/index.php>) databases and were therefore considered to be novel lncRNAs. These data suggest that, although many studies have identified lncRNAs from rice tissues, most cheRNAs escaped detection using conventional sequencing methods, possibly due to their low abundance and specific subcellular localization. Indeed, only 27.4% of che-lincRNAs are located within 500 bp away from coding genes. Che-lincRNAs exhibited specific strand bias from their putative transcription start sites (TSSs), and the H3K4me3, H3K27ac, and H4K12ac (Additional file 4: Table S3) which were reported to be enriched at gene TSSs showed enrichment at the TSSs of cheRNAs [12, 23] (Additional file 1: Fig. S2F, G). These results suggest that they might not be the byproducts of read-through transcription from upstream genes [12]. One thousand one hundred seventeen of the che-lincRNAs contain repeat elements and might be TE-derived che-lincRNAs. Higher suppressive histone marks (H3K27me3 and H3K9me2) and lower active histone marks (H3K27ac and H3K4me3) were observed at the TE-derived che-lincRNAs loci compared with other che-lincRNAs (Additional file 1: Fig. S2H; Additional file 3: Table S2).

As somatic cell regeneration might be controlled by conserved mechanisms, we investigated whether the che-lincRNAs were more conserved than the annotated lncRNAs. Our analysis showed that the level of conservation of the exons of che-lincRNAs in different plant species is similar to that of coding exons. The level of conservation is relatively modest compared to that of coding exons in different rice varieties, but greater than that of introns and rice lncRNAs (Fig. 2B), suggesting that cheRNAs might be subject to stronger evolutionary pressure than previously characterized lncRNAs.

We also observed many snoRNA-like transcripts on chromatin, including 201 C/D box and 113 H/ACA box snoRNAs. Similarly, chromatin-associated snoRNAs were also identified in mammals [24–26] (Fig. 2C). The 85-nt peak on the length distribution map of intermediate-sized cheRNAs represents box C/D snoRNAs, and the 145-nt peak

represents box H/ACA snoRNAs (Fig. 1F; Additional file 1: Fig. S2I). In total, we identified 367 expressed snoRNAs in our dataset. A comparison of these snoRNA-like transcripts with previously annotated snoRNAs in the Rfam database v14.5 and the literature [27] revealed that 49 are completely new snoRNA candidate genes, whereas 35 are chromatin-enriched snoRNAs (Fig. 2C). We analyzed the genomic organization of the novel snoRNAs and found that the majority (277 of 367, 75.5%) of the expressed snoRNAs detected are organized into 96 gene clusters, including 49 intergenic clusters, 35 intronic clusters, and 12 exonic clusters (Fig. 2C), which is in accordance with data previously obtained in rice [27, 28].

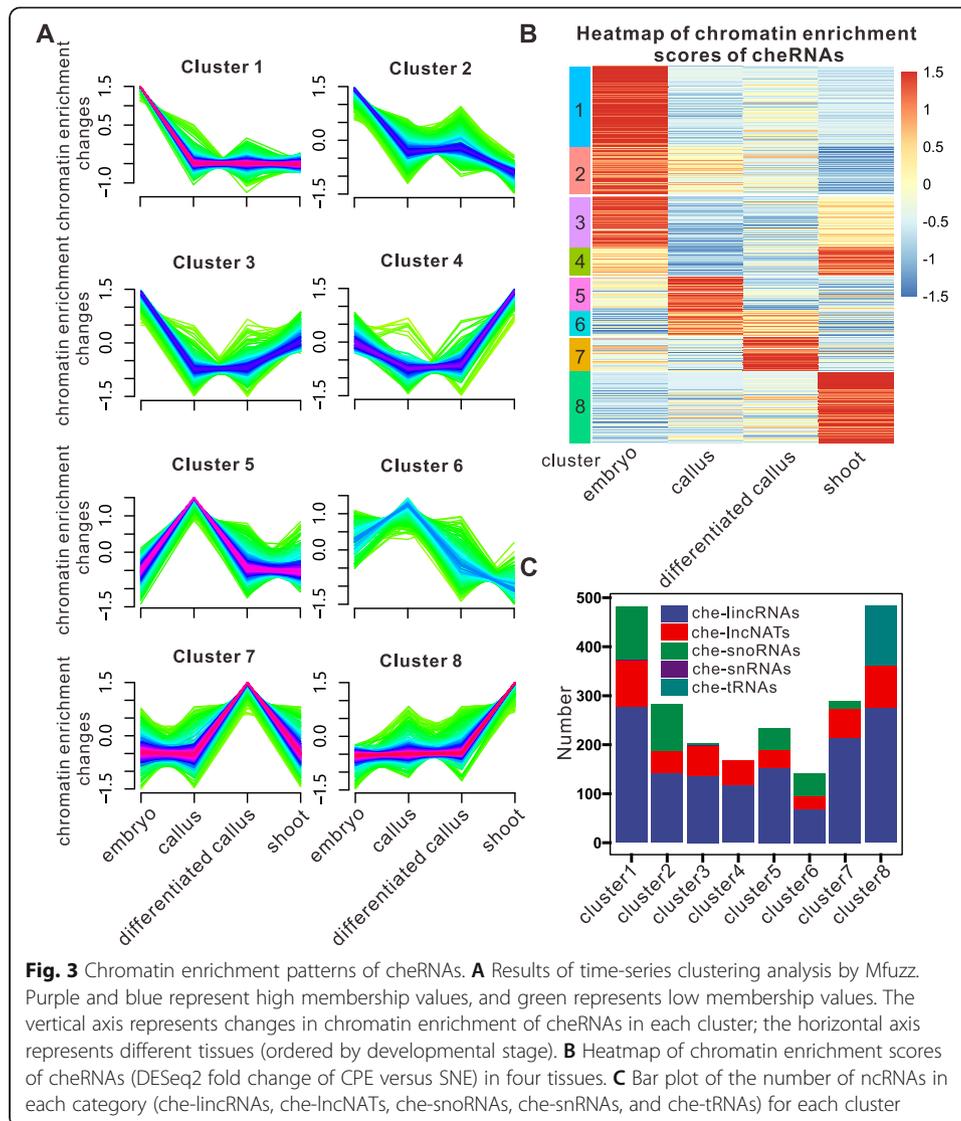
We then compared the evolutionary conservation between non-chromatin enriched snoRNAs and che-snoRNAs in different rice varieties, Arabidopsis, Sorghum, and Brachypodium. Similar to che-lincRNAs, che-snoRNAs were significantly more conserved than non-chromatin-enriched snoRNAs (Fig. 2D). When we predicted the potential strong interacted targets of the che-box C/D snoRNAs using PLEXY, 71 of the 237 identified boxC/D snoRNAs had no predicted complementary targets, suggesting that these che-snoRNAs might have different functions from traditional rice snoRNAs.

Collectively, these findings indicate that cheRNAs consist of ncRNAs that previously escaped detection using conventional sequencing methods and could not be annotated. The lengths and GC contents of the cheRNAs are different from those of annotated ncRNAs. In particular, cheRNAs appear to be subject to higher evolutionary pressure than previously characterized lincRNAs and snoRNAs. These characteristics suggest that cheRNAs represent a subclass of rice ncRNAs that might be functional during somatic cell regeneration, an inherent capacity of plants.

### **cheRNA dynamics during cellular reprogramming**

Having demonstrated that cheRNAs display developmental stage-specific enrichment patterns and may shuttle between chromatin and the nucleoplasm during somatic cell regeneration (Fig. 2A), we next asked how the changes in the chromatin enrichment patterns of cheRNAs are related to cellular reprogramming. To address this issue, we systematically identified the transcriptional shift and chromatin enrichment variation of each cheRNA. A fuzzy *c*-means soft clustering analysis of the cheRNAs grouped them into eight clusters (Fig. 3A, B; Additional file 3: Table S2). For this analysis, we used the scaled chromatin enrichment score (CPE versus SNE fold-changes resulting from differential expression analysis using DESeq2) for calculation with the R package Mfuzz. Only 7.4% of the cheRNAs were associated with chromatin in all four tissues (Fig. 2A, B). The majority shuttled from the CPE to the SNE fraction during somatic cell reprogramming, pointing to their regulatory roles throughout this process (Fig. 2A, B).

When a mature embryo dedifferentiates into pluripotent callus, the closed-chromatin state transforms into an open-chromatin state to allow massive gene reprogramming, conferring various possibilities for differentiation [7]. During this process, there were approximately three times more cheRNAs with declining chromatin enrichment than cheRNAs with increased chromatin enrichment (Fig. 3A, B), implying that more cheRNAs might function in maintaining the differentiated states of somatic cells. During callus differentiation and plant regeneration, only slightly more cheRNAs showed



increased chromatin enrichment than declining enrichment (Fig. 3A, B). Thus, cheRNAs consist of both positive and negative regulators of plant regeneration.

Among the eight clusters, the cheRNAs in clusters 1 and 2 were highly chromatin enriched on chromatin in embryos. The cheRNAs in clusters 3 and 4 showed reduced chromatin enrichment during callus induction from embryos and increased chromatin enrichment during plant regeneration, pointing to their roles in plant differentiation. The cheRNAs in clusters 5 and 6 exhibited enrichment patterns opposite to those of cluster 3 and 4 cheRNAs—they were highly enriched on chromatin in callus and showed declining enrichment during differentiation—implying that they might be required for pluripotent cell fate. Cluster 7 and 8 cheRNAs were highly enriched on chromatin in differentiated callus or shoot tissue (Fig. 3A, B). These data indicate that cheRNAs as a group have multiple functions during somatic cell regeneration.

We next analyzed the types of cheRNAs in each cluster. As shown in Fig. 3C, each cluster consists of different types of cheRNAs. che-lincRNAs were distributed in all eight clusters, with a modest preference for clusters 1 (20.6%) and 8 (19.9%). 63.5% of

the che-lncRNAs were enriched on chromatin in only a single tissue; whereas 16.6% were enriched on chromatin in three or more tissues, with only the degree of enrichment varying during somatic cell regeneration (Fig. 3C). The chromatin enrichment patterns of the che-snoRNAs were less tissue specific, as 31.2% of che-snoRNAs were enriched on chromatin in only one tissue. Most of these che-snoRNAs are found in clusters 1, 2, and 5 (Fig. 3C). che-tRNAs were only enriched on chromatin in shoots; thus, they are all in cluster 8 (Fig. 3A, C).

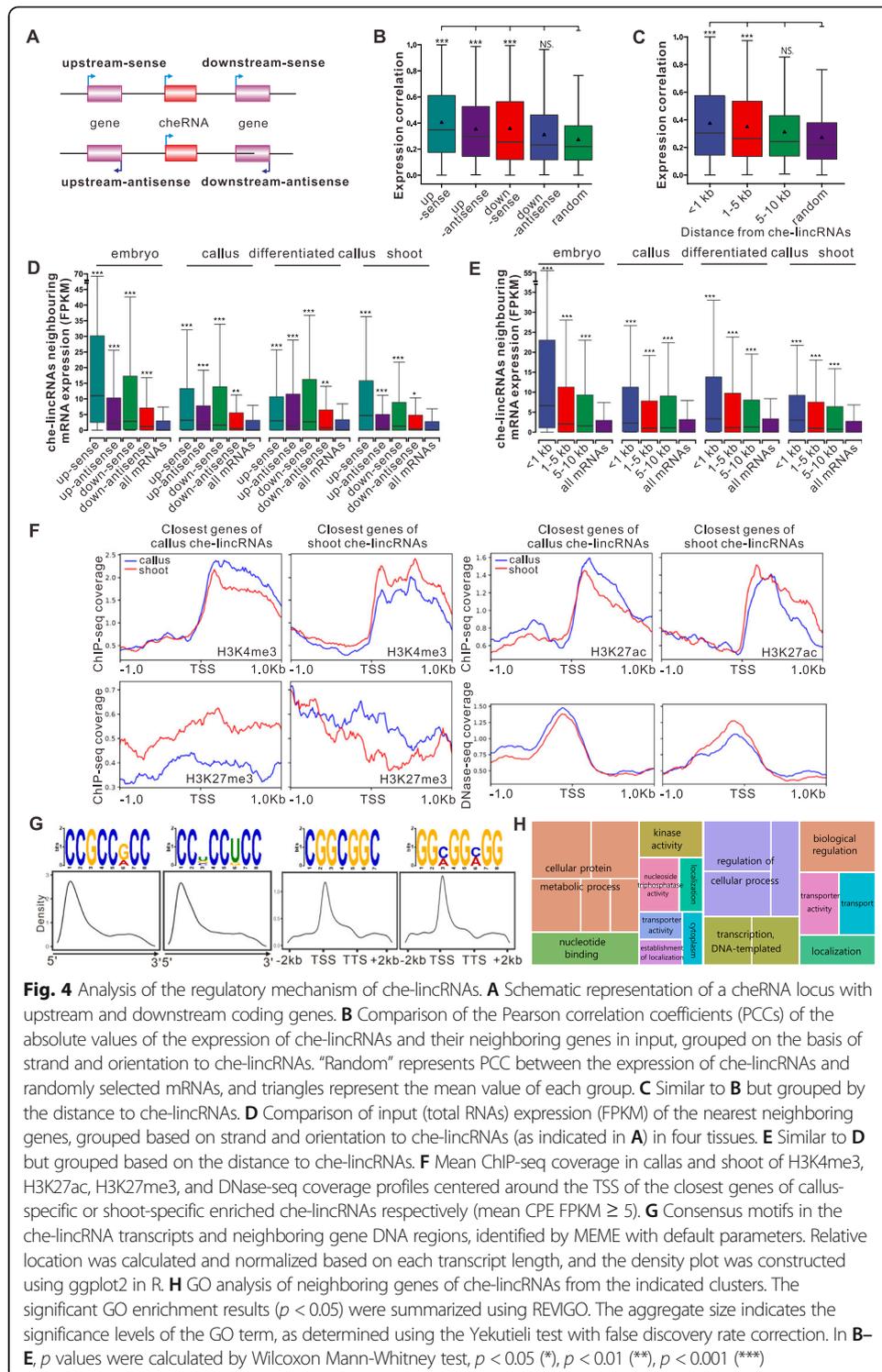
Taken together, these results suggest that the cheRNAs are tissue specific during somatic cell reprogramming, and they could shuttle from chromatin to the nucleoplasm. Thus, cheRNAs might function as regulators required for cell dedifferentiation or differentiation.

### **Mechanisms underlying the roles of cheRNAs in regulating cellular reprogramming and the expression of differentiation-related genes**

We analyzed the potential mechanisms of cheRNA function during cellular reprogramming and differentiation. Previous studies suggested that cheRNAs might act *in cis* or *in trans* to regulate gene expression [29]. Numerous che-lincRNAs might regulate the expression of their neighboring protein-coding genes [12]. To investigate the *cis*-regulatory activities of che-lincRNAs, we compared the expression patterns of che-lincRNAs and their neighboring genes based on total RNAs.

che-lincRNAs were divided by strand sense and orientation relative to their nearest coding genes (Fig. 4A), and their distribution showed no significant bias. However, che-lincRNA expression levels were more highly correlated with those of nearby protein-coding genes than with those of randomly chosen genes, and che-lincRNAs downstream of their neighbors displayed even stronger expression correlation than che-lincRNAs from other orientation (Fig. 4B; Additional file 5: Table S4). By contrast, the expression levels of che-lncNATs were more highly correlated with the expression levels of neighboring genes on the antisense strands (Additional file 1: Fig. S3). The correlation between the expression levels of che-lincRNAs and their neighboring genes gradually decreased with increasing distance from the che-lincRNAs (Fig. 4C), suggesting that che-lincRNAs might function as local enhancers that affect the expression of multiple genes. We also compared the che-lincRNAs with previously identified enhancers in rice. Fifty-two che-lincRNAs overlapped with enhancers identified by STARR-seq [30], and 274 overlapped with enhancers predicted by DHS [31] (Additional file 1: Fig. S4A), suggesting at least part of them might be *cis*-regulatory elements that function during cellular reprogramming.

Thus, we further analyzed the expression levels of the neighboring genes of che-lncRNAs to investigate the roles of che-lncRNAs in regulating gene expression. While the che-lncNATs did not significantly promote the expression of their neighboring genes (Additional file 1: Fig. S3), a higher correlation with che-lincRNAs tended to result in higher expression of neighboring protein-coding genes (Fig. 4D, E). The effect of che-lincRNAs in promoting neighboring gene expression was significantly higher than those of SNE lincRNAs, annotated lincRNAs, and lncNATs (Additional file 1: Fig. S4B). It has been reported that lncRNAs could regulate gene expression by mediating post-translational modification of histones [18]. Thus, we further analyzed the relationship



between che-lincRNAs and the epigenetic activities by using the published data on DNA methylation and histone modification sequencing. The results showed that che-lincRNAs are positively correlated with the H3K4me3 and H3K2ac modifications and chromatin accessibility and negatively correlated with the H3K27me3 modification of their neighboring genes, whereas other histone modifications and DNA methylations

are not affected by che-lincRNAs (Fig. 4F; Additional file 1: Fig. S4C, D). lncRNAs undergo sequence-specific interactions with DNA via triple helix (triplex) formation both *in cis* and *in trans*, which allows them to recruit protein complexes to specific genomic regions and regulate gene expression. To analyze whether che-lincRNAs could bind to the DNA regions around the neighboring genes or at *trans* genomic loci, we looked for potential triplexes between che-lincRNAs and the DNA regions of the gene bodies and the regions 1000 bp upstream or downstream of their neighboring genes or across the genome using Triplex Domain Finder (TDF) or Triplexator analysis. This indicated that 390 che-lincRNAs have predicted binding sites on the DNA regions around their neighboring genes (TDF *p* value < 0.05). In addition, che-lincRNAs also have predicted *trans* binding sites across the genome, which inclined to around the TSS of coding genes (Additional file 1: Fig. S4E; Additional file 6: Table S5), pointing to their *in trans* functions. Our data indicate that che-lincRNAs might promote gene expression *in cis*, hinting that che-lincRNAs may have a role in regulating the expression of reprogramming-specific genes rather than genes with basal functions.

We then looked for potential functional motifs in the che-lincRNAs, neighboring genes of the che-lincRNAs, and the genes predicted to form triplexes with che-lincRNAs. Two short motifs, “CCGCCWCC” (H = A or G) and “CCWCCMCC” (W = U or G or C, M = U or G), were identified in 294 and 461 che-lincRNAs, respectively (Fig. 4G). These motifs were also enriched in che-lincRNAs with predicted DNA-binding sites. The motifs were mainly present at the 5'-ends of che-lincRNAs (Fig. 4G). lncRNAs form DNA-RNA hybrids via complementary base pairing [32]. We also identified two short motifs, “CGGCGGC” and “GGNGGNGG” (N = C or A), that were mainly present around the transcription start sites of 619 and 661 neighboring genes of che-lincRNAs, respectively, and genes that were predicted to form triplexes with che-lincRNAs; these sequences are complementary to the motifs enriched in che-lincRNAs (Fig. 4G). In addition, these two motifs share sequence similarity with the reported DNA-binding motifs of lncRNAs in animals [33], pointing to a conserved regulatory mechanism in both plants and animals.

We then examined the coding genes proximal to che-lincRNAs that might be regulated *in cis* by che-lincRNAs in clusters 1 and 2, 3 and 4, 5 and 6, or 7 and 8. The proximal coding genes of che-lincRNAs from different clusters were enriched in different biological processes and functions (Fig. 4H; Additional file 1: Fig. S4F; Additional file 7: Table S6). Cell dedifferentiation and reprogramming lead to comprehensive transcriptional changes [7, 8]. che-lincRNAs in clusters 5 and 6 tended to be chromatin enriched during callus induction and distributed in the SNE fraction during plant regeneration. The proximal coding genes of che-lincRNAs in clusters 5 and 6 primarily included genes encoding proteins required for gene transcription (Fig. 4H; Additional file 7: Table S6). Besides transcriptional regulation, protein phosphorylation is another key factor in plant regeneration, which is essential for hormone signaling pathways. che-lincRNAs in clusters 3 and 4 showed opposite chromatin enrichment patterns from those in clusters 5 and 6. The proximal coding genes of che-lincRNAs in cluster 3 and 4 che-lincRNAs are mainly involved in protein phosphorylation; 15% encode kinases (Fig. 4H; Additional file 7: Table S6). Thus, our data suggest that che-lincRNAs are positive regulators of genes related to somatic cell reprogramming. Collectively, these results suggest that che-lincRNAs might function *in cis* or *in trans* to regulate the

expression of specific groups of genes via complementary sequences shared with their target genes.

### **cheRNAs are associated with crop traits**

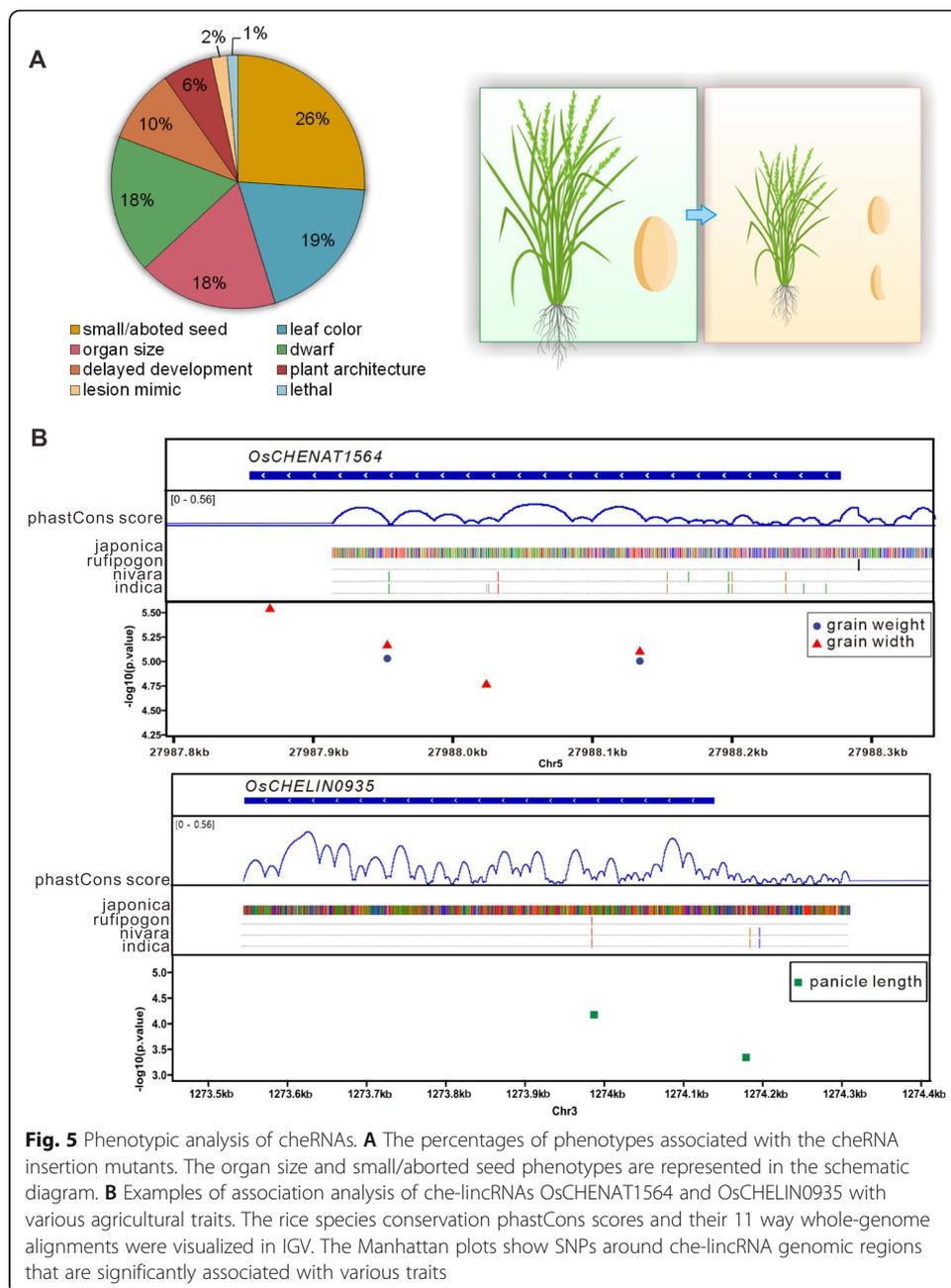
The results described above suggest that cheRNAs might regulate somatic cell reprogramming. As extensive regeneration ability is required by plants to ensure their post-embryonic development and survival, we investigated the possible roles of cheRNAs in crop development in vivo. Nine public rice mutant databases are currently available (affjp [34, 35], cirad [36, 37], gsnu, ostid [38], pfg [39], rmd [40], ship, trim, and ucd databases), three of which (affjp, cirad, and trim) describe the phenotypes of each mutant. The annotated phenotypes cover all stages of rice development. We performed a preliminary functional analysis of all the identified che-lincRNAs and che-snoRNAs using the nine rice mutant databases. che-lincRNAs were not selected for mutant analysis because they partially overlap with protein-coding genes, which might produce false positive results. Our strategy was to perform BLAST analysis of the flanking sequence tags (FSTs) included in each mutant database against the che-lincRNAs and che-snoRNAs and their 1-kb upstream regions (potential promoter regions) separately. A total of 531 cheRNAs were represented by insertional mutants in these databases; these mutants could contribute to the functional analysis of individual cheRNAs.

Annotated phenotypic data were available for the insertional mutants of 206 cheRNAs. We summarized the phenotypes of these mutants and performed statistical analysis. The most frequently occurring phenotypes were small/aborted seeds (26%), leaf color (19%), and altered organ size (18%) (Fig. 5A; Additional file 1: Fig. S4G; Additional file 8: Table S7). These results indicate that cheRNAs have important functions in determining organ size, especially seed size, as well as plant metabolism or plastid development. Notably, these traits directly affect crop yield.

Next, we investigated the single-nucleotide polymorphisms (SNPs) in the cheRNAs to identify trait-associated SNPs in a wide range of rice varieties from the 3K Rice Genomes Project [41, 42]. We identified 343 SNPs in 62 cheRNAs that are significantly associated with grain or panicle traits. For example, che-lincRNA *OsCHENAT1564* contains three SNPs which were significantly associated with grain and panicle size (Fig. 5B). One of these SNPs differentiated during rice domestication. In indica varieties and *Oryza nivara*, the SNP is a A allele, whereas in japonica varieties and *Oryza rufipogon*, it is a G allele (Fig. 5B). Importantly, this SNP was significantly associated with grain width and weight (Fig. 5B). Similarly, a SNP associated with panicle length was identified in the conserved region of che-lincRNA *OsCHELIN0935* (Fig. 5B). These data further emphasize the importance of cheRNAs for rice development and their potential practical value.

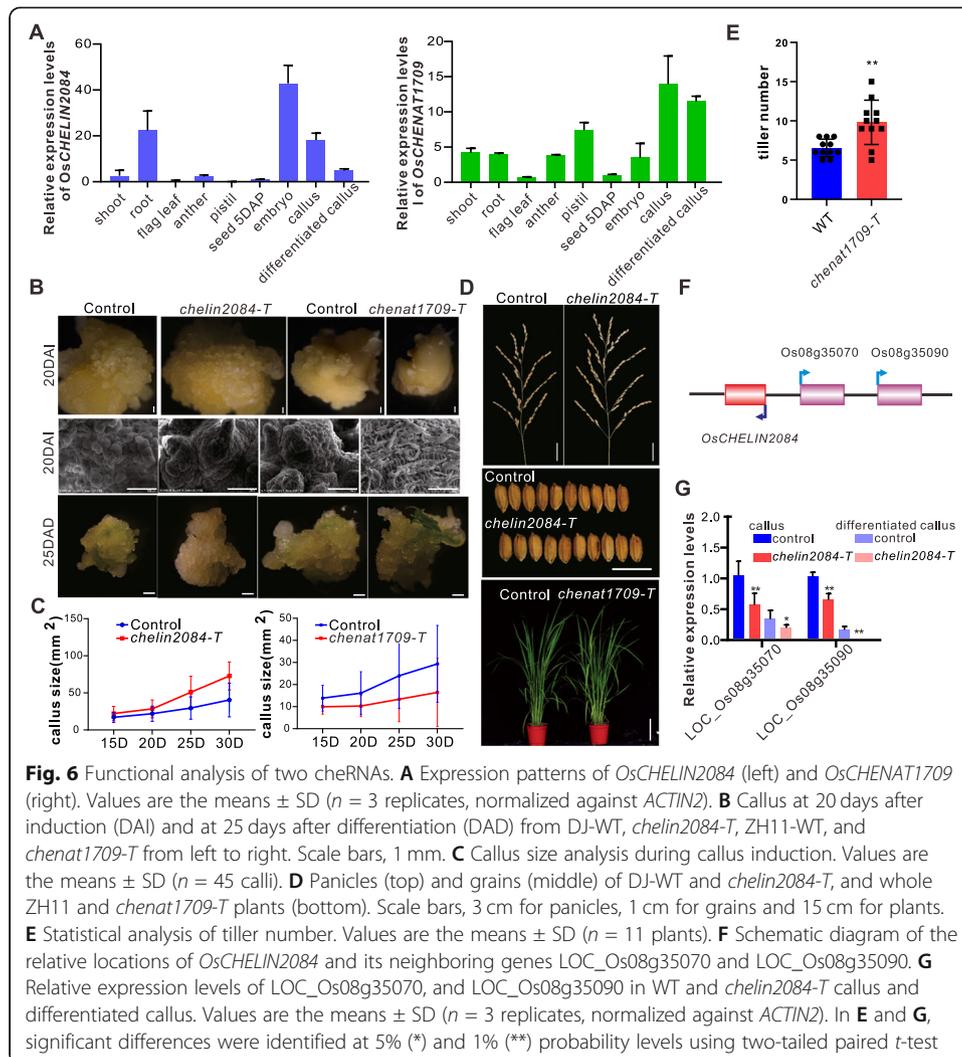
### **Loss of function of che-lincRNAs impairs cell dedifferentiation and plant regeneration ability**

To examine the functions of the cheRNAs, we further analyzed two che-lincRNAs (che-lincRNA *OsCHELIN2084* and che-lincRNA *OsCHENAT1709*) with no annotated phenotypes and compared the phenotypes of their T-DNA insertion mutants and RNAi transgenic plants (Additional file 1: Fig. S5A). *OsCHENAT1709* is highly expressed in



callus but expressed at lower levels in differentiated tissues, whereas *OsCHELIN2084* is highly expressed in embryos but more weakly expressed in callus (Fig. 6A). We used T<sub>2</sub> seeds of loss-of-function T-DNA insertion transgenic plants and RNAi transgenic plants of *OsCHENAT1709* and loss-of-function T-DNA insertion transgenic plants of *OsCHELIN2084* for phenotypic analysis (Additional file 1: Fig. S5A). The three stages that are typically observed during early callus differentiation are the formation of calli, green spots, and shoot primordia. We observed these three stages in the two insertional mutants.

Notably, the calli that were regenerated from the *OsCHELIN2084* loss-of-function mutant (*chelin2084-T*) and the *OsCHENAT1709* loss-of-function and RNAi mutant



(*chenat1709-T* and *chenat1709-RNAi*) had opposite phenotypes (Fig. 6B; Additional file 1: Fig. S5B, C, D). When callus formation was induced from embryos, callus formation from *chenat1709-T* and *chenat1709-RNAi* was restrained, whereas callus proliferation from *chelin2084-T* was more rapid than that of wild-type (WT) plants (Fig. 6B; Additional file 1: Fig. S5B - E). After 30 days of induction, the average callus size was 40.41 mm<sup>2</sup> for WT Dongjin, 72.80 mm<sup>2</sup> for *chelin2084-T*, and 29.37 mm<sup>2</sup> for WT Zhonghua11, but only 16.42 mm<sup>2</sup> for *chenat1709-T* (Fig. 6B, upper panel; Fig. 6C). We have also transfected calli with the RNAi vector of *OsCHELIN2084* and analyzed the phenotypes. The calli transfected with *chelin2084-RNAi* vectors showed rapid proliferation, which is similar with that of the T-DNA insertion mutant of *OsCHELIN2084* (Additional file 1: Fig. S5F, G). Scanning electron microscopy (SEM) revealed that *chelin2084-T* calli consisted of many globular nodules with turgid cells, whereas WT Dongjin and Zhonghua11 calli contained fewer globular nodules and the cells were flaccid, and *chenat1709-T* cells were extremely enlarged (Fig. 6B, middle panel). *chelin2084-T* callus cells were globular and compact, which is characteristic of embryogenic calli [43, 44], whereas *chenat1709-T* callus cells showed the characteristics of non-embryogenic callus [44] (Fig. 6B, upper panel). These results indicate that che-

lncNAT *OsCHENAT1709* is required for cell pluripotency, whereas che-lincRNA *OsCHELIN2084* suppresses cell dedifferentiation.

Next, we examined the regeneration process of the mutants. Green spots emerged much earlier in *chenat1709-T* than that of WT plants, whereas *chelin2084-T* showed opposite, the green spots in *chelin2084-T* formed later than that of WT plants (Fig. 6B; Additional file 1: Fig. S5H). Moreover, the green spots of *chenat1709-T* appeared lustrous and were covered with sickle-shaped trichomes, which could eventually transform into shoots, while the green spots of *chelin2084-T* were unorganized, pale green, and covered with white hairs that did not give rise to shoots (Fig. 6B, bottom panel). These results demonstrate that che-lncNAT *OsCHENAT1709* suppresses plant regeneration while che-lincRNA *OsCHELIN2084* is required for regeneration; this is consistent with the expression patterns of these cheRNAs (Fig. 6A). Specifically, *OsCHENAT1709* was highly expressed in callus, and loss of function of this cheRNA impaired cellular pluripotency, whereas *OsCHELIN2084* was highly expressed in embryos, and its loss of function reduced the likelihood of plant regeneration. In addition, we observed significant differences in plant architecture between both the mutants and WT plants. The *OsCHELIN2084* mutant plants (*chelin2084-T*) had longer panicles and wider seeds than the WT (Fig. 6D; Additional file 1: Fig. S5I), while the *OsCHENAT1709* mutant plants (*chenat1709-T* and *OsCHENAT1709-RNAi*) had more tillers (Fig. 6E; Additional file 1: Fig. S5J), suggesting that this cheRNA might regulate cell division or bud formation and development. These findings further support the roles of these cheRNAs in both cellular reprogramming and crop traits.

Together, our findings demonstrate that *OsCHENAT1709* and *OsCHELIN2084* are associated with the process of somatic embryogenesis, in which embryogenic-competent cells respond to environmental and phytohormone signals in culture medium and develop into somatic embryos. These cheRNAs also regulate grain size and panicle size.

Finally, to investigate the potential mechanisms employed by these two cheRNAs in regulating somatic embryogenesis, we examined their neighboring genes. Notably, we detected the “CCGCCWCC” and “CCWCCMCC” motifs in che-lincRNA *OsCHELIN2084* and the “CGGCGGC” and “GGNGGNGG” motifs in the promoter regions of its neighboring genes, encoding Ubiquitin-protein ligase (LOC\_Os08g35070) and Subtilisin-like serine protease (LOC\_Os08g35090) (Fig. 6F; Additional file 1: Fig. S6A), and their family members have been reported to regulate organ development [45–47]. No che-RNA-associated motif was detected in che-lncNAT *OsCHENAT1709*. Thus, we analyzed the expression correlation between *OsCHELIN2084* and its neighboring genes LOC\_Os08g35070 and LOC\_Os08g35090, and we found that their expression patterns were positively correlated (Fig. 6G). We further analyzed the phenotypes of the knock-out transgenic plants of LOC\_Os08g35070 and LOC\_Os08g35090 respectively and found that the loss of function of LOC\_Os08g35090 showed more rapid callus proliferation and wider seeds than that of the control plants transferred with empty vector (Additional file 1: Fig. S6B, C, D), which is similar with that of the loss of function mutant of *OsCHELIN2084*, while loss of function of LOC\_Os08g35070 promoted callus proliferation but not significantly affected seed size (Additional file 1: Fig. S6B, C, D). These results are consistent with the roles of che-lincRNAs in promoting the expression of their neighboring genes, further supporting the hypothesis that che-lincRNAs function as *cis*-regulatory elements during cellular reprogramming.

Collectively, these data suggest that che-lincRNA loci act as transcriptional regulators *in cis* and are required for embryo regeneration.

## Discussion

Plants have the remarkable ability to generate a pluripotent cell mass that acquires competence for subsequent tissue regeneration [48, 49]. This cell fate transition is accompanied by epigenetic changes [6]. Global reprogramming of DNA methylation, histone modification, and chromatin remodeling is required for the cell fate transition [7, 50–52]. Therefore, global changes in the chromatin landscape define gene expression patterns. For example, during callus formation, the loss of DNA methylation deregulates the expression of protein-coding genes involved in certain biological processes. How the global reprogramming of chromatin is regulated during the cell fate transition is not fully understood. In animals, cell totipotency is thought to rely primarily on the unique chromatin of totipotent cells or on an RNA-centric posttranscriptional regulation program [53]. cheRNAs are thought to function as epigenetic regulators that play important roles in creating and/or maintaining chromatin states that influence changes in gene expression during development [11–14]. In this study, we identified cheRNAs from mature rice embryos, callus induced from mature embryos, regenerated greenish calli, and shoots and showed that the cheRNAs likely regulate the expression patterns of specific genes during somatic cell regeneration.

A total of 2284 cheRNAs were identified, which mainly consisted of lincRNAs and snoRNAs. The composition of rice cheRNAs is similar to that reported in mammals, indicating that the roles of lincRNAs and snoRNAs in regulating chromatin status are conserved between plants and animals. In addition, these cheRNAs have different characteristics from other ncRNAs, especially their level of conservation: che-lincRNAs and che-snoRNAs are highly conserved across plant species and in different rice varieties, pointing to the strong evolutionary pressure on cheRNAs and their fundamental functions. For example, in the che-lincRNA *OsCHENAT1564* and che-lincRNA *OsCHELIN0935*, several SNPs are significantly associated with rice traits. Thus, it is important to further analyze the functions of individual cheRNAs.

Another characteristic of cheRNAs is that their enrichment on chromatin is dynamic during cellular reprogramming, suggesting that they might shuttle between chromatin and the nucleoplasm. Their dynamic chromatin enrichment patterns might be associated with their roles in regulating gene expression during the cell fate transition. For example, chromatin associated lincRNA *XIST* *in cis* regulates X chromosome inactivity over long genomic distance; and chromatin associated lincRNA *FIRRE* has both *trans*- and *cis*-acting effects on epigenetic features [54]. In addition, the dissociation of lincRNAs from chromatin is also important for their regulatory roles, such as lincRNA *A-ROD* was shown to enhance its upstream gene *DKK1* transcription at its release from chromatin [55]. We have observed correlations between the expression patterns of che-lincRNAs and their adjacent genes, and examined the specific functions of these adjacent genes. We found, for example, that the adjacent genes of che-lincRNAs that are enriched on chromatin during cell dedifferentiation but dissociate from chromatin during plant regeneration primarily include genes encoding proteins required for gene transcription; by contrast, the adjacent genes of che-lincRNAs with the opposite enrichment pattern are mainly involved in protein phosphorylation (Fig. 4G; Additional file 6:

Table S5). These pathways might be essential for *in vitro/in planta* regeneration [7, 8]. Thus, cheRNAs could function as important components of the regulatory networks of somatic cell reprogramming.

Previous studies have showed that lncRNAs could regulate chromatin remodeling by mediating post-translational modification which is mostly related to histones [18]. For example, three lncRNAs *COLD ASSISTED INTRONIC NONCODING RNA (COLDAIR)*, *COOLAIR*, and *COLDWRAP* regulate *FLOWERING LOCUS C (FLC)* transcription by mediating H3K27me3 and H3K4me3 deposition [56–59]. lncRNA *AUXIN REGULATED PROMOTER LOOP (APOLO)* and *MARNERAL SILENCING (MARS)* negatively regulate H3K27me3 deposition [60, 61], whereas NAT-lncRNA *MADS AFFECTING FLOWERING4 (MAS)* [62] and lncRNA *LRK Antisense Intergenic RNA (LAIR)* [63] positively regulate H3K4me3 deposition. Intriguingly, our data showed that che-lncRNAs are positively correlated with the active histone marks H3K4me3 and H3K27ac modifications and negatively correlated with the suppressive histone mark H3K27me3 modification of their neighboring genes, whereas other histone modifications and DNA methylations are not affected by che-lncRNAs. These data implied that chromatin remodeling regulatory lncRNAs might be inclined to regulate target gene expression through mediating H3K27me3, H3K4me3, and/or H3K27ac modifications. In addition to mediate post-translational modifications, cheRNAs might also recruit transcriptional factors (TFs) to regulate gene expression, as TFs have capacity of binding snRNAs and lncRNAs [64–67].

Lastly, the extensive regeneration abilities of plants are important for their survival. Sustained stem cell activity in meristems ensures that plants undergo unlimited growth to optimize the use of resources and to heal local damage via tissue regeneration [8, 48, 49]. Thus, cheRNAs involved in somatic cell reprogramming could also play roles in organ development and stress responses. We indeed observed correlations between cheRNAs and crop traits by performing large-scale analysis of the phenotypes of mutants and transgenic plants. Most of these cheRNAs regulate tissue size, including seed size and panicle size, which are essential for grain yield. Considering their high conservation across rice varieties, these cheRNAs have great potential for use in crop trait improvement and crop breeding in the future.

## Conclusions

We systematically investigated cheRNAs in rice during callus induction, proliferation, and regeneration. These cheRNAs, which are highly conserved across plant species, shuttle between chromatin and the nucleoplasm during somatic cell regeneration. They regulate the expression of neighboring genes via specific RNA motifs, and mutant analysis implies they might be associated with plant size and seed morphology. Investigation of the functions of two che-lncRNAs supported their roles in *cis*-regulating, plant regeneration and rice traits regulation.

## Methods

### Extraction of chromatin-enriched RNAs

Three-gram samples (mature embryos, undifferentiated embryogenic callus, differential callus, and shoots) were ground with liquid nitrogen into fine powder and transferred

into an ice-cold 50 ml tube with 20 ml cell lysis buffer (20 mM Tris-HCl, pH 7.4, 20 mM KCl, 2 mM EDTA, 2.5 mM MgCl<sub>2</sub>, 25% glycerol, 250 mM sucrose, and 5 mM DTT, cocktail plant protease inhibitor, 5 U/ml RNase inhibitor). After homogenization by vortexing, the extracts were kept on ice for 15 min. Then, the homogenate was filtered through two layers of Miracloth. After centrifugation at 4 °C and 2500g for 10 min, the supernatant was removed and collected as the cytoplasmic fraction for western blot, and the pellet was resuspended and washed once with 2 ml cell lysis buffer. The pellet was then resuspended in 5 ml resuspension buffer (20 mM Tris-HCl, pH 7.4, 25% glycerol, 2.5 mM MgCl<sub>2</sub>, 0.2% Triton X-100, and 5 mM DTT, 1 U/ml RNase inhibitor) and centrifuged at 4 °C, 2500g for 10 min. The pellet was washed three times using resuspension buffer. The supernatant was completely removed, and the nuclei were resuspended with 500 µl gradient buffer 1 (10 mM Tris-HCl, pH 8.0, 250 mM sucrose, 10 mM MgCl<sub>2</sub>, 1% Triton X-100, and 5 mM β-mercaptoethanol, cocktail plant protease inhibitor, 10 U/ml RNase inhibitor). A 2-ml tube with round bottom was prepared, and 500 µl gradient buffer 2 (10 mM Tris-HCl, pH 8.0, 1.7 M sucrose, 2 mM MgCl<sub>2</sub>, 0.15% Triton X-100, and 5 mM β-mercaptoethanol, cocktail plant protease inhibitor, 10 U/ml RNase inhibitor) was added. Gradient buffer 1 containing samples was transferred carefully on the top of gradient buffer 2 and centrifuged at 4 °C for 10 min at 12000 rpm. The supernatant was thoroughly discarded and resuspended with 500 µl glycerol buffer (20 mM Tris-HCl, pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT, 0.125 mM PMSE, 10 mM β-mercaptoethanol, and 125 U/ml RNase inhibitor). The suspension was transferred into 500 µl urea buffer (10 mM HEPES, pH 7.6, 7.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.3 M NaCl, 1 M urea, 1% NP-40, 1 mM DTT, 0.5 mM PMSE, cocktail plant protease inhibitor, 10 mM β-mercaptoethanol, and 125 U/ml RNase inhibitor), vortexed, and kept in ice for 5 min. It was then centrifuged at 4 °C, 13,000 rpm for 2 min, and the supernatant was collected as the nucleoplasmic fraction. The pellet was washed again with glycerol buffer and urea buffer as mentioned above. The pellet was retained as the chromatin fraction. Several nucleoplasmic and chromatin fractions were collected for western blot analysis.

For RNA extraction, the pellet was resuspended in 1 ml TRIzol. The nucleoplasmic fraction was mixed with 2.632 volumes of RNA precipitation solution (ethanol containing 0.15 M sodium acetate, pH 5.5), vortexed thoroughly, and kept at - 20 °C overnight. The pellet was vortexed and centrifuged at 4 °C, 18,000g for 15 min. The supernatant was discarded and the pellet air-dried. Then, 1 ml TRIzol was added to lyse the pellet. Two hundred microliters of chloroform was then added, vortexed for 10 s, and kept at room temperature for 5 min. The mixture was centrifuged at 4 °C and 12,000g for 15 min. The supernatant was transferred to a new tube, and 1.5 volume of GXP2 buffer (HiPure HP Plant RNA Mini Kit, Magen, China) was added. The solution was vortexed and transferred into a Spin Column (Plant/Fungi Total RNA Purification Kit, NORGEN, Canada). The extraction procedures were performed according to the manufacturer's instructions (Plant/Fungi Total RNA Purification Kit, NORGEN, Canada). The RNA samples were quantified using a Nanodrop 2000 and stored at - 80 °C.

To verify the purity of each fraction, the total protein, cytoplasmic, nucleoplasmic, and chromatin protein fractions were subsequently analyzed using western blot. For immunoblot analysis, antibodies against GAPDH (BPI, AbP80006-A-SE) and Histone H3 (Abcam, Ab1791) were used for cytoplasmic and chromatin fraction-specific markers, respectively.

### Library construction and sequencing

The extracted RNA was prepared for RNA sequencing and deep sequencing of intermediate-size RNAs (50 to 300 nt) with two biological replicates. For RNA-seq, the total RNA quantity and purity were analyzed using a Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, CA, USA) with RIN number > 7.0. The preparation of whole-transcriptome libraries and deep sequencing were performed by the Annoroad Gene Technology Corporation. Libraries were controlled for quality and quantitated using the BioAnalyzer 2100 system and qPCR (Kapa Biosystems, Woburn, MA). The resulting libraries were sequenced initially on a HiSeq 2000 instrument that generated paired end reads of 150 nt. For intermediate-size RNA-seq, the library size selection was performed by gel electrophoresis with a range of 50–300 bp. Approximately 1 µg of total RNA was used to prepare the library according to the protocol of the TruSeq Small RNA Sample Prep Kits (Illumina, San Diego, USA). The libraries were subsequently sequenced on the Illumina HiSeq2500 platform at LC-BIO (Hangzhou, China) following the manufacturer's instructions, and the full-length pair-end reads were obtained. The datasets generated during the current study are available in the SRA database of NCBI (SRP338667) [68].

### Sequencing data processing and novel ncRNA identification

For transcriptome sequencing data, the read quality was inspected using FastQC v0.11.9 and then aligned to the *Oryza sativa* genome assembly (MSU RGAP Release 7 [69]) using TopHat v2.1.1 [70]. The transcript from each dataset was de novo assembled independently using Cufflinks v2.2.1 [71]. The CPE and SNE transcripts from all samples were pooled and merged to generate a single final GTF file using the Cuffmerge program, and the abundance of all transcripts was estimated by Cuffdiff based on the final GTF file.

For intermediate-sized RNA sequencing data, the adapters in raw reads were removed using Cutadapt v3.0 [72], and the untrimmed paired reads were merged using PEAR v0.9.6 [73], combining the trimmed first-end reads into single read FastQ files. Reads were then aligned against intermediate-size ncRNAs for perfect matches using the STAR v2.7.5 [74] program with the following priority: rRNA (RAP-DB [75]), tRNA (RAP-DB), snRNA (Rfam database v14.5 [76]), and annotated snoRNA (Rfam database v14.5 and published articles [27]). Reads that could not be mapped to either class above were converted into FastA format using the fastx\_colapser program from FASTX-Toolkit v0.0.14 ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)) and then aligned to genome assembly using Bowtie v2.4.1 [77]. The novel snoRNAs were identified using snoSeekerNGS-1.0 [78] against the alignment files, and the prediction results were gathered and filtered with the conserved box motif, resulting in the novel snoRNA candidates. All the mapped reads above were then aligned to the annotated intermediate-size ncRNAs and novel snoRNA candidates and filtered by the length coverage using the manual Perl script. The effective reads were aligned to the genome and counted using featureCounts v2.0.1 [79]. Only ncRNAs with more than 3 supported reads in at least 2 samples or more than 10 supported reads in at least 1 sample were kept. The expression quantity of the intermediate-size ncRNAs was normalized by RPM (reads per million).

Raw count matrixes were counted by featureCounts, and differential expression analysis was performed by DESeq2 v1.32.0 [80] in R (version 4.1.0), setting an adjusted  $p$  value less than 0.05 as the cutoff for statistical significance. The ncRNAs with CPE versus SNE fold change  $> 1.2$  were classified as chromatin-enriched RNAs (che-RNAs), while those with fold change  $< 0.8$  were classified as soluble nuclear extract enriched RNAs (sne-RNAs). The che-lincRNAs and che-lincNATs were identified by estimating the coding potential using CPC2 [81] and comparing the genomic coordinate and strand with the MSU7.0 annotated mRNA transcripts using intersectBed (bedtools v2.29.2 [82]). The possible TEs derived che-lincRNAs (TE che-lincRNA) were identified by overlapping with known rice TEs. The rice TE annotations used in this study are obtained from the outputs of the RepeatMasker which were filtered to remove some non-TE elements, including low complexity, satellites, simple repeats, and ncRNAs. Identified cheRNAs transcripts were extracted and added to the MSU7.0 mRNA annotation, and the expression abundance in Input samples was estimated using Cuffdiff.

#### Constructs for genetic transformation

To construct the RNAi transformation plasmid, 300 nt DNA fragments of chr8 22101135 to 22100836 for *OsCHELIN2084* and chr6 23891471 to 23898765 for *OsCHENATI709* were ligated to modified pRTV vector [83]. And the pRHCas9 vector [83] was used to construct the knock out mutant. The sgRNA target sites by CRISPR-cas9 are chr8 22107716 to 22107735 for LOC\_Os08g35070 and chr8 22111291 to 22111310 for LOC\_Os08g35090 respectively.

#### Chromatin RNA immunoprecipitation (ChRIP)

1.5 g callus and shoot were crosslinked in 30 ml of 1.0% formaldehyde under vacuum for 30 min in a desiccator attached to a vacuum pump. Then, quench cross-linking in 0.125M Glycine solution for an additional 5 min was done. Wash the samples with distilled water three times, and then ground the samples into fine powders. Nucleus were isolated, lysed and sonicated into 1 kb fragments, immunoprecipitated with histone H3 antibody (Abcam) or with IgG (Millipore). The chromatin-associated RNA was extracted using TRIzol (Invitrogen, USA), and DNase I treatment was conducted to remove DNA contamination. Then, the chromatin-associated RNA was reverse-transcribed into cDNA and qPCR reactions were performed for RNAs of interest using H3 and IgG pull-down fractions.

#### Analysis of cheRNA neighboring gene expression and genomic features

Comparisons of neighboring gene were performed using closestBed (bedtools v2.29.2) with MSU7.0-annotated non-TE mRNA transcripts relative to different genomic features. The average FPKM expression of input samples was used, and all boxplots were plotted using the R package ggplot2 v3.3.3 in R. The Pearson correlation coefficient (PCC) was calculated between the expression levels of cheRNAs and their neighboring genes in R. The  $p$  values were calculated using a Wilcoxon Mann-Whitney test.

### Analysis of epigenetic activities

The H3K27ac, H3K4me3, H3K27me3 ChIP-seq [84], and DNase-seq [31] analyzed data were downloaded from Plant Chromatin State Database (PCSD) [85]. The raw sequencing data of H4K12ac and H3K9me2 ChIP-seq and Bisulfite sequencing (BS-seq) were downloaded from the NCBI database (PRJNA386513 [86], PRJNA142153 [31], GSE126436 [87], GSE42410 [50]); all raw reads adapters were removed using cutadapt. The ChIP-seq reads were aligned to the *Oryza sativa* genome assembly (MSU RGAP Release 7 [49]) using bowtie2. The mapped reads were converted to bigwig format for visualization using bamCoverage from deepTools v3.5.1 [88]. The BS-seq reads mapping and methylation extraction were conducted using Bismark v0.23.1 [89]. The DNA methylation levels were calculated by averaging the DNA methylation ratios of all cytosine sites with coverage larger than 5 in 20 bp windows. All region profiles were computed and plotted using deepTools v3.5.1 commands. All other public datasets used in the study were listed in Additional file 4: Table S3.

### Clustering and Gene Ontology analysis

The CPE versus SNE fold change of cheRNAs was defined as the chromatin enrichment score and used to perform a time-series cluster with a series of embryo, callus, differentiated callus, and shoot sample. The time-series soft clustering analysis was conducted by the fuzzy *c*-means method in the Mfuzz [90] package v2.52.0 to identify the different chromatin enrichment variation patterns. The neighboring genes of cheRNA with different chromatin enrichment patterns were extracted, and a GO enrichment analysis was performed with AgriGOv2 [91] (<http://systemsbiology.cau.edu.cn/agriGOv2/>). The significant GO enrichment results ( $p < 0.05$ ) were summarized using the REVIGO [92] website (<http://revigo.irb.hr/>). The aggregate size indicates the significance levels of the GO terms, as determined using the Yekutieli test with false discovery rate correction. PCA (principal component analysis) was conducted with the normalized abundance FPKM of indicated RNAs using the R package factoextra v1.0.7 and plotted by ggplot2.

### Whole-genome alignment and conservation analysis

Pairwise whole-genome alignments with the *Oryza sativa* japonica genome were generated for each *Oryza* species following the UCSC pipeline. Specifically, the other *Oryza* species genomes were downloaded from the NCBI Assembly database, including *O. sativa* indica group (PRJNA353946), *O. rufipogon* (PRJEB4137), *O. nivara* (PRJNA48107), *O. barthii* (PRJNA30379), *O. glaberrima* (PRJNA13765), *O. glumaepatula* (PRJNA48429), *O. meridionalis* (PRJNA48433), *O. punctata* (PRJNA13770), *O. brachyantha* (PRJNA70533), and *L. perrieri* (PRJNA163065). All the repetitive DNA was masked from genomes using RepeatMasker v4.0.8. Each pairwise alignment was conducted using the RunLastzChain.sh script from UCSC Kent Utils setting a “Near” parameter, and the Netting and Maffing steps were performed using the UCSC pipeline program with manual scripts. All of the above computations were run in parallel in a Linux cluster. The reference-guided multiple alignments were conducted by the Roast v3 program [93]. A phylogenetic model was fitted based on the multiple alignment of the 11 *Oryza* genomes using the phyloFit program [94]. The conservation scores of each

base were calculated from the 11-way alignments based on the fitted model using the phastCons program.

#### **SnoRNA genome organization and target RNA prediction**

The identified novel snoRNAs were combined with all annotated snoRNAs, and snoRNA clusters and genome organization were determined as previously described in the literature [27]. SnoRNAs with less than 500-bp gene intervals were classified into the same cluster. All genomic, family, and chromatin enrichment information of all expressed snoRNAs was gathered and plotted into a sankey diagram using the ggplot2 extension ggforce in R. The snoRNA modification targets were predicted by PLEXY [95] for CD box snoRNAs and RNAsnoop [96] v2.4.17 for HACA box snoRNAs, and the target RNA sequences (rRNAs and snRNAs) were obtained with the annotation downloaded from the RAP-DB and Rfam databases.

#### **Insertion mutant and crop trait-associated SNP analysis**

The T-DNA insertion mutant analysis was performed as previously described [97]. The T-DNA insertion site of *OsCHELIN2084* is at chr8 22101165, and the T-DNA insertion site of *OsCHENAT1709* is at chr6 23896685. The crop trait-associated SNP GWAS (Genome Wide Association Study) data were downloaded from the Rice SNP-Seek Database [42] ([https://snp-seek.irri.org/\\_gwas.zul](https://snp-seek.irri.org/_gwas.zul)), setting the minimum  $-\log_{10}(p)$  value as 4 and the subpopulation option as “all varieties.” The crop trait-associated SNP genomic locations were compared with the che-lincRNAs’ genomic coordinates using intersectBed.

#### **Triple helix formation prediction and motif analysis**

The potential DNA:DNA:RNA triple helix sites of che-lincRNAs were predicted using Triplexator v1.3.2 [98]. The predictions were performed with the parameters of “-l 20 -e 5” and other defaulted parameters. The triplex-forming target site DNA regions were annotated and plotted using the ChIPseeker [99] v1.28.3 package across the MSU7.0 gene annotation. The triplex formation of che-lincRNAs and their neighboring gene DNA regions was tested using the Triplex Domain Finder region test program [100] (rgt-TDF v0.13.2 from the Regulatory Genomics Toolbox).

De novo motif analysis was conducted using the MEME suite [101] v4.11.2. The motif distributions were scanned using Fimo from the MEME suite, and the relative location was calculated and normalized for each transcript length. The density distribution was plotted using ggplot2 in R.

#### **Total RNA extraction and qRT-PCR**

Total RNA extraction and qRT-PCR were performed as described previously [97]. The results were presented as the relative expression levels normalized to the expression of *OsActin2*. For the semi-quantitative RT-PCR analysis, the amplification was performed in a 20- $\mu$ l reaction volume containing diluted cDNA, 0.4 mM primers, diethylpyrocarbonate-treated water, and TB Green® Premix Ex Taq™ (TAKARA). The PCR conditions were as follows: 95 °C for 30 s, followed by various cycles according to different genes of 95 °C for 10 s. The PCR products were electrophoresed in a 2.5%

agarose gel, and the images were captured. Each qRT-PCR was performed for three biological replicates. The primers used for both semi-quantitative RT-PCR and qRT-PCR are listed in Additional file 9: Table S8.

#### Tissue culture procedure

Mature, healthy seeds were sterilized by immersion in 70% ethanol for ~2 min, followed by 2.5% sodium hypochlorite solution for 30 min with shaking, and rinsed five or six times with sterile water on an ultraclean workbench. N6 was used as the main callus induction medium, and 2 mg/L 2,4-D and 30 g/L sucrose were added. The pH of the medium was adjusted to 5.8, and 3.0 g/L phytigel was added to the medium before boiling. Approximately 90 mature seeds per line, evenly distributed among two dishes, were incubated in induction medium for 15 days at 28 °C. The induced calli were transferred to subculture medium and incubated at 28 °C for 15 days. After 1 month culture, the calli were transferred to differential medium (MS, 2 mg/L 6BA, 2 mg/L KT, 0.2 mg/L IAA, 0.2 mg/L NAA, and 30 g/L sucrose; the pH of the medium was adjusted to 5.8, and 3.0 g/L phytigel was added to the medium before boiling) and incubated for 30 days at 28 °C.

#### Phenotype observations

Images of calli during the induction stage were taken by a LEICA M205FA (Germany). The images were taken after the calli were transferred to subculture medium for 0, 5, 10, and 15 days, and the sizes were calculated using ImageJ by measuring the mean size.

#### Scanning electron microscopy

To prepare histological sections, calli that had been cultured for 25 days were fixed in FAA fixative solution (50% alcohol: acetic acid: formaldehyde = 89:6:5) for 30 min under vacuo, and post-fixed in the same buffer overnight. After being dehydrated through an ethanol series and dried using a carbon dioxide critical-point dryer, the calli were cleaned with ethanol and dried at 45 °C. The dry calli were gold plated and photographed under a Hitachi S-3400 N scanning electron microscope (Japan).

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02608-y>.

**Additional file 1: Supplementary Figures 1-6.**

**Additional file 2: Table S1.** Sequencing information.

**Additional file 3: Table S2.** Information of cheRNAs.

**Additional file 4: Table S3.** Public datasets used in this study.

**Additional file 5: Table S4.** Neighboring genes which significantly correlated with cheRNAs.

**Additional file 6: Table S5.** The predicted DNA binding sites of che-lincRNAs.

**Additional file 7: Table S6.** GO terms of the neighboring genes of the che-lincRNAs from different clusters.

**Additional file 8: Table S7.** phenotype analysis of the cheRNA T-DNA insertion mutants.

**Additional file 9: Table S8.** Primers used for qRT-PCR and plasmid construct.

**Additional file 10.** Uncropped images.

**Additional file 11.** Review history.

#### Acknowledgements

We are grateful to Dr. Yumeng Sun in the School of Life Science of Sun Yat-sen University for sample collection and chromatin separation.

### Review history

The review history is available as Additional file 11.

### Peer review information

Wenjing She was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Authors' contributions

Y.C.Z., Y.F.Z., and Y.C. designed and performed the research, analyzed the data, and wrote the manuscript. J.H.H., J.P.L., L.Y., R.R.H. M.Q.L., Y.W.L., C.Y. W.L.Z., and X.S. performed the research and analyzed data. Y.C. Z. and Y.Q.C. designed the research, analyzed the data, and wrote the manuscript. The authors read and approved the final manuscript.

### Funding

This research was supported by the National Natural Science Foundation of China (No. 91940301, U1901202, 32070624 and 32100437) and the grants from Guangdong Province (2019JC05N394).

### Availability of data and materials

The datasets generated during the current study are available in the SRA database of NCBI (SRP338667) [68]. The published data used in this study were downloaded from the NCBI database (PRJNA386513 [86], PRJNA142153 [31], GSE126436 [87], GSE42410 [50]).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing financial interests.

#### Author details

<sup>1</sup>Guangdong Provincial Key Laboratory of Plant Resources, State Key Laboratory for Biocontrol, School of Life Science, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China. <sup>2</sup>MOE Key Laboratory of Gene Function and Regulation, Sun Yat-sen University, Guangzhou 510275, People's Republic of China.

Received: 22 September 2021 Accepted: 12 January 2022

Published online: 19 January 2022

### References

1. He G, Elling AA, Deng XW. The epigenome and plant development. *Annu Rev Plant Biol.* 2011;62(1):411–35. <https://doi.org/10.1146/annurev-arplant-042110-103806>.
2. Wagner D. Chromatin regulation of plant development. *Curr Opin Plant Biol.* 2003;6(1):20–8. <https://doi.org/10.1016/S1369-526602000079>.
3. Goodrich J, Tweedie S. Remembrance of things past: chromatin remodeling in plant development. *Annu Rev Cell Dev Biol.* 2002;18(1):707–46. <https://doi.org/10.1146/annurev.cellbio.18.040202.114836>.
4. Henderson IR, Jacobsen SE. Epigenetic inheritance in plants. *Nature.* 2007;447(7143):418–24. <https://doi.org/10.1038/nature05917>.
5. Xu Y, Zhang M, Li W, Zhu X, Bao X, Qin B, et al. Transcriptional control of somatic cell reprogramming. *Trends Cell Biol.* 2016;26(4):272–88. <https://doi.org/10.1016/j.tcb.2015.12.003>.
6. Xu L, Huang H. Genetic and epigenetic controls of plant regeneration. *Curr Top Dev Biol.* 2014;108:1–33. <https://doi.org/10.1016/B978-0-12-391498-9.00009-7>.
7. Lee K, Seo PJ. Dynamic epigenetic changes during plant regeneration. *Trends Plant Sci.* 2018;23(3):235–47. <https://doi.org/10.1016/j.tplants.2017.11.009>.
8. Feher A. Somatic embryogenesis - stress-induced remodeling of plant cell fate. *BBA.* 2015;1849(4):385–402. <https://doi.org/10.1016/j.bbaggm.2014.07.005>.
9. Wierzbicki AT, Blevins T, Swiezewski S. Long noncoding RNAs in plants. *Annu Rev Plant Biol.* 2021;72(1):245–71. <https://doi.org/10.1146/annurev-arplant-093020-035446>.
10. Yu Y, Zhang Y, Chen X, Chen Y. Plant Noncoding RNAs: Hidden players in development and stress responses. *Annu Rev Cell Dev Biol.* 2019;35(1):407–31. <https://doi.org/10.1146/annurev-cellbio-100818-125218>.
11. GH D, Kelley DR, Tenen D, Bernstein B, Rinn JL. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* 2016;17(1):28. <https://doi.org/10.1186/s13059-016-0878-3>.
12. Werner MS, Ruthenburg AJ. Nuclear fractionation reveals thousands of chromatin-tethered noncoding RNAs adjacent to active genes. *Cell Rep.* 2015;12(7):1089–98. <https://doi.org/10.1016/j.celrep.2015.07.033>.
13. Werner MS, Sullivan MA, Shah RN, Nadadur RD, Grzybowski AT, Galat V, et al. Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. *Nat Struct Mol Biol.* 2017;24(7):596–603. <https://doi.org/10.1038/nsmb.3424>.
14. Xiao R, Chen JY, Liang Z, Luo D, Chen G, Lu ZJ, et al. Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. *Cell.* 2019;178(1):107–21 e118. <https://doi.org/10.1016/j.cell.2019.06.001>.

15. Akhtar A, Zink D, Becker PB. Chromodomains are protein-RNA interaction modules. *Nature*. 2000;407(6802):405–9. <https://doi.org/10.1038/35030169>.
16. Maison C, Bailly D, Peters AH, Quivy JP, Roche D, Taddei A, et al. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet*. 2002;30(3):329–34. <https://doi.org/10.1038/ng843>.
17. Acharya S, Hartmann M, Erhardt S. Chromatin-associated noncoding RNAs in development and inheritance. *Wiley Interdiscip Rev RNA*. 2017;8(6):e1435.
18. Fonouni-Farde C, Ariel F, Crespi M. Plant Long noncoding RNAs: new players in the field of post-transcriptional regulations. *Noncoding RNA*. 2021;7(1):12.
19. Belcheva A, Mishkova R. Histamine content in lymph nodes from patients with malignant lymphomas. *Inflamm Res*. 1995;44(Suppl 1):S86–7. <https://doi.org/10.1007/BF01674409>.
20. Jin J, Lu P, Xu Y, Li Z, Yu S, Liu J, et al. PLncDB V2.0: a comprehensive encyclopedia of plant long noncoding RNAs. *Nucleic Acids Res*. 2021;49(D1):D1489–95. <https://doi.org/10.1093/nar/gkaa910>.
21. Sweeney BA, Hoksza D, Nawrocki EP, Ribas CE, Madeira F, Cannone JJ, et al. R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nat Commun*. 2021;12(1):3494. <https://doi.org/10.1038/s41467-021-23555-5>.
22. Zhou B, Ji B, Liu K, Hu G, Wang F, Chen Q, et al. EVLncRNAs 2.0: an updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res*. 2021;49(D1):D86–91. <https://doi.org/10.1093/nar/gkaa1076>.
23. Hu Y, Lai Y, Chen X, Zhou DX, Zhao Y. Distribution pattern of histone marks potentially determines their roles in transcription and RNA processing in rice. *J Plant Physiol*. 2020;249:153167. <https://doi.org/10.1016/j.jplph.2020.153167>.
24. Bonetti A, Agostini F, Suzuki AM, Hashimoto K, Pascarella G, Gimenez J, et al. RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat Commun*. 2020;11(1):1018. <https://doi.org/10.1038/s41467-020-14337-6>.
25. Li X, Zhou B, Chen L, Gou LT, Li H, Fu XD. GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol*. 2017;35(10):940–50. <https://doi.org/10.1038/nbt.3968>.
26. Schubert T, Pusch MC, Diermeier S, Benes V, Kremmer E, Imhof A, et al. Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. *Mol Cell*. 2012;48(3):434–44. <https://doi.org/10.1016/j.molcel.2012.08.021>.
27. Liu TT, Zhu D, Chen W, Deng W, He H, He G, et al. A global identification and analysis of small nucleolar RNAs and possible intermediate-sized non-coding RNAs in *Oryza sativa*. *Mol Plant*. 2013;6(3):830–46. <https://doi.org/10.1093/mp/sss087>.
28. Chen CL, Liang D, Zhou H, Zhuo M, Chen YQ, Qu LH. The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*. *Nucleic Acids Res*. 2003;31(10):2601–13. <https://doi.org/10.1093/nar/gkg373>.
29. Li X, Fu XD. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat Rev Genet*. 2019;20(9):503–19. <https://doi.org/10.1038/s41576-019-0135-1>.
30. Sun J, He N, Niu L, Huang Y, Shen W, Zhang Y, et al. Global quantitative mapping of enhancers in rice by STARR-seq. *Genom Proteom Bioinf*. 2019;17(2):140–53. <https://doi.org/10.1016/j.gpb.2018.11.003>.
31. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, et al. High-resolution mapping of open chromatin in the rice genome. *Genome Res*. 2012;22(1):151–62. <https://doi.org/10.1101/gr.131342.111>.
32. Mas AM, Huarte M. lncRNA-DNA hybrids regulate distant genes. *EMBO Rep*. 2020;21(3):e50107.
33. Li Y, Syed J, Sugiyama H. RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem Biol*. 2016;23(11):1325–33. <https://doi.org/10.1016/j.chembiol.2016.09.011>.
34. Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, et al. Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell*. 2003;15(8):1771–80. <https://doi.org/10.1105/tpc.012559>.
35. Miyao A, Iwasaki Y, Kitano H, Itoh J, Maekawa M, Murata K, et al. A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol Biol*. 2007;63(5):625–35. <https://doi.org/10.1007/s11103-006-9118-7>.
36. Sallaud C, Gay C, Larmande P, Bes M, Piffanelli P, Piegu B, et al. High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. *Plant J*. 2004;39(3):450–64. <https://doi.org/10.1111/j.1365-313X.2004.02145.x>.
37. Droc G, Ruiz M, Larmande P, Pereira A, Piffanelli P, Morel JB, et al. OryGenesDB: a database for rice reverse genetics. *Nucleic Acids Res*. 2006;34(Database issue):D736–40. <https://doi.org/10.1093/nar/gkj012>.
38. van Enckevort LJ, Droc G, Piffanelli P, Greco R, Gagneur C, Weber C, et al. EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice. *Plant Mol Biol*. 2005;59(1):99–110. <https://doi.org/10.1007/s11103-005-8532-6>.
39. Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J, et al. T-DNA insertional mutagenesis for functional genomics in rice. *Plant J*. 2000;22(6):561–70. <https://doi.org/10.1046/j.1365-313x.2000.00767.x>.
40. Zhang J, Li C, Wu C, Xiong L, Chen G, Zhang Q, et al. RMD: a rice mutant database for functional analysis of the rice genome. *Nucleic Acids Res*. 2006;34(Database issue):D745–8. <https://doi.org/10.1093/nar/gkj016>.
41. Wang CC, Yu H, Huang J, Wang WS, Faruquee M, Zhang F, et al. Towards a deeper haplotype mining of complex traits in rice with RFBG v2.0. *Plant Biotechnol J*. 2020;18(1):14–6. <https://doi.org/10.1111/pbi.13215>.
42. Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, et al. Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res*. 2017;45(D1):D1075–81. <https://doi.org/10.1093/nar/gkw1135>.
43. Bevitori R, Papielarska-Konieczna M, dos Santos EM, Grossi-de-Sa MF, Petrofeza S. Morpho-anatomical characterization of mature embryo-derived callus of rice (*Oryza sativa* L.) suitable for transformation. *Protoplasma*. 2014;251(3):545–54. <https://doi.org/10.1007/s00709-013-0553-4>.
44. Lopez-Ruiz BA, Juarez-Gonzalez VT, Sandoval-Zapotitla E, Dinkova TD. Development-related miRNA expression and target regulation during staggered in vitro plant regeneration of Tuxpeno VS-535 maize cultivar. *Int J Mol Sci*. 2019;20(9):2079.
45. Schardon K, Hohl M, Graff L, Pfannstiel J, Schulze W, Stintzi A, et al. Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases. *Science*. 2016;354(6319):1594–7. <https://doi.org/10.1126/science.aai8550>.

46. Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, et al. A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development* (Cambridge, England). 2001;128(23):4681–9.
47. Kim B, Piao R, Lee G, Koh E, Lee Y, Woo S, et al. OsCOP1 regulates embryo development and flavonoid biosynthesis in rice (*Oryza sativa* L.). *Theor Appl Genet*. 2021;134(8):2587–601. <https://doi.org/10.1007/s00122-021-03844-9>.
48. Birnbaum KD, Sanchez AA. Slicing across kingdoms: regeneration in plants and animals. *Cell*. 2008;132(4):697–710. <https://doi.org/10.1016/j.cell.2008.01.040>.
49. Zimmerman JL. Somatic embryogenesis: a model for early development in higher plants. *Plant Cell*. 1993;5(10):1411–23. <https://doi.org/10.2307/3869792>.
50. Stroud H, Ding B, Simon SA, Feng S, Bellizzi M, Pellegrini M, et al. Plants regenerated from tissue culture contain stable epigenome changes in rice. *eLife*. 2013;2:e00354. <https://doi.org/10.7554/eLife.00354>.
51. Williams L, Zhao J, Morozova N, Li Y, Avivi Y, Grafi G. Chromatin reorganization accompanying cellular dedifferentiation is associated with modifications of histone H3, redistribution of HP1, and activation of E2F-target genes. *Dev Dyn*. 2003;228(1):113–20. <https://doi.org/10.1002/dvdy.10348>.
52. De-la-Pena C, Nic-Can Gl, Galaz-Avalos RM, Avilez-Montalvo R, Loyola-Vargas VM. The role of chromatin modifications in somatic embryogenesis in plants. *Front Plant Sci*. 2015;6:635. <https://doi.org/10.3389/fpls.2015.00635>.
53. Seydoux G, Braun RE. Pathway to totipotency: lessons from germ cells. *Cell*. 2006;127(5):891–904. <https://doi.org/10.1016/j.cell.2006.11.016>.
54. Fang H, Bonora G, Lewandowski JP, Thakur J, Filippova GN, Henikoff S, et al. Trans- and cis-acting effects of Firre on epigenetic features of the inactive X chromosome. *Nat Commun*. 2020;11(1):6053. <https://doi.org/10.1038/s41467-020-19879-3>.
55. Ntini E, Louloupi A, Liz J, Muino JM, Marsico A, Orom UAV. Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. *Nat Commun*. 2018;9(1):1636. <https://doi.org/10.1038/s41467-018-04100-3>.
56. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*. 2011;331(6013):76–9. <https://doi.org/10.1126/science.1197349>.
57. Kim DH, Sung S. Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. *Dev Cell*. 2017;40(3):302–12 e304. <https://doi.org/10.1016/j.devcel.2016.12.021>.
58. Csorba T, Questa JJ, Sun Q, Dean C. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc Natl Acad Sci U S A*. 2014;111(45):16160–5. <https://doi.org/10.1073/pnas.1419030111>.
59. Tian Y, Zheng H, Zhang F, Wang S, Ji X, Xu C, et al. PRC2 recruitment and H3K27me3 deposition at FLC require FCA binding of COOLAIR. *Sci Adv*. 2019;5(4):eaau7246.
60. Ariel F, Jegu T, Latrasse D, Romero-Barrios N, Christ A, Benhamed M, et al. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol Cell*. 2014;55(3):383–96. <https://doi.org/10.1016/j.molcel.2014.06.011>.
61. 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):1055–65 e1054. <https://doi.org/10.1016/j.molcel.2019.12.015>.
62. Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat Commun*. 2018;9(1):5056. <https://doi.org/10.1038/s41467-018-07500-7>.
63. Wang Y, Luo X, Sun F, Hu J, Zha X, Su W, et al. Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat Commun*. 2018;9(1):3516. <https://doi.org/10.1038/s41467-018-05829-7>.
64. Higuchi T, Anzai K, Kobayashi S. U7 snRNA acts as a transcriptional regulator interacting with an inverted CCAAT sequence-binding transcription factor NF-Y. *BBA*. 2008;1780(2):274–81. <https://doi.org/10.1016/j.bbagen.2007.11.005>.
65. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011;43(7):621–9. <https://doi.org/10.1038/ng.848>.
66. Holmes ZE, Hamilton DJ, Hwang T, Parsonnet NV, Rinn JL, Wuttke DS, et al. The Sox2 transcription factor binds RNA. *Nat Commun*. 2020;11(1):1805. <https://doi.org/10.1038/s41467-020-15571-8>.
67. Long Y, Wang X, Youmans DT, Cech TR. How do lncRNAs regulate transcription? *Sci Adv*. 2017;3(9):eaao2110.
68. Zhang Y, Cheng Y. Genome-wide analysis and functional annotation of chromatin-enriched noncoding RNAs in rice during somatic cell regeneration. *Datasets*. NCBI. <https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP338667>. (2021).
69. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* (N Y). 2013;6(1):4.
70. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
71. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*. 2011;27(17):2325–9. <https://doi.org/10.1093/bioinformatics/btr355>.
72. Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J Comput Biol*. 2017;24(11):1138–43. <https://doi.org/10.1089/cmb.2017.0096>.
73. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*. 2014;30(5):614–20. <https://doi.org/10.1093/bioinformatics/btt593>.
74. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
75. Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, et al. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol*. 2013;54(2):e6. <https://doi.org/10.1093/pcp/pcs183>.
76. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res*. 2021;49(D1):D192–200. <https://doi.org/10.1093/nar/gkaa1047>.
77. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
78. Yang JH, Zhang XC, Huang ZP, Zhou H, Huang MB, Zhang S, et al. snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. *Nucleic Acids Res*. 2006;34(18):5112–23. <https://doi.org/10.1093/nar/gkl672>.
79. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30. <https://doi.org/10.1093/bioinformatics/btt656>.

80. 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. <https://doi.org/10.1186/s13059-014-0550-8>.
81. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 2017;45(W1):W12–6. <https://doi.org/10.1093/nar/gkx428>.
82. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
83. He F, Zhang F, Sun W, Ning Y, Wang GL. A versatile vector toolkit for functional analysis of rice genes. *Rice (N Y).* 2018; 11(1):27.
84. Zhang K, Xu W, Wang C, Yi X, Zhang W, Su Z. Differential deposition of H2AZ in combination with histone modifications within related genes in *Oryza sativa* callus and seedling. *Plant J.* 2017;89(2):264–77. <https://doi.org/10.1111/tpj.13381>.
85. Liu Y, Tian T, Zhang K, You Q, Yan H, Zhao N, et al. PCSD: a plant chromatin state database. *Nucleic Acids Res.* 2018; 46(D1):D1157–67. <https://doi.org/10.1093/nar/gkx919>.
86. Zhao D, Hamilton JP, Vaillancourt B, Zhang W, Eizenga GC, Cui Y, et al. The unique epigenetic features of Pack-MULEs and their impact on chromosomal base composition and expression spectrum. *Nucleic Acids Res.* 2018;46(5):2380–97. <https://doi.org/10.1093/nar/gky025>.
87. Wang M, Chen M. Evolution of heterochromatin and heterochromatin genes in the *Oryza* genomes reveals a new heterochromatin-euchromatin boundary [ChIP-Seq]. *Datasets. Gene Expression Omnibus.* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=%20GSE126436>. (2019).
88. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):W160–5. <https://doi.org/10.1093/nar/gkw257>.
89. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011;27(11):1571–2. <https://doi.org/10.1093/bioinformatics/btr167>.
90. Kumar L, M EF. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation.* 2007;2(1):5–7. <https://doi.org/10.6026/97320630002005>.
91. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 2017;45(W1):W122–9. <https://doi.org/10.1093/nar/gkx382>.
92. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6(7):e21800. <https://doi.org/10.1371/journal.pone.0021800>.
93. Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat Genet.* 2018;50(2):285–96. <https://doi.org/10.1038/s41588-018-0040-0>.
94. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 2005;15(8):1034–50. <https://doi.org/10.1101/gr.3715005>.
95. Kehr S, Bartschat S, Stadler PF, Tafer H. PLEXY: efficient target prediction for box C/D snoRNAs. *Bioinformatics.* 2011;27(2): 279–80. <https://doi.org/10.1093/bioinformatics/btq642>.
96. Tafer H, Kehr S, Hertel J, Hofacker IL, Stadler PF. RNAsnoop: efficient target prediction for H/ACA snoRNAs. *Bioinformatics.* 2010;26(5):610–6. <https://doi.org/10.1093/bioinformatics/btp680>.
97. Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, et al. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 2014;15(12):512. <https://doi.org/10.1186/s13059-014-0512-1>.
98. Buske FA, Bauer DC, Mattick JS, Bailey TL. Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.* 2012;22(7):1372–81. <https://doi.org/10.1101/gr.130237.111>.
99. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics.* 2015;31(14):2382–3. <https://doi.org/10.1093/bioinformatics/btv145>.
100. Kuo CC, Hanzelmann S, Senturk Cetin N, Frank S, Zajzon B, Derks JP, et al. Detection of RNA-DNA binding sites in long noncoding RNAs. *Nucleic Acids Res.* 2019;47(6):e32. <https://doi.org/10.1093/nar/gkz037>.
101. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(Web Server issue):W202–8.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

