

Mutation of a major CG methylase alters genome-wide lncRNA expression in rice

Juzuo Li ^{1,†}, Ning Li ^{1,†}, Ling Zhu,¹ Zhibin Zhang,¹ Xiaochong Li,¹ Jinbin Wang,¹ Hongwei Xun,² Jing Zhao,¹ Xiaofei Wang,¹ Tianya Wang,¹ Hongyan Wang,^{3,*} Bao Liu,¹ Yu Li,⁴ and Lei Gong^{1,*}

¹Key Laboratory of Molecular Epigenetics of the Ministry of Education (MOE), Northeast Normal University, Changchun 130024, China

²Jilin Provincial Key Laboratory of Agricultural Biotechnology, Jilin Academy of Agricultural Sciences, Changchun 130033, China

³Laboratory of Plant Epigenetics and Evolution, School of Life Science, Liaoning University, Shenyang 110036, China

⁴Engineering Research Center of the Ministry of Education (MOE) for Edible and Medicinal Fungi, Jilin Agricultural University, Changchun, Jilin 130118, China

[†]These authors contributed equally to this work.

*Corresponding authors: Key Laboratory of Molecular Epigenetics of the Ministry of Education (MOE), Northeast Normal University, Changchun 130024, China. gongl100@nenu.edu.cn (L.G.); Laboratory of Plant Epigenetics and Evolution, School of Life Science, Liaoning University, Shenyang 110036, China. hongyan2003@126.com (H.W.)

Abstract

Plant long non-coding RNAs (lncRNAs) function in diverse biological processes, and lncRNA expression is under epigenetic regulation, including by cytosine DNA methylation. However, it remains unclear whether 5-methylcytosine (^{5m}C) plays a similar role in different sequence contexts (CG, CHG, and CHH). In this study, we characterized and compared the profiles of genome-wide lncRNA profiles (including long intergenic non-coding RNAs [lincRNAs] and long noncoding natural antisense transcripts [lncNATs]) of a null mutant of the rice DNA methyltransferase 1, *OsMET1-2* (designated *OsMET1-2*^{-/-}) and its isogenic wild type (*OsMET1-2*^{+/+}). The En/Spm transposable element (TE) family, which was heavily methylated in *OsMET1-2*^{+/+}, was transcriptionally de-repressed in *OsMET1-2*^{-/-} due to genome-wide erasure of CG methylation, and this led to abundant production of specific lncRNAs. In addition, RdDM-mediated CHH hypermethylation was increased in the 5'-upstream genomic regions of lncRNAs in *OsMET1-2*^{-/-}. The positive correlation between the expression of lincRNAs and that of their proximal protein-coding genes was also analyzed. Our study shows that CG methylation negatively regulates the TE-related expression of lncRNA and demonstrates that CHH methylation is also involved in the regulation of lncRNA expression.

Keywords: long non-coding RNAs (lncRNAs); DNA methylation; transposable element; *OsMET1-2*; small interference RNA (siRNA); RNA-directed DNA methylation (RdDM)

Introduction

Long non-coding RNAs (lncRNAs) are mRNA-like long RNA transcripts (usually >200 nt in length) that do not encode proteins because they lack discernible open-reading frames (Zhu and Wang 2012; Quinn and Chang 2016; Kopp and Mendell 2018). lncRNAs are expressed across diverse plant and animal species and are involved in the regulation of various biological processes, such as reproduction (Lee and Bartolomei 2013; Zhang et al. 2014), nutrient absorption (Franco-Zorrilla et al. 2007), and response to stimuli (Bhan et al. 2017; Qin et al. 2017). With the development of high-throughput sequencing technologies, many lncRNA transcripts have been identified in different species by transcriptome reassembly (Liu et al. 2012; Wang et al. 2015; Kyriakou et al. 2016; Uszyczynska-Ratajczak et al. 2018; Akay et al. 2019). lncRNAs can be classified into long intergenic non-coding RNAs (lincRNAs) and long noncoding natural antisense transcripts (lncNATs) according to their genomic locations and transcriptional direction relative to the closest neighboring protein-coding genes (PCgenes) (Derrien et al. 2012).

Following the advancing steps of lncRNA identification and characterization in animal models (Wang et al. 2004; Bakhtiarzadeh et al. 2016; Scott et al. 2017; Wang et al. 2017b), many studies have explored tissue lncRNA in different plant species, including representative angiosperms and gymnosperms (Liu et al. 2012; Wang et al. 2014; Zhang et al. 2014; Wang et al. 2015; Lu et al. 2016; Jain et al. 2017; Wang et al. 2017a; Deng et al. 2018; Huang et al. 2018; Wang et al. 2018; Xu et al. 2018; Yuan et al. 2018; Zhang et al. 2018; Zhao et al. 2018a; Deng et al. 2019; Hou et al. 2019; Jiang et al. 2019; Zheng et al. 2019). The features of lncRNAs in these plant species have been extensively characterized in terms of their biogenesis, intrinsic regulation, responses to stresses, regulation of PCgene expression, and involvement in speciation. Plant lncRNAs are typically transcribed by RNA polymerase II, which is similar to that characterized in animal species; additionally, lncRNAs can also be transcribed by plant-specific RNA polymerase V (Wierzbicki et al. 2008). In terms of intrinsic regulation, most lncRNAs exhibit lower expression levels and strong tissue-specific

Received: October 15, 2020. Accepted: February 08, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

expression patterns relative to PCgenes (Liu et al. 2012; Wang et al. 2015). It is also recognized that whole-genome expression of plant lncRNAs is responsive to multiple stress conditions (Wang et al. 2014; Lu et al. 2016; Deng et al. 2018; Yuan et al. 2018) and specific lncRNAs function as novel positive regulators of plants response to different abiotic and biotic stresses (Jain et al. 2017; Qin et al. 2017; Wang et al. 2017a; Zhang et al. 2018). Another special type of stress, the genomic shock that results from genome merger and doubling in allopolyploid plant species, also induces changes in the lncRNA expression profile (Zhao et al. 2018a). Another intriguing dimension involves the regulation by lncRNAs of their PCgene expression (Huang et al. 2018; Xu et al. 2018). Finally, from an evolutionary viewpoint, lncRNA profiles phylogenetically related species suggest that abundant genome-specific and/or lineage-specific lncRNAs show weak evolutionary conservation throughout plant speciation (Liu et al. 2012; Zhao et al. 2018a; Zheng et al. 2019).

The close association between transposable elements (TEs) and lncRNA expression has inspired a number of investigations into the regulation of lncRNA expression by DNA methylation (Wang et al. 2015; Yan et al. 2018; Chen et al. 2019). Most of these studies have characterized the DNA methylation (in CG, CHG, and CHH contexts) around genomic regions that generate lncRNAs and have reached a consistent conclusion: CG and CHG methylation tends to be negatively correlated with lncRNA expression (Wang et al. 2015; Xu et al. 2018; Yan et al. 2018). Notably, because no detailed analysis of DNA methylation mutants were involved, these previous studies are based on correlation analyses only and therefore do not reveal a causal relationship. In addition, although the loss function of DDM1 (decrease in DNA methylation 1, required for CG and CHG methylation of heterochromatic regions) was used to probe the effects of methylation on the expression of transcripts in some plant species (Corem et al. 2018; Tan et al. 2018; Long et al. 2019); this approach could not distinguish the specific effect of CG methylation from that of CHG methylation on lncRNA expression. Overall, the question of whether and how contextual methylation (i.e., CG, CHG, and CHH) affects lncRNA expression remains unanswered.

In this study, we characterized and compared genome-wide lncRNA profiles between a rice loss-of-function mutant for DNA methyltransferase 1, *OsMET1-2* (*OsMET1-2^{-/-}*), and its isogenic wild type (*OsMET1-2^{+/+}*). We show that genome-wide CG hypomethylation in *OsMET1-2^{-/-}* (Hu et al. 2014) leads to massive generation of specific lincRNAs and lncNATs. We demonstrate that these novel lincRNAs and lncNATs derive primarily from hypomethylated En/Spm TEs that are heavily methylated in the wild type. We also find that RNA-directed DNA methylation (RdDM)-mediated CHH hypermethylation in the 5'-upstream genomic regions of lincRNAs is associated with their elevated transcription in *OsMET1-2^{-/-}*. Using paired samples of *OsMET1-2^{-/-}* and *OsMET1-2^{+/+}*, we consistently show that the expression of cis-acting lincRNAs is positively correlated with that of their paired PCgenes in rice.

Materials and methods

Plant materials

The homozygous null mutant of *OsMET1-2* (*OsMET1-2^{-/-}*) and its isogenic wild type (*OsMET1-2^{+/+}*) of *Oryza sativa* L. ssp. *japonica* cv. Nipponbare (Hu et al. 2014) were used in this study. *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* seeds were germinated and grown on plates with Murashige and Skoog (MS) medium in a plant incubator under controlled conditions of 24°C/16 h light and 20°C/8 h dark.

Three biological replicates of each genotype, each consisting of five pooled 11-day-old seedlings, were collected and prepared for RNA isolation.

Library construction and next-generation sequencing

Total RNA was isolated from each biological replicate following standard procedures using the TRIzol reagent (Invitrogen). High-quality RNA was used for the subsequent library constructions. Strand-specific whole transcriptome sequencing (containing both coding and non-coding RNAs) and small RNA sequencing libraries were constructed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, USA) and the NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB, USA.). The resulting libraries were sequenced on the Illumina HiSeq 2500 platform in paired-end 150bp and single-end 50bp mode, respectively, at the Novogene Company in Beijing.

Identification of lncRNAs and their adjacent PCgenes

Low-quality raw sequencing reads were filtered out, and contaminating adaptors within the reads were trimmed, thereby producing clean reads for mapping to the rice reference genome (MSU7.0; <http://rice.plantbiology.msu.edu/>) with HISAT2 (version 2.1.0; no mismatches allowed) (Kim et al. 2015). The transcriptome was assembled and transcripts were quantified by StringTie (version 1.3.4d) (Pertea et al. 2015). GffCompare (version 0.11.4, <http://ccb.jhu.edu/software/stringtie/gffcompare.shtml>) was used to compare the assembled transcripts to the rice annotation profiles and generate a classification code for each transcript, including "i/u/x" coded transcripts (Zhao et al. 2018b). Based on previous definitions and characterizations of lncRNAs (Derrien et al. 2012), transcripts that originated from existing genes were removed, although they were retained if they were located on the opposite strand. In addition, transcripts <200nt in length, transcripts expressed in only one replicate of each genotype, and transcripts with TPM (Transcripts Per Million as calculated by StringTie) <1 were also removed. After these initial filtering steps, blastx was used to evaluate the similarity of candidate transcripts to annotated proteins in rice genome (abbreviated as rice-proteins) and the uniref90 (<https://www.uniprot.org/help/uniref/>) protein database (*e*-value <0.001). Furthermore, minimap2 (with default parameter) and TransDecoder (*e*-value <0.001) were used to scan the Rfam (<http://rfam.xfam.org/>) and Pfam (<http://pfam.xfam.org/>) databases. Candidate transcripts with matches in the aforementioned databases were excluded. In addition, the potential coding ability of novel transcripts was estimated using the CPC2 (<http://cpc2.cbi.pku.edu.cn/>) and CNCI programs (Sun et al. 2013), and novel transcripts with potential coding ability were also removed. A final list of candidate lncRNA transcripts identified from each genotype with their originating genomic locations was used for further analyses. Based on their genomic locations, lncRNAs were further classified into lincRNAs and lncNATs. lncRNA located completely within intergenic regions of the rice genome and that did not intersect with PCgenes were defined as lincRNAs. By contrast, lncRNAs situated on the opposite strand from protein coding genes and that intersected with PCgenes by more than one nucleotide were defined as lncNATs.

For each lincRNA, the closest PCgene within ± 5 kb of its genomic position was defined as its paired PCgene. For each lncNAT, the PCgene on the opposite strand with which it intersected by at least one base was defined as its paired PCgene.

Experimental validation of lincRNA

Twenty lincRNAs randomly selected from mutant-specific lincRNAs and lincNATs and from common lincRNAs and lincNATs were validated by reverse transcription polymerase chain reaction (RT-PCR) followed by Sanger sequencing. In brief, reverse transcription of total RNAs extracted from each genotype was performed using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech). Primer pairs were designed to specifically amplify the reverse-transcribed cDNA of the target lincRNAs using Primer Premier 5 software (Lalitha 2000) (Supplementary Table S1). After amplification and electrophoresis, the PCR products were collected, cloned, and sequenced by Sangon Biotech (Shanghai, China).

To verify the differential expression of selected lincRNAs (DElincRNA; see details in the following sections) in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}*, quantitative real-time PCR (qRT-PCR) was performed on 20 randomly selected common DElincRNAs: 10 lincRNAs and 10 lincNATs. All qRT-PCR primers were designed using an Integrated DNA Technologies online tool (<https://sg.idtdna.com/scitools/Applications/RealTimePCR/>; Supplementary Table S2). Reverse-transcribed cDNA from each biological replicate of each genotype was used as a template for individual qRT-PCR amplification to quantify the lincRNA expression level. The $2^{-\Delta\Delta C_t}$ method was used to estimate relative expression, and ACTIN was used as the internal control gene.

Differential expression of PCgenes and lincRNAs

To identify differentially expressed lincRNAs (DElincRNAs) and PCgenes (DEPCgenes) in *OsMET1-2^{+/+}* vs *OsMET1-2^{-/-}*, DESeq2 (Love et al. 2014) was used to calculate their normalized expression values in RPKM (reads per kilobase per million mapped reads) and assessed their differential expression based on raw reads counts. DElincRNAs and DEPCgenes were defined based on a twofold expression difference between the genotypes and a false discovery rate-adjusted $P < 0.05$.

Small RNA data analysis

Raw small RNA sequencing data (merged from three biological replicates per genotype) were filtered by removing adaptor contamination and low-quality reads. Reads derived from rRNA, tRNA, and were removed using SILVA (<https://www.arb-silva.de/>), GtRNAdb (<http://gtndb.ucsc.edu/>), Rfam, and snoPY (<http://snoopy.med.miyazaki-u.ac.jp/>). All potential miRNA reads were identified using miRDeep-P prediction tool (Yang and Li 2011) and by blastn searches against known pre-miRNAs in the miRbase (version 22.1) (Kozomara et al. 2019). After removing potential miRNA reads, the remaining small interference RNAs (siRNAs) reads were used as input for subsequent analyses. All siRNAs were mapped to the rice reference genome (MSU7.0) using Bowtie1 (Langmead et al. 2009). To compare the siRNA abundance in *OsMET1-2^{+/+}* vs *OsMET1-2^{-/-}*, the counts of mapped 21–24 nt siRNAs from each genotype were normalized into RPM values (reads per million base pair).

Analysis of whole genome bisulfite sequencing data

Whole genome bisulfite sequencing (WGBS) data from *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* were published previously and have been deposited at NCBI under the accession no. SRP043447 (Hu et al. 2014). We estimated context-specific DNA methylation profiles and differentially methylated regions (DMRs) as described in our previous studies (Hu et al. 2014, 2020). The weighted mean DNA

methylation levels in CG, CHG, and CHH contexts within and around genomic regions that contained expressed lincRNAs, PCgenes, and TEs were calculated.

Anchoring paralogs of DEPCgenes

Paralogous gene duplicates in the rice genome were downloaded from the Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/>). The PCgenes of DElincRNAs, which had no paralogous duplicates, were discarded. Only the PCgenes with paralogous duplicates being not any neighbor PCgenes of any other lincRNAs were retained.

Results

Genome-wide identification and characterization of lincRNAs in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}*

Strand-specific RNA-sequencing and a stringent prediction pipeline were used to identify the long non-coding RNA (lincRNA) in a homozygous mutant of *OsMET1-2* (*OsMET1-2^{-/-}*) and its isogenic wild type (*OsMET1-2^{+/+}*). After the removal of the low-quality raw reads, 378 and 380 million paired-end reads were obtained for *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* and were used as input for the prediction pipeline (Figure 1A). In brief, 81,842 and 139,425 transcripts were obtained from *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* using HISAT2 and StringTie as mapping and assembly tools, respectively. Following the removal of unqualified transcripts similar to annotated genic transcripts, transcripts of unexpectedly short length, and transcripts with very low expression, 38,611 and 38,795 transcripts remained in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}*. To ensure the non-coding features of the identified lincRNAs, a final filtration step was performed to exclude transcripts with known and predicted coding potential. Finally, 932 and 1104 lincRNAs were identified in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* (Figure 1A; Supplementary File S1).

The genomic locations and transcription directions of the lincRNAs relative to their nearest neighboring PCgenes were determined, and the lincRNAs were then categorized into long intergenic non-coding RNA (lincRNA) and long non-coding natural antisense transcript (lincNAT). As shown in the Venn diagrams (Figure 1B), the 932 lincRNAs in *OsMET1-2^{+/+}* consisted of 729 lincRNAs and 203 lincNATs, and the 1104 lincRNAs of *OsMET1-2^{-/-}* consisted of 880 lincRNAs and 224 lincNATs. Most lincRNAs and lincNATs were shared by the two genotypes (719 common lincRNAs and 201 common lincNATs; Figure 1B). However, there were a limited number of genotype-specific lincRNAs and lincNATs (10 and 2 wild type-specific lincRNAs and lincNATs; 23 mutant-specific lincNATs; Figure 1B). An exceptionally large number (161) of mutant-specific lincRNAs were identified (Figure 1B). Compared with their respective PCgenes, both types of lincRNAs usually contained fewer exons (most consisted of single exon; Figure 1C), produced shorter transcripts (Figure 1D), and had lower expression levels (Figure 1E). These lincRNA characteristics are consistent with those reported in other plant species (Li et al. 2014; Wang et al. 2015; Xu et al. 2018).

RT-PCR and qRT-PCR analyses confirmed the existence of randomly selected lincRNAs and validated their relative expression levels, further verifying the accuracy of our lincRNA predictions (Supplementary Figures S1 and S2).

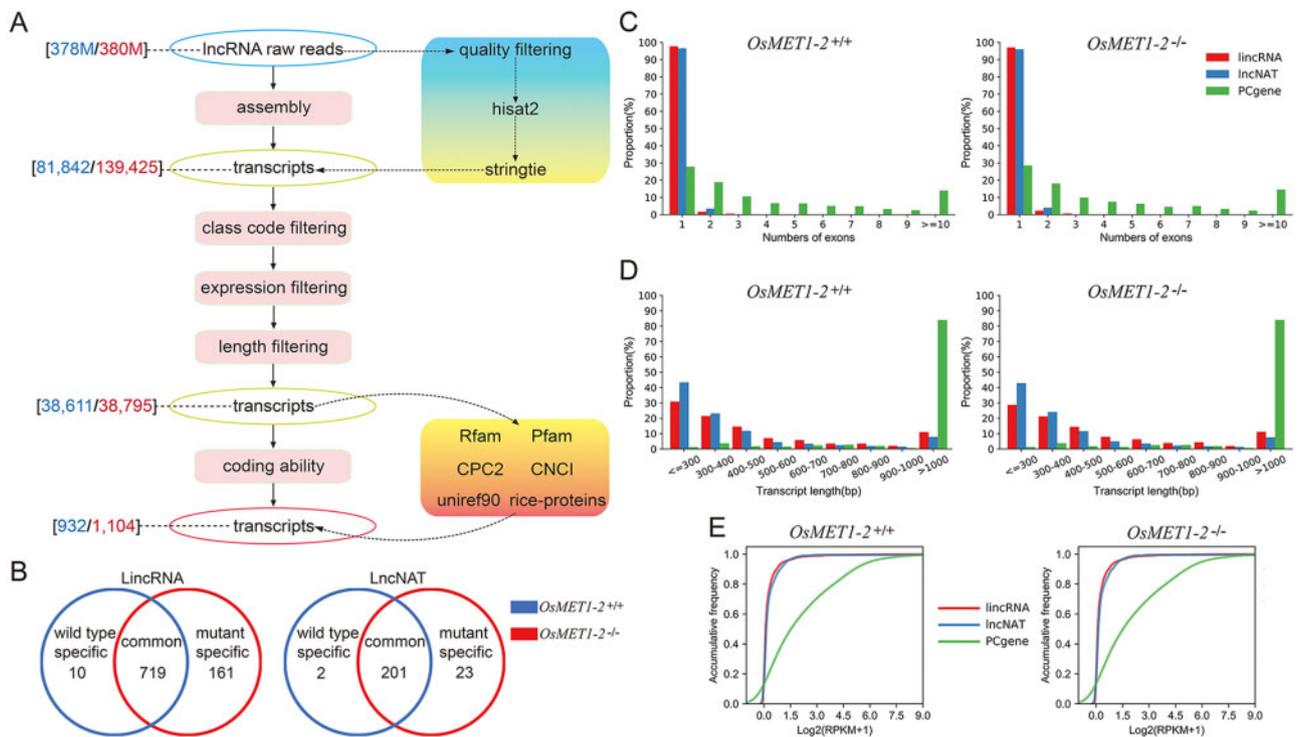


Figure 1 Identification and characterization of long non-coding RNA (lncRNAs) in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}*. (A) The workflow of lncRNA identification pipeline developed in this study. The parenthesized numbers in blue and red denote the respective number of reads or transcripts input into the following step. The frames in gradient colors specify the detailed database(s) and/or tools adopted in respective step. (B) The Venn diagrams tabulating the numbers of lincRNA and lincNAT shared (common) in *OsMET1-2^{+/+}* (blue) and *OsMET1-2^{-/-}* (red) and specifically identified in respective sample (wild type and mutant specific). The exact number of lincRNAs in each category is listed beneath respective category name. (C) Proportions of lincRNA transcripts (lincRNAs and lincNATs) and the adjacent PCGenes in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* categorized in terms of the exon numbers. (D) Proportions of lincRNA transcripts (lincRNAs and lincNATs) and the adjacent PCGenes in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* categorized in terms of the transcript length. (E) Cumulative frequency curves of the transcript abundances of lincRNA, lincNAT, and PCGenes. The x-axis tabulates each transcript category with respective log₂FC (fold change) of Reads Per Kilobase per Million mapped reads (RPKM); the y-axis tabulates the accumulative frequency after adding each transcript category.

Genomic regions that generated mutant-specific lncRNAs showed greater hypomethylation than those that generated common lncRNAs

In addition to the large number of common lncRNAs shared between *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}*, sets of lincRNAs (18.30%; 161/880) and lincNATs (10.27%; 23/224) were specifically expressed in *OsMET1-2^{-/-}* (Figure 1B). In our previous study, the loss-of-function mutation of *OsMET1-2* caused genome-wide CG and CHG hypomethylation (Hu et al. 2014). To test for an association between novel lncRNAs expression and CG and CHG hypomethylation, we compared the CG and CHG methylation patterns of genomic regions that expressed novel or common lncRNAs in *OsMET1-2^{-/-}* with their corresponding regions in *OsMET1-2^{+/+}* (Figure 2A; Supplementary Figure S3). As expected, the overall CG and CHG methylation level of lncRNA genomic regions was lower in *OsMET1-2^{-/-}* than in *OsMET1-2^{+/+}* for both common and mutant-specific lncRNAs (Figure 2A; Supplementary Figure S3A). Genomic regions that generated mutant-specific lncRNAs in *OsMET1-2^{-/-}* had higher CG and CHG methylation levels in *OsMET1-2^{+/+}* than regions that generated common lncRNAs (Figure 2A). This difference was confirmed statistically by a random sampling method in which the CG and CHG methylation levels of regions that encoded mutant-specific lncRNAs in *OsMET1-2^{+/+}* were significantly higher than those of randomly sampled regions (Figure 2B; Supplementary Figure S3B). Furthermore, CG and CHG methylation levels of genomic regions that generated mutant-specific lncRNAs were hypomethylated

more than the regions that generated common lncRNAs in *OsMET1-2^{-/-}* (Figure 2A; Supplementary Figure S3A).

To obtain further support, we also calculated the numbers of common and mutant-specific lncRNAs that co-localizing with CG and CHG DMRs in *OsMET1-2^{-/-}* for each type of lncRNAs (Figure 2C; Supplementary Figure S3C). Relative to the number of randomly bootstrap-sampled intergenic and anti-sense genic regions that overlapped with DMRs (i.e., the reference distribution), the mutant-specific lincRNAs and lincNATs occurred in CG DMRs at significantly higher frequencies than expected, but a similar result was not found for common lncRNAs (Figure 2C). However, the result for CHG DMRs was more complicated: both mutant-specific and common lincRNAs were statistically enriched in CHG DMRs (Supplementary Figure S3C), but mutant-specific lincNATs were not. These observations suggest a potential association between novel lncRNA expression and CG hypomethylation. Nonetheless, there was a lack of statistical evidences to support an association between novel lncRNAs expression and CHG hypomethylation in this study.

TE-derived lncRNAs were de-repressed in *OsMET1-2^{-/-}*

Genomic features that generated lincRNAs and lincNATs in both *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* were further characterized. First, the two types of lncRNAs were categorized into four groups based on their locations in genic/intergenic regions with/without TEs (Supplementary Figure S4; the lack of coding ability of

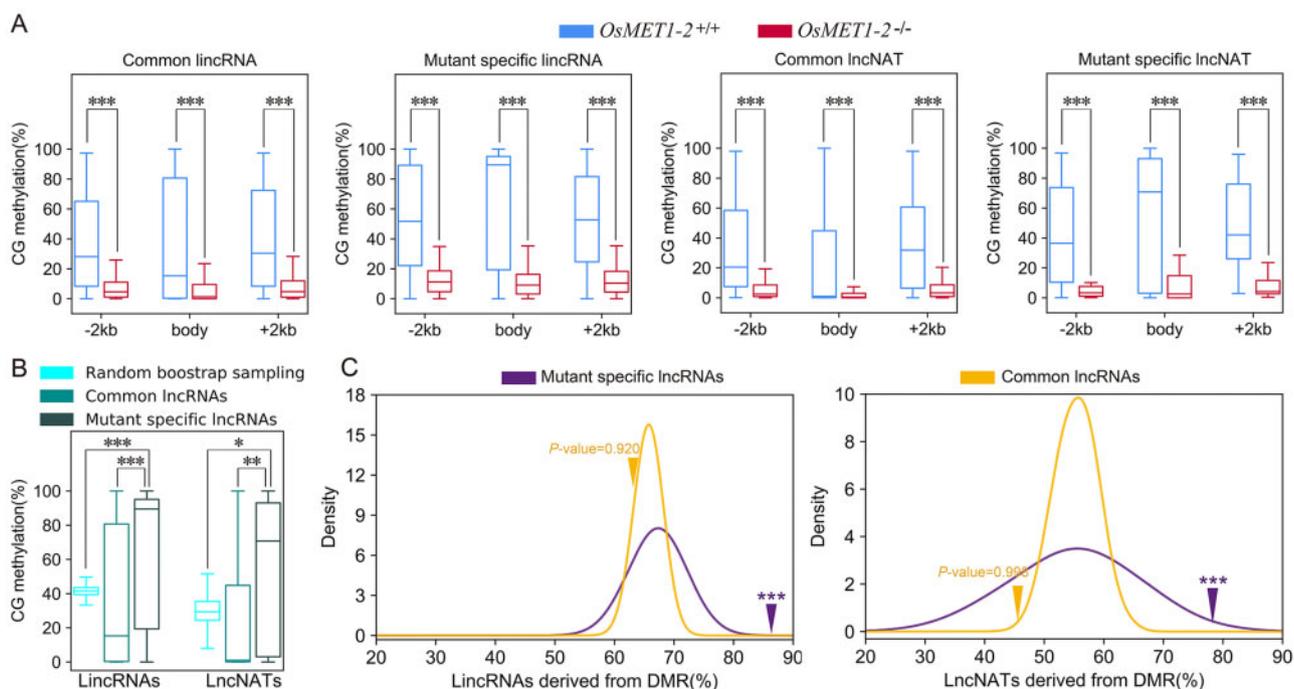


Figure 2 Genomic regions of CG hypomethylation in *OsMET1-2*^{+/+} expressing mutant-specific lincRNAs after null-mutation of *OsMET1-2* gene. (A) The boxplots depict the CG methylation levels of genomic regions (core body and up-/downstream 2 kb flanking regions) expressing common and mutant-specific lincRNAs (including lincRNA and lincNAT) in respective *OsMET1-2*^{+/+} and *OsMET1-2*^{-/-}. Wilcoxon test is adopted to test the statistical significance for paired two sample sets. One asterisk (*), two asterisks (**), and three asterisks (***) denote the significant P-values at the levels of 0.05, 0.01, and 0.001, respectively. (B) Boxplots of weighted mean CG methylation levels of random bootstrap sampled genomic regions and genomic regions expressing common and mutant-specific lincRNAs (lincRNAs and lincNATs) in *OsMET1-2*^{+/+}. Independent two-sample t-test is used, in which significance levels are also denoted at the same cutoff P-values as above. (C) Density curves of the percentages of random bootstrap sampled intergenic (left) and anti-sense genic regions (right) overlapping with DMRs and arrow-marked observed percentage of common and mutant-specific lincRNAs (lincRNAs and lincNATs) derived from the DMRs. Within respective bootstrapping test, we randomly re-sample 1000 sets of genomic regions, the number and length of which are identical with respective lincRNAs (lincRNAs and lincNATs). Within each re-sampled set of genomic regions, the proportion of regions overlapping with DMRs is calculated. Respective 1000 proportions are summarized in each density curve. The original observed proportion of lincRNA occurred in DMRs is denoted by the arrow and respective statistical P-value for each bootstrapping test is also specified nearby each arrow.

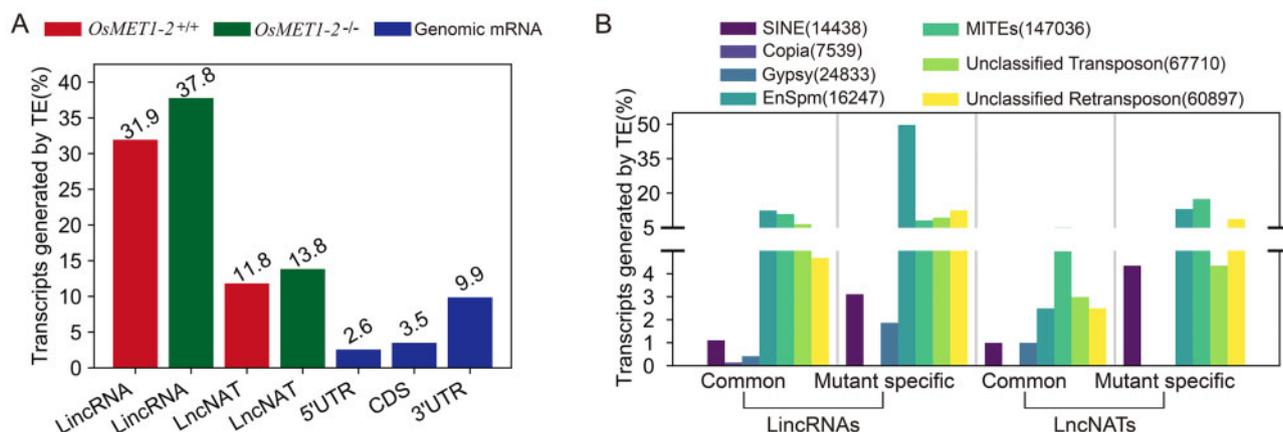


Figure 3 LincRNA and mRNA transcripts generated by TEs in *OsMET1-2*^{+/+} and *OsMET1-2*^{-/-}. (A) Proportions of lincRNA transcripts (lincRNA and lincNAT) and genomic mRNA with at least one exon overlapping with TEs (at least 10 bp). (B) Proportions of common and mutant-specific lincRNAs (lincRNAs and lincNATs) overlapping with respective type of TEs (at least 10 bp). The parenthesized number denotes the total number of respective TE type in the genome.

autonomous TE-related lincRNAs was confirmed by checking their incomplete ORFs; see Materials and methods section). Relative to mRNA regions (separated into 5' UTR, CDS, and 3' UTR), significantly more lincRNAs (especially lincRNAs) were generated by genomic regions associated with TEs (genic and intergenic TEs) in both *OsMET1-2*^{+/+} and *OsMET1-2*^{-/-} (Figure 3A).

Next, the proportions of common and mutant-specific lincRNAs expressed by specific TE types were summarized (Figure 3B). Overall, more lincRNAs were generated by Type II transposons (DNA transposons) than by Type I transposons (retro-transposons) for both common and mutant-specific lincRNAs (Figure 3B). In addition, mutant-specific lincRNAs and lincNATs

Table 1 The weighted mean cytosine DNA methylation levels of protein coding genes, TE-related genes, all TE types, and each specific type of TEs in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}*

Category	CG			CHG			CHH		
	<i>OsMET1-2^{+/+}</i> (%)	<i>OsMET1-2^{-/-}</i> (%)	Decreased (%)	<i>OsMET1-2^{+/+}</i> (%)	<i>OsMET1-2^{-/-}</i> (%)	Decreased (%)	<i>OsMET1-2^{+/+}</i> (%)	<i>OsMET1-2^{-/-}</i> (%)	Decreased (%)
Protein coding genes	26.40	3.30	-87.60	9.20	6.70	-26.50	2.10	0.70	-65.50
TE-related genes	85.40	18.40	-78.40	64.70	51.20	-20.90	4.90	2.20	-53.90
Total repeats	83.60	18.30	-78.10	54.10	43.80	-19.10	26.30	10.70	-59.50
Retrotransposons (Class I/retro TE)	89.60	21.10	-76.50	65.00	51.00	-21.50	10.20	6.10	-40.60
Copia	87.50	19.30	-78.00	61.00	47.00	-23.00	7.60	4.90	-36.10
Gypsy	90.80	21.70	-76.10	68.70	53.70	-21.80	7.00	5.50	-20.60
LTR-other	84.90	23.30	-72.50	59.30	48.80	-17.70	10.90	9.50	-12.90
Cassandra	94.40	28.90	-69.40	74.20	60.10	-19.00	20.50	13.10	-36.20
Caulimovirus	94.60	24.90	-73.70	81.70	73.90	-9.60	3.30	4.30	29.10
LINE	82.70	17.20	-79.20	61.20	55.30	-9.70	4.60	2.50	-45.60
SINE	87.30	18.90	-78.40	54.60	42.90	-21.40	23.90	7.90	-66.90
Transposons (Class II/DNA TE)	78.20	16.60	-78.80	47.20	38.40	-18.50	22.00	9.20	-58.30
En/Spm	90.50	19.00	-79.00	54.10	37.00	-31.50	10.00	10.20	2.70
MITEs	83.40	18.00	-78.50	52.90	43.20	-18.30	33.80	12.90	-61.80
hAT	79.60	14.50	-81.80	37.50	23.10	-38.30	12.30	5.20	-57.80
Harbinger	80.90	17.50	-78.40	53.10	46.30	-12.90	30.20	12.80	-57.70
Stowaway	77.20	17.20	-77.70	45.70	37.40	-18.10	25.40	9.20	-63.50
Tourist	79.40	18.40	-76.80	50.30	44.60	-11.40	24.50	9.50	-61.10
MuDR	87.50	21.10	-75.90	53.50	47.10	-12.00	16.30	6.10	-62.80
DNA-other	59.40	10.10	-83.00	34.60	29.40	-14.90	14.40	5.90	-59.30

Within each category, the proportion of reduction in DNA methylation level (in CG, CHG, and CHH context) in *OsMET1-2^{-/-}* relative to respective level in the *OsMET1-2^{+/+}* is recorded as "Increase or Decreased (%)," which is calculated as $(\text{OsMET1-2}^{-/-} - \text{OsMET1-2}^{+/+}) / \text{OsMET1-2}^{+/+}$.

were more highly expressed than common lincRNAs (Figure 3B). Notably, 49.69% of the mutant-specific lincRNAs were generated by the En/Spm DNA transposon family, significantly higher than the corresponding percentage of common lincRNA (12.40%) (Chi-square test, $P < 0.001$) (Figure 3B). Although miniature inverse-repeated TEs (MITEs) were the most abundant TE types in the rice genome (Figure 3B), MITEs did not generate significantly more mutant-specific lincRNAs than common lincRNAs (Chi-square test, $P = 0.09$). Furthermore, detailed characterization of DNA methylation (in CG, CHG, and CHH contexts) of all TE types in *OsMET1-2^{+/+}* revealed that En/Spm harbored higher CG methylation levels than other DNA TEs (Table 1). Taken together, these results imply that CG-methylated TE types (e.g., En/Spm) may be more likely to be de-repressed and to express lincRNAs in the *OsMET1-2^{-/-}*.

RdDM-mediated CHH hypermethylation in the 5'-upstream regions of transcriptionally upregulated lincRNAs in *OsMET1-2^{-/-}*

To examine the link between DNA methylation and lincRNA expression in different contexts (i.e., CG, CHG, and CHH), genomic regions that contained differentially expressed lincRNA (DELincRNA [lincRNA and lincNAT]) transcripts and their ± 2 kb upstream and downstream regulatory regions were examined in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}* (Figure 4A; Supplementary Figures S5 and S6). Genomic regions with statistically significantly upregulated and downregulated lincRNAs and with common and mutant-specific lincRNAs were considered separately (Figure 4A; Supplementary Figures S5 and S6).

For DNA methylation in CG and CHG contexts, all lincRNA-related genomic regions were consistently hypomethylated in *OsMET1-2^{-/-}*, and there were no region-specific DNA methylation changes (Supplementary Figures S5 and S6). The genomic regions that generated lincRNAs exhibited CHH hypomethylation in

OsMET1-2^{-/-} (Figure 4A). Specifically, CHH hypomethylation occurred in genomic regions that generated downregulated common lincRNAs and lincNATs (Figure 4A; Supplementary Figure S6). By contrast, in genomic regions that generated upregulated common and mutant-specific lincRNAs, CHH sites were hypermethylated in the 5'-upstream regulatory regions (~250 bp) adjacent to transcription starting sites in *OsMET1-2^{-/-}* (Figure 4A). This phenomenon was not observed in regions that generated lincNATs (Supplementary Figure S6). Considering the important role of siRNAs in the establishment of CHH methylation by the RdDM pathway (Matzke and Mosher 2014), we sought to test whether these CHH hypermethylated 5'-upstream regions were targeted by siRNAs. As expected, our small RNA sequencing and mapping results revealed significantly abundant siRNAs that co-localized with the special hypermethylated regions associated with upregulated common and mutant-specific lincRNAs (Figure 4A). Among the mutant-specific lincRNAs, 61.49% (99/161) displayed CHH hypermethylation in their 5'-upstream region, 74.53% (120/161) harbored enriched siRNAs in their 5'-upstream region, and 52.17% (84/161) exhibited concomitant CHH hypermethylation and abundant siRNAs in their 5'-upstream regions. However, such high proportions were not observed for non-differentially expressed lincRNAs (hyper^mCHH 32.99%, 193/585; abundant siRNAs 39.15%, 229/585; concomitant hyper^mCHH and abundant siRNAs 14.19%, 83/585).

Given our previous findings of En/Spm enrichment in mutant-specific lincRNAs (Figure 3B), we also characterized the weighted mean CHH methylation levels of En/Spm genomic regions that expressed upregulated common and mutant-specific lincRNAs transcripts. Concomitant CHH hypermethylation and siRNA abundance was once again observed 5'-upstream of En/Spm genomic regions that expressed upregulated common and mutant-specific lincRNAs (Figure 4B). This observation was also supported by compensatory CHH methylation that occurred

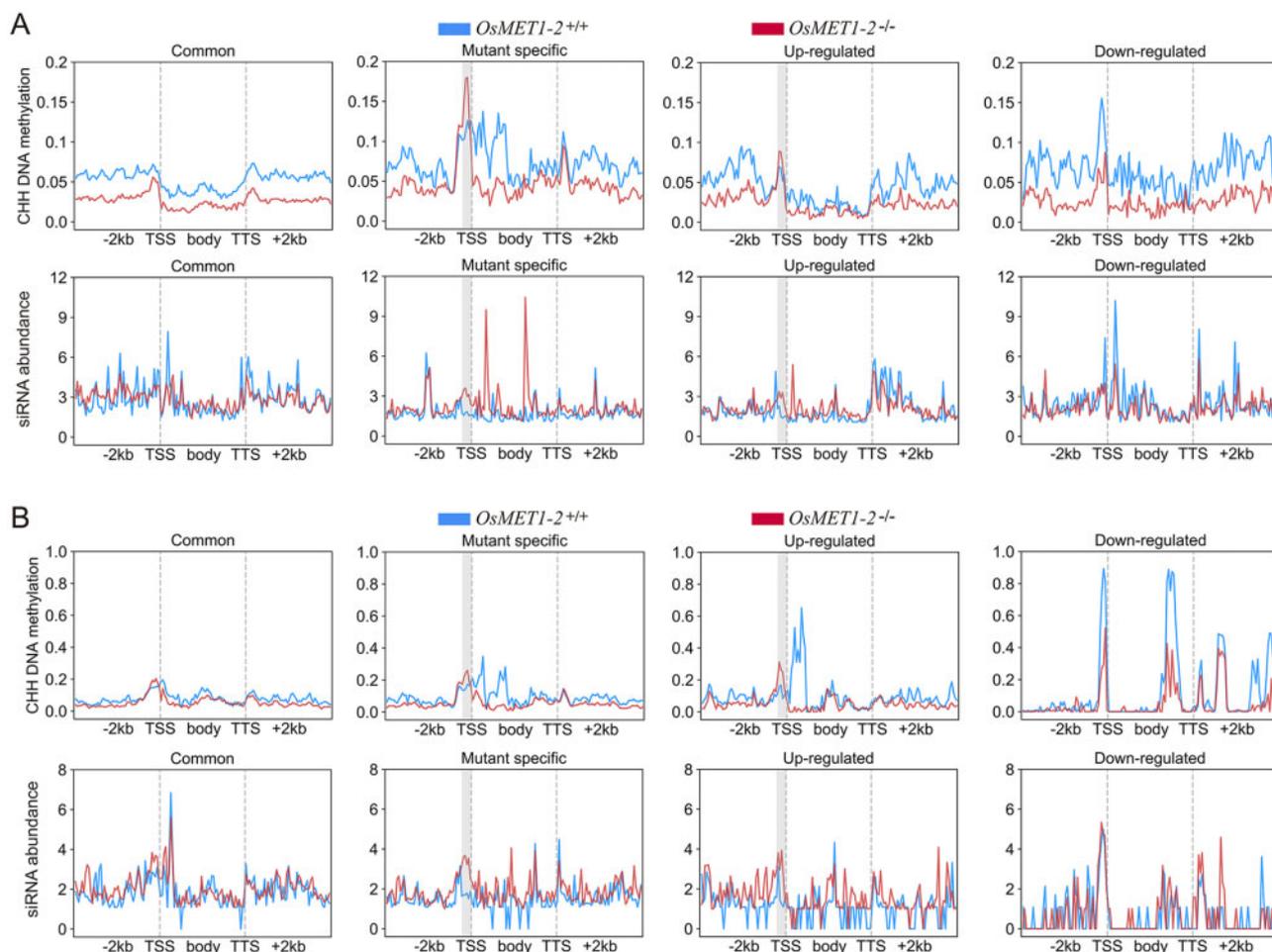


Figure 4 Weighted mean CHH DNA methylation and siRNA abundance (Log2 transformed) of genomic regions (lincRNA bodies and their up-/downstream [+2kb] regulative regions) expressing common, mutant-specific, and differentially up- and downregulated lincRNA in *OsMET1-2^{+/+}* and *OsMET1-2^{-/-}*. (A) Weighted mean CHH DNA methylation and siRNA abundance of genomic regions expressing respective featured lincRNAs. (B) Weighted mean CHH DNA methylation and siRNA abundance of genomic regions expressing En/Spm-derived featured lincRNAs. The gray blocks denote the 5'-upstream (~250 bp upstream of transcription starting site) regulative regions with co-localization of hypermethylated CHH and abundant siRNAs.

specifically in En/Spm TEs after the null mutation of the *OsMET1-2* gene (Table 1). All these results indicate that RdDM can produce compensatory CHH methylation within the 5'-upstream regulative genomic regions (especially in the En/Spm TE regions) of transcriptionally upregulated lincRNAs in *OsMET1-2^{-/-}*.

Expression of cis-acting lincRNAs is positively correlated with that of their paired PCgenes

Our earlier study reported extensive differential PCgene expression in *OsMET1-2^{-/-}* relative to *OsMET1-2^{+/+}* (Hu et al. 2014). Based on the DELincRNAs in the same sample set, it was possible to explore potential cis-regulatory effects of lincRNAs on the expression of their neighboring PCgenes. Specifically, we characterized the correlation between expression fold changes of DELincRNAs (including both common and mutant-specific lincRNAs and lincNATs) and those of their corresponding differentially expressed of PCgenes (DEPCgenes) (Figure 5). To exclude intrinsic noise effects from other factors (including the adjacent TEs and local differential methylation) that may have mediated an indirect correlation, we categorized the lincRNAs into four subgroups based on their locations relative to genomic TEs and CG DMRs. Subsequently, we calculated Pearson's correlations and

corresponding *P*-values for each subgroup of lincRNAs and lincNATs (Figure 5C). After excluding the effects of adjacent TEs and CG DMRs associated with the null mutation of *OsMET1-2* gene, the fold changes of DELincRNA expression in *OsMET1-2^{-/-}* relative to *OsMET1-2^{+/+}* were significantly correlated with those of their corresponding DEPCgenes ($n = 184$; Pearson's correlation = 0.604, $P < 0.001$; Figure 5, A and C). There was no significant correlation between the fold changes of DELincNAT and those of their corresponding DEPCgenes (Figure 5, B and C). PCgenes paired with DELincRNA are enriched in arabinan/xylan catabolic process and sodium ion transmembrane transport. Both arabinan and xylan are present abundantly in plant cell walls (Verhertbruggen et al. 2009; Grantham et al. 2017). These enrichments indicate that the correlation between lincRNA and PCgene expression may be involved in the abnormal growth of the mutant.

To further verify the potential positive correlation between expression of cis-acting lincRNAs and that of their paired PCgenes, another two groups of PCgenes were selected as negative controls. One included paralogs of the lincRNA-related PCgenes (see Materials and methods section), and the other included randomly selected rice genes. If the expression of lincRNAs was positively correlated with that of their PCgenes,

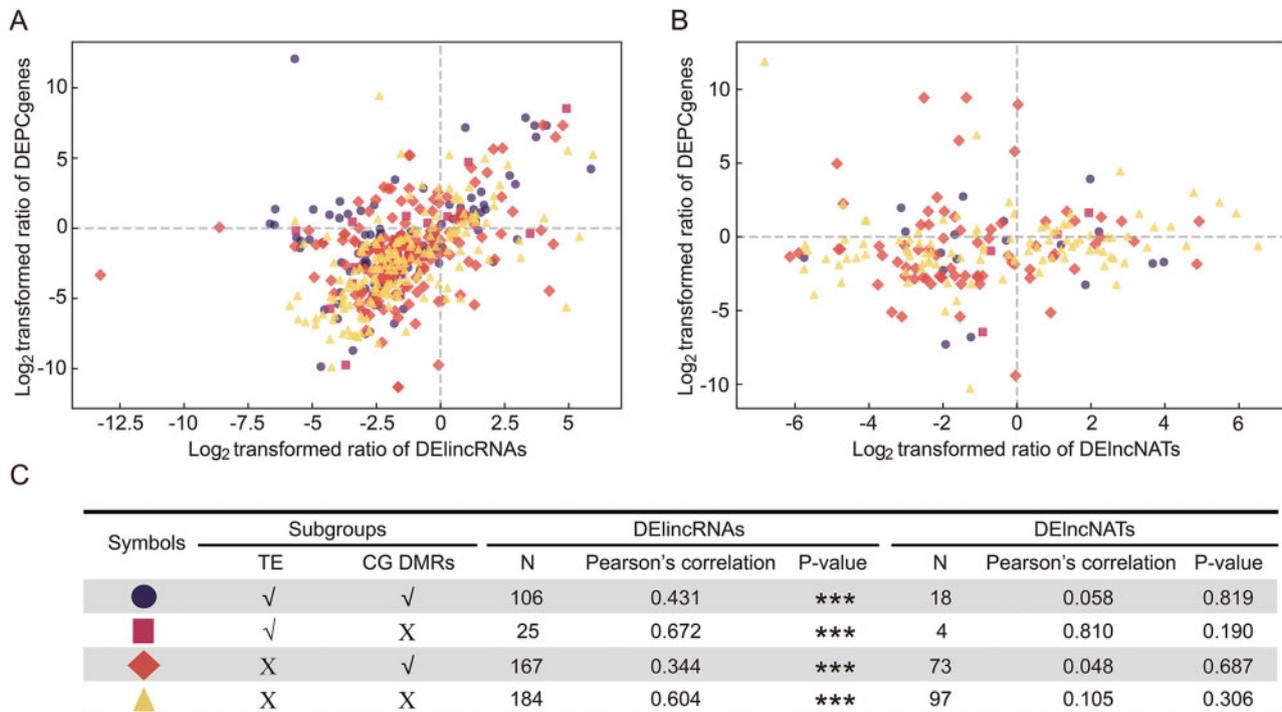


Figure 5 Cis-acting lincRNAs is positively correlated with expression of their neighboring PCgenes. (A) Scatter plot illustrating the positive correlation between the fold changes of DElincRNA (differential expression of lincRNA in *OsMET1-2^{-/-}* vs in *OsMET1-2^{+/+}*; log₂ transformed on the x-axis) and those of respective DEPCgenes (differential expression of lincRNA-related PCgenes in *OsMET1-2^{-/-}* vs in *OsMET1-2^{+/+}*; log₂ transformed on the y-axis). The detailed Pearson's correlation indices and respective statistical significances are tabulated in panel C of this figure. (B) Scatter plot illustrating no correlation between the fold changes of DElincNAT (differential expression of lincNAT in *OsMET1-2^{-/-}* vs in *OsMET1-2^{+/+}*; log₂ transformed on the x-axis) and those of respective DEPCgenes (differential expression of lincNAT-related PCgenes in *OsMET1-2^{-/-}* vs in *OsMET1-2^{+/+}*; log₂ transformed on the y-axis). The detailed Pearson's correlation indices and respective statistical significances are tabulated in panel C of this figure. (C) Different lincRNA and lincNAT subgroups are categorized in terms of their relative positions to TEs and CG DMRs, in which the circles denote the lincRNAs co-localizing with the TEs and CG DMRs; the squares denote the lincRNAs uniquely co-localizing with the TEs; the diamonds denote the lincRNAs uniquely co-localizing with the CG DMRs; and the triangles denote the lincRNAs neither co-localizing with the TEs nor CG DMRs. Pearson's correlation is calculated for paired lincRNA and PCgenes in each subgroup. Three asterisks (***) represent the significant *P*-values at the level of 0.001; and raw non-significant *P*-values (>0.05) are specified.

such a positive correlation should be present between lincRNAs and their PCgenes but absent in the two negative control groups. This hypothesis was tested using the same method described above (Figure 5), and a significant correlation was found only between the cis-acting lincRNAs and their corresponding paired PCgenes (Figure 6).

Discussion

High-throughput sequencing technology has enabled researchers to characterize a large number of lincRNAs from various eukaryotic species (Kyriakou et al. 2016; Wang et al. 2017b; Akay et al. 2019). Major questions about lincRNA composition, biogenesis, tissue-specific expression, function, and association with epigenetic modifications have been explored and mostly answered in plant species (Liu et al. 2012; Wang et al. 2015; Hu et al. 2020). Nonetheless, little evidence exists for participation of context-specific DNA methylation in the regulation of plant lincRNA expression (Wang et al. 2017a; Xu et al. 2018; Chen et al. 2019). We therefore characterized and compared lincRNA expression (lincRNAs and lincNATs) between wild-type rice (*OsMET1-2^{+/+}*) and its homozygous mutant *OsMET1-2^{-/-}*, in which CG methylation has been dramatically reduced by null mutation of the *OsMET1-2* gene (Hu et al. 2014). In addition to clarifying the elusive relationship between CG methylation and lincRNA expression, we also demonstrated the involvement of CHH methylation in the

regulation of lincRNA expression. Notably, compared with the *OsDDM1* mutant that exhibits a simultaneous decrease in CG and CHG methylation (Tan et al. 2018), the limited CHG methylation variation in our rice *OsMET1-2^{-/-}* mutant allows us to specifically exclude any potential mixed effects from CHG methylation in our association analyses.

Use of the wild type *OsMET1-2^{+/+}* and its *OsMET1-2^{-/-}* mutant enabled us to provide strong evidence for the regulation of lincRNA expression by CG methylation: the heavily CG-methylated regions in *OsMET1-2^{+/+}* were induced to express novel mutant-specific lincRNAs in *OsMET1-2^{-/-}* (Figure 2).

Given that the CG methylation level was higher in TE regions than in genic regions (Table 1) (Feng et al. 2010), we hypothesized that the novel mutant-specific lincRNAs may have originated from TE-rich regions. To test this hypothesis, we investigated the composition of genomic regions that generated mutant-specific lincRNAs. A specific group of DNA transposons, the En/Spm DNA transposons, expressed more mutant-specific lincRNAs after the erasure of CG methylation in *OsMET1-2^{-/-}* (Figure 3). Here, it is necessary to emphasize that the role of CHG methylation in the regulation of lincRNA expression is still ambiguous as characterized in the current study system. Future investigation in other mutants with abolished CHG methylation (e.g., the *cmt3* mutant) could provide additional insight.

Another intriguing question arises: why does this specific type of TE promote the active expression of lincRNAs in response to

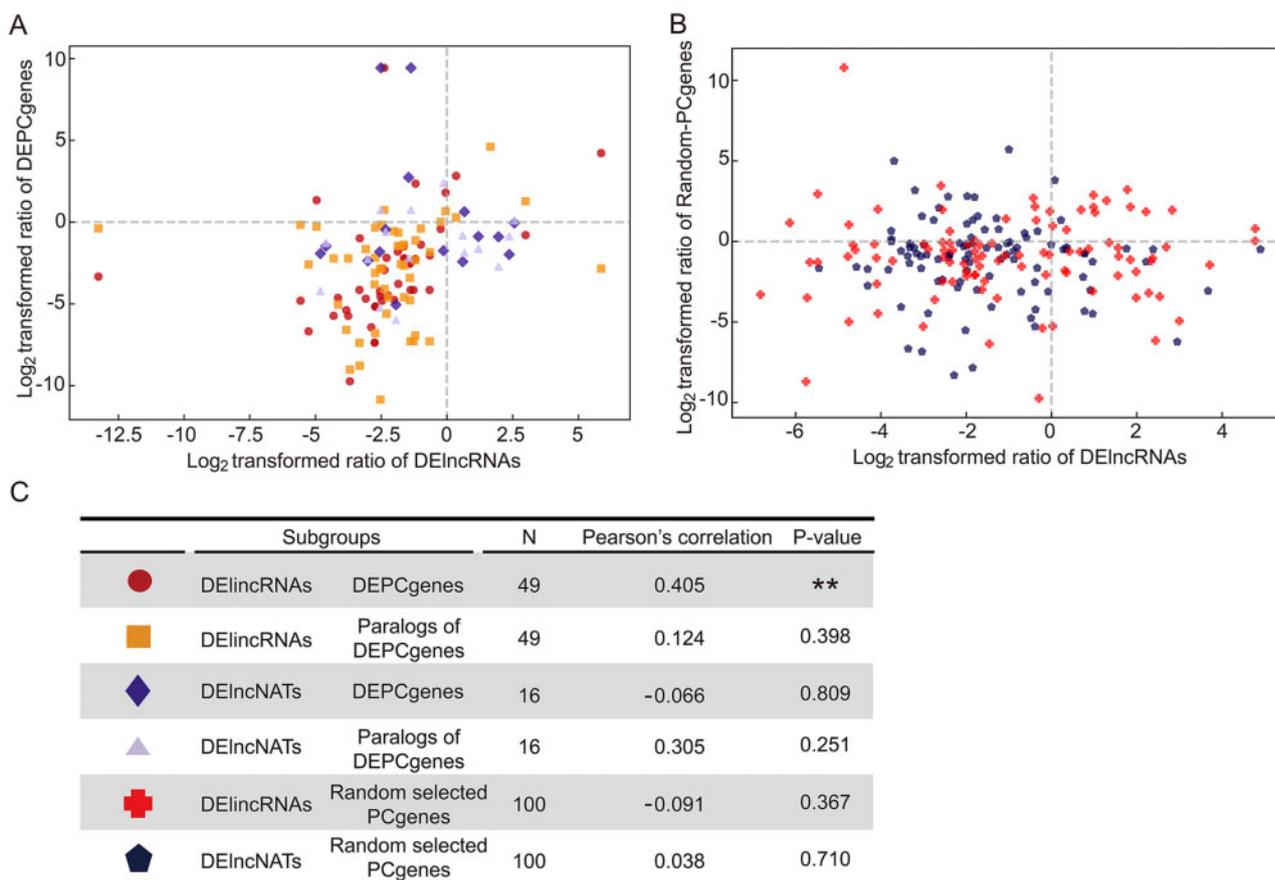


Figure 6 Scatter plot illustrating the unique positive correlation of cis-acting lincRNA with the expression of their neighboring PCgenes rather than respective paralogs of PCgenes and random selected PCgenes for the correlation. (A) Positive correlation between the fold changes of DElincRNA (differential expression of lincRNA in *OsMET1-2^{-/-}* vs in *OsMET1-2^{+/+}*; log₂ transformed on the x-axis) and those paralogs of DEPCgenes and DEPCgene (differential expression of lincRNA-related PCgenes and their paralogs in *OsMET1-2^{-/-}* vs in *OsMET1-2^{+/+}*; log₂ transformed on the y-axis). No corresponding correlation is detected between DElincNAT and their DEPCgenes and respective paralogs of DEPCgene. The detailed Pearson's correlation indices and respective statistical significances are tabulated in panel C of this figure. (B) No significant correlation is detected between the lincRNA and their random selected PCgenes. Detailed Pearson's correlation indices and categories are tabulated in panel C of this figure. (C) Pearson's correlation indices between the fold changes of DElincRNA (differential expression of lincRNAs and lincNATs in *OsMET1-2^{-/-}* vs in *OsMET1-2^{+/+}*) and those of DEPCgenes, paralogs of respective DEPCgenes (differential expression of lincRNA- and lincNAT-related PCgenes and their paralogs in *OsMET1-2^{-/-}* vs in *OsMET1-2^{+/+}*), and random selected respective PCgenes are tabulated with corresponding supporting statistical P-values. Different lincRNA and lincNAT subgroups are categorized in terms of their PCgenes, paralogs of respective DEPCgenes, and random selected PCgenes, in which the circles denote the lincRNAs paired with their respective DEPCgenes; the squares denote the lincRNAs paired with their respective paralogs of DEPCgenes; the diamonds denote the lincNATs paired with their respective DEPCgenes; the triangles denote the lincNATs paired with their respective paralogs of DEPCgenes; the crosses denote the lincRNAs paired with random selected PCgenes; and the pentagons denote the lincNATs with random selected PCgenes. Pearson's correlation is calculated for each subgroup. Two asterisks (**) represent the significant P-values at the level of 0.01; and raw non-significant P-values (>0.05) are specified.

the removal of CG methylation? Given the smaller number of En/Spm transposons relative to those of other TE types in the rice genome (Figure 3B), the contribution of En/Spm transposons to lincRNA transcription does not correlate with their genomic abundance. This suggests that active lincRNA expression by En/Spm transposons must be determined by other intrinsic properties. Although both En/Spm transposons and MITEs are enriched in intergenic regions (Ouyang and Buell 2004), significant mutant-specific lincRNA expression is derived by En/Spm transposons but not by MITEs, implying that a biased distribution within intergenic regions is not the intrinsic factor either. Given the marked decrease in CG methylation in regions expressing mutant-specific En/Spm transposons in *OsMET1-2^{-/-}* (79.00%, Table 1; Figure 2, B and C), greater erasure of CG methylation from En/Spm transposons than from other TE types may be one relevant intrinsic factors. However, SINE retrotransposons exhibited a degree of CG methylation erasure similar to that of En/Spm

transposons (79.20%; Table 1), but they did not express more mutant-specific lincRNAs in *OsMET1-2^{-/-}*. This suggests that other unknown intrinsic features of En/Spm transposons and/or other regulatory process(es) involved in their de-repression must influence mutant-specific lincRNA expression after the null mutation of *OsMET1-2*. In addition to the previously reported colocalization of TEs with expressed lincRNAs in rice and other plant species (Wang et al. 2017a; Yan et al. 2018), this study provides a clear example of the direct negative regulation of lincRNA expression by CG methylation of TEs in a monocot species.

In addition to enriched CG methylation, CHH methylation established by siRNAs through the RdDM pathway is another prominent epigenetic feature of plant intergenic TE regions (Xu et al. 2018; Yan et al. 2018). As previously reported (Hu et al. 2014) and also illustrated in our study (Table 1; Figure 4), a decrease in CHH methylation within the bodies and regulatory regions of most TEs is accompanied by the erasure of CG methylation.

However, an exceptional contrasting case is the compensatory increase in CHH methylation in the En/Spm transposons (2.70%; Table 1). The *prima facie* coincidence of lincRNA expression and compensatory CHH methylation in the same group of En/Spm transposons after null mutation is contradicted by the observed co-occurrence of siRNA enrichment and increased CHH methylation in the 5'-upstream regulatory regions of mutant-specific and upregulated common lincRNA transcripts (Figure 4). Our observations suggest that together with CG methylation, CHH methylation mediated by the RdDM pathway is also involved in regulating lincRNA expression, especially for lincRNAs. However, in contrast to the clear negative effects of CG methylation on lincRNA expression discussed above, the potential role of compensatory CHH methylation remains unclear. According to canonical theory on the silencing effects of CHH methylation on TE transcription (Matzke and Mosher 2014), it is deduced that our observed CHH hypermethylation in lincRNA regulative regions could compensatively silence the TE transcription in the absence of inhibitive CG methylation. Such a prediction is consistent with the previously reported association between 5'-upstream CHH methylation and the expression of downstream neighboring PCgenes in other plant species (Gent et al. 2013; Li et al. 2015; Secco et al. 2015). However, based on the recent recognition of RdDM-mediated CHH methylation as a signal that recruits certain transcriptional anti-silencers (Harris et al. 2018), another possible scenario is that CHH methylation around the intergenic TE regions may counteract the repressive effects of CG methylation on lincRNA expression. Comparisons of lincRNA profiles from additional RdDM rice mutants will be necessary to determine whether intergenic lincRNAs expression increases (supporting the former "collaborative negative model") or decreases (supporting the latter "counteracting active model") when the RdDM pathway is abolished. The exact role of CHH methylation in the regulation of lincRNA expression will then be made clear.

LincRNA has been reported to regulate the expression of both neighboring (*cis*) and distal (*trans*) PCgenes in animal models (Pauli et al. 2012; Casero et al. 2015; Zhu et al. 2015). As in some other plant model species (Huang et al. 2018; Xu et al. 2018), *cis*-acting lincRNAs exhibited positive correlations with their neighboring PCgenes in our rice materials. Given the abnormal phenotypes of *OsMET1-2^{-/-}* (Hu et al. 2014), it will be interesting to construct lincRNA and/or PCgene mutants with which to characterize the specific functions of lincRNAs in the regulation of PCgene expression and to identify their potential roles in underpinning the observed phenotypes. As in other plant studies (Li et al. 2014; Li et al. 2017; Huang et al. 2018; Gao et al. 2020), the potential *trans*-acting functions of lincRNAs in the regulation of gene expression at independent or distant loci were not explored in this study. Any potential *trans*-action of lincRNAs on their partners, any possible physical interactions between them, and any effects of DNA methylation on these processes deserve further detailed exploration.

Data availability

The non-coding RNA sequencing data and small RNA sequencing data had been deposited and available in the NCBI (PRJNA629903). LincRNA (lincRNA and lincNAT) profiles with information about location and coding ability are available in Supplementary File S1. Supplementary material available at figshare: <https://doi.org/10.25387/g3.14034515>.

Acknowledgments

We appreciate the knowledge and trainings given by the course of Evolution Biology in Northeast Normal University.

Funding

This work was supported by the National Natural Science Foundation of China (grant nos. 31670220 and 31700187), the Recruitment Program of Global Youth Experts, and the Program of Changbai Mountain Scholar.

Conflicts of interest: None declared

Literature cited

- Akay A, Jordan D, Navarro IC, Wrzesinski T, Ponting CP et al. 2019. Identification of functional long non-coding RNAs in *C. elegans*. *BMC Biol.* 17:14.
- Bakhtiarzadeh MR, Hosseinpour B, Arefnezhad B, Shamabadi N, Salami SA. 2016. In silico prediction of long intergenic non-coding RNAs in sheep. *Genome.* 59:263–275.
- Bhan A, Soleimani M, Mandal SS. 2017. Long noncoding RNA and cancer: a new paradigm. *Cancer Res.* 77:3965–3981.
- Casero D, Sandoval S, Seet CS, Scholes J, Zhu Y et al. 2015. Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat Immunol.* 16:1282–1291.
- Chen R, Li M, Zhang H, Duan L, Sun X et al. 2019. Continuous salt stress-induced long non-coding RNAs and DNA methylation patterns in soybean roots. *BMC Genomics.* 20:730.
- Corem S, Doron-Faigenboim A, Jouffroy O, Maumus F, Arazi T et al. 2018. Redistribution of CHH methylation and small interfering RNAs across the genome of tomato *ddm1* mutants. *Plant Cell.* 30: 1628–1644.
- Deng F, Zhang X, Wang W, Yuan R, Shen F. 2018. Identification of *Gossypium hirsutum* long non-coding RNAs (lincRNAs) under salt stress. *BMC Plant Biol.* 18:23.
- Deng N, Hou C, Ma F, Liu C, Tian Y. 2019. Single-molecule long-read sequencing reveals the diversity of full-length transcripts in leaves of *Gnetum* (Gnetales). *Int J Mol Sci.* 20:6350.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22: 1775–1789.
- Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M et al. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci.* 107:8689–8694.
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI et al. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet.* 39:1033–1037.
- Gao C, Sun J, Dong Y, Wang C, Xiao S et al. 2020. Comparative transcriptome analysis uncovers regulatory roles of long non-coding RNAs involved in resistance to powdery mildew in melon. *BMC Genomics.* 21:125.
- Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y et al. 2013. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res.* 23:628–637.
- Grantham NJ, Wurman-Rodrich J, Terrett OM, Lyczakowski JJ, Stott K et al. 2017. An even pattern of xylan substitution is critical for interaction with cellulose in plant cell walls. *Nat. Plants.* 3:859–865.

- Harris CJ, Scheibe M, Wongpalee SP, Liu W, Cornett EM et al. 2018. A DNA methylation reader complex that enhances gene transcription. *Science*. 362:1182–1186.
- Hou C, Deng N, Su Y. 2019. PacBio long-read sequencing reveals the transcriptomic complexity and Aux/IAA. *Gene Evolution in Gnetum (Gnetales)*. *Forests*. 10:1043
- Hu L, Li N, Xu C, Zhong S, Lin X et al. 2014. Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc Natl Acad Sci*. 111:10642–10647.
- Hu L, Li N, Zhang Z, Meng X, Dong Q et al. 2020. CG hypomethylation leads to complex changes in DNA methylation and transpositional burst of diverse transposable elements in callus cultures of rice. *Plant J*. 101:188–203.
- Huang L, Dong H, Zhou D, Li M, Liu Y et al. 2018. Systematic identification of long non-coding RNAs during pollen development and fertilization in Brassica rapa. *Plant J*. 96:203–222.
- Jain P, Sharma V, Dubey H, Singh PK, Kapoor R, ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi-110012, India et al. 2017. Identification of long non-coding RNA in rice lines resistant to rice blast pathogen *Maganaporthe oryzae*. *Bioinformatics*. 13:249–255.
- Jiang H, Jia Z, Liu S, Zhao B, Li W et al. 2019. Identification and characterization of long non-coding RNAs involved in embryo development of Ginkgo biloba. *Plant Signal Behav*. 14:1674606.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 12:357–360.
- Kopp F, Mendell JT. 2018. Functional classification and experimental dissection of long noncoding RNAs. *Cell*. 172:393–407.
- Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 47: D155–D162.
- Kyriakou D, Stavrou E, Demosthenous P, Angelidou G, San Luis B-J et al. 2016. Functional characterisation of long intergenic non-coding RNAs through genetic interaction profiling in *Saccharomyces cerevisiae*. *BMC Biol*. 14:106.
- Lalitha S. 2000. Primer premier 5. *Biotech Software Internet Rep*. 1: 270–272.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10:R25.
- Lee JT, Bartolomei MS. 2013. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*. 152:1308–1323.
- Li L, Eichten SR, Shimizu R, Petsch K, Yeh C-T et al. 2014. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol*. 15:R40.
- Li Q, Gent JI, Zynnda G, Song J, Makarevitch I et al. 2015. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proc Natl Acad Sci USA*. 112:14728–14733.
- Li S, Yu X, Lei N, Cheng Z, Zhao P et al. 2017. Genome-wide identification and functional prediction of cold and/or drought-responsive lncRNAs in cassava. *Sci. Rep*. 7:45981.
- Liu J, Jung C, Xu J, Wang H, Deng S et al. 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell*. 24:4333–4345.
- Long JC, Xia AA, Liu JH, Jing JL, Wang YZ et al. 2019. Decrease in DNA methylation 1 (DDM1) is required for the formation of mCHH islands in maize. *J Integr Plant Biol*. 61:749–764.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.
- Lu X, Chen X, Mu M, Wang J, Wang X et al. 2016. Genome-wide analysis of long noncoding RNAs and their responses to drought stress in cotton (*Gossypium hirsutum* L.). *PloS One*. 11: e0156723.
- Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet*. 15: 394–408.
- Ouyang S, Buell CR. 2004. The TIGR plant repeat databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Res*. 32:D360–D363.
- Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL et al. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 22:577–591.
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 33:290–295.
- Qin T, Zhao H, Cui P, Albeshier N, Xiong L. 2017. A nucleus-localized long non-coding RNA enhances drought and salt stress tolerance. *Plant Physiol*. 175:1321–1336.
- Quinn JJ, Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*. 17:47–62.
- Scott EY, Mansour T, Bellone RR, Brown CT, Mienaltowski MJ et al. 2017. Identification of long non-coding RNA in the horse transcriptome. *BMC Genomics*. 18:511.
- Secco D, Wang C, Shou H, Schultz MD, Chiarenza S et al. 2015. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *eLife*. 4:e09343.
- Sun L, Luo H, Bu D, Zhao G, Yu K et al. 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 41:e166–e166.
- Tan F, Lu Y, Jiang W, Wu T, Zhang R et al. 2018. DDM1 represses non-coding RNA expression and RNA-directed DNA methylation in heterochromatin. *Plant Physiol*. 177:1187–1197.
- Uszczynska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R. 2018. Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet*. 19:535–548.
- Verhertbruggen Y, Marcus SE, Haeger A, Verhoef R, Schols HA et al. 2009. Developmental complexity of arabinan polysaccharides and their processing in plant cell walls. *Plant J*. 59: 413–425.
- Wang D, Qu Z, Yang L, Zhang Q, Liu Z-H et al. 2017. Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *Plant J*. 90:133–146.
- Wang H, Chung PJ, Liu J, Jang I-C, Kean MJ et al. 2014. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome Res*. 24: 444–453.
- Wang J, Zhang J, Zheng H, Li J, Liu D et al. 2004. Mouse transcriptome: neutral evolution of ‘non-coding’ complementary DNAs. *Nature*. 431:1–757.
- Wang L, Ma X, Xu X, Zhang Y. 2017c. Systematic identification and characterization of cardiac long intergenic noncoding RNAs in zebrafish. *Sci Rep*. 7:1250.
- Wang L, Xia X, Jiang H, Lu Z, Cui J et al. 2018. Genome-wide identification and characterization of novel lncRNAs in *Ginkgo biloba*. *Trees*. 32:1429–1442.
- Wang M, Yuan D, Tu L, Gao W, He Y et al. 2015. Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp). *New Phytol*. 207:1181–1197.
- Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*. 135:635–648.

- Xu W, Yang T, Wang B, Han B, Zhou H *et al.* 2018. Differential expression networks and inheritance patterns of long non-coding RNAs in castor bean seeds. *Plant J.* 95:324–340.
- Yan H, Bombarely A, Xu B, Frazier TP, Wang C *et al.* 2018. siRNAs regulate DNA methylation and interfere with gene and lncRNA expression in the heterozygous polyploid switchgrass. *Biotechnol Biofuels.* 11:208.
- Yang X, Li L. 2011. miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics.* 27: 2614–2615.
- Yuan J, Li J, Yang Y, Tan C, Zhu Y *et al.* 2018. Stress-responsive regulation of long non-coding RNA polyadenylation in *Oryza sativa*. *Plant J.* 93:814–827.
- Zhang L, Wang M, Li N, Wang H, Qiu P *et al.* 2018. Long noncoding RNAs involve in resistance to *Verticillium dahliae*, a fungal disease in cotton. *Plant Biotechnol J.* 16:1172–1185.
- Zhang Y-C, Liao J-Y, Li Z-Y, Yu Y, Zhang J-P *et al.* 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 15:512.
- Zhao T, Tao X, Feng S, Wang L, Hong H *et al.* 2018a. LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. *Genome Biol.* 19:195.
- Zhao X, Li J, Lian B, Gu H, Li Y *et al.* 2018b. Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat Commun.* 9:5056.
- Zheng XM, Chen J, Pang HB, Liu S, Gao Q *et al.* 2019. Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication. *Sci Adv.* 5:eaax3619.
- Zhu L, Zhu J, Liu Y, Chen Y, Li Y *et al.* 2015. Methamphetamine induces alterations in the long non-coding RNAs expression profile in the nucleus accumbens of the mouse. *BMC Neurosci.* 16:18.
- Zhu Q-H, Wang M-B. 2012. Molecular functions of long non-coding RNAs in plants. *Genes.* 3:176–190.

Communicating editor: B. Andrews