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

Juzuo Li, Ning Li, Ling Zhu, Zhibin Zhang, Xiaochong Li, Jinbin Wang, Hongwei Xun, Jing Zhao, Xiaofei Wang, Tianya Wang, Hongyan Wang, Bao Liu, Yu Li, and Lei Gong

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.

Abstract

Plant long non-coding RNAs (lncRNAs) function in diverse biological processes, and IncRNA expression is under epigenetic regulation, including by cytosine DNA methylation. However, it remains unclear whether 5-methylcytosine (5mC) plays a similar role in different sequence contexts (CG, CHG, and CHH). In this study, we characterized and compared the profiles of genome-wide IncRNA 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 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 that of their proximal protein-coding genes was also analyzed. Our study shows that CG methylation negatively regulates the TE-related expression of IncRNA and demonstrates that CHH methylation is also involved in the regulation of IncRNA 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 IncRNA transcripts have been identified in different species by transcriptome reassembly (Liu et al. 2012; Wang et al. 2015; Kyriakou et al. 2016; Uszczynska-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 IncRNA identification and characterization in animal models (Wang et al. 2004; Bakhtiarizadeh et al. 2016; Scott et al. 2017; Wang et al. 2017b), many studies have explored tissue IncRNA 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 IncRNAs 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 IncRNAs exhibit lower expression levels and strong tissue-specific
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; Yang 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/C0), and its isogenic wild type (OsMET1-2/C+). We show that genome-wide CG hypomethylation in OsMET1-2/C0 (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/C0. Using paired samples of OsMET1-2/C+ and OsMET1-2/C0, we consistently show that the expression of cis-acting lincRNAs is positively correlated with that of their paired PCgenes.

**Materials and methods**

**Plant materials**

The homozygous null mutant of OsMET1-2 (OsMET1-2/C0) and its isogenic wild type (OsMET1-2/C+) of Oryza sativa L. ssp. japonica cv. Nipponbare (Hu et al. 2014) were used in this study. OsMET1-2/C+ and OsMET1-2/C0 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 150 bp and single-end 50 bp 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://ecb.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 <200 nt 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 (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 that intersected with PCgenes by more than one nucleotide were defined as lncNATs.

For each lncRNA, 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 lncRNA

Twenty lincRNAs randomly selected from mutant-specific lincRNAs and lncNATs and from common lincRNAs and lncNATs 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 (Laithia 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 lncNATs. 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−ΔΔCt method was used to estimate relative expression, and ACTIN was used as the internal control gene.

Differential expression of PCgenes and lncRNAs

To identify differentially expressed PCgenes (DEPCgenes) and lncRNAs (DElncRNAs) 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://gtRNAdb.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 lncRNAs, 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 DElncRNAs, 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 lncRNAs in OsMET1-2+/+ and OsMET1-2−/−

Strand-specific RNA-sequencing and a stringent prediction pipeline were used to identify the long non-coding RNA (lncRNA) 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 annotatedgenic 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 lncRNAs, a final filtration step was performed to exclude transcripts with known and predicted coding potential. Finally, 932 and 1104 lncRNAs were identified in OsMET1-2+/+ and OsMET1-2−/− (Figure 1A; Supplementary File S1).

The genomic locations and transcription directions of the lncRNAs relative to their nearest neighboring PCgenes were determined, and the lncRNAs were then categorized into long intergenic non-coding RNA (lincRNA) and long non-coding natural antisense transcript (lncNAT). As shown in the Venn diagrams (Figure 1B), the 932 lncRNAs in OsMET1-2+/+ consisted of 729 lincRNAs and 203 lncNATs, and the 1104 lncRNAs of OsMET1-2−/− consisted of 880 lincRNAs and 224 lncNATs. Most lincRNAs and lncNATs were shared by the two genotypes (719 common lincRNAs and 201 common lncNATs; Figure 1B). However, there were a limited number of genotype-specific lincRNAs and lncNATs (10 and 2 wild type-specific lincRNAs and lncNATs; 23 mutant-specific lncNATs; 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).
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 lncRNAs (18.30%; 161/880) and lncNATs (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 lncRNA expression 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 lncRNAs and lncNATs 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 lncRNAs were statistically enriched in CHG DMRs (Supplementary Figure S3C), but mutant-specific lncNATs 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 lncRNAs and lncNATs 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...
autonomous TE-related lncRNAs 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 lncRNAs (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 lncRNAs expressed by specific TE types were summarized (Figure 3B). Overall, more lncRNAs were generated by Type II transposons (DNA transposons) than by Type I transposons (retro-transposons) for both common and mutant-specific lncRNAs (Figure 3B). In addition, mutant-specific lincRNAs and lncNATs...
were more highly expressed than common lncRNAs (Figure 3B). Notably, 49.69% of the mutant-specific lncRNAs were generated by the En/Spm DNA transposon family, significantly higher than the corresponding percentage of common lncRNA (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 lncNATs than common lncNATs (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 lncRNAs in the OsMET1-2\(^{-/+}\).

### RdDM-mediated CHH hypermethylation in the 5'-upstream regions of transcriptionally upregulated lncRNAs in OsMET1-2\(^{-/-}\)

To examine the link between DNA methylation and lncRNA expression in different contexts (i.e., CG, CHG, and CHH), genomic regions that contained differentially expressed lncRNA (DELncRNA [lncRNA and lncNAT]) transcripts and their ±2kb 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 lncRNAs and with common and mutant-specific lncRNAs were considered separately (Figure 4A; Supplementary Figures S5 and S6).

For DNA methylation in CG and CHG contexts, all lncRNA-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 lncRNAs exhibited CHH hypomethylation in OsMET1-2\(^{-/-}\) (Figure 4A). Specifically, CHH hypomethylation occurred in genomic regions that generated downregulated common lncRNAs and lncNATs (Figure 4A; Supplementary Figure S6). By contrast, in genomic regions that generated upregulated common and mutant-specific lncRNAs, 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 lncNATs (Supplementary Figure S6).

### Table 1

| Category       | Protein coding genes | TE-related genes | Total repeats | Retrotransposons (Class I/retro TE) | Transposons (Class II/DNA TE) |
|----------------|---------------------|-----------------|--------------|-----------------------------------|-------------------------------|
|                | OsMET1-2\(^{-/+}\)  | OsMET1-2\(^{-/-}\) | Decreased    | OsMET1-2\(^{-/+}\)  | OsMET1-2\(^{-/-}\) | Decreased    | OsMET1-2\(^{-/+}\)  | OsMET1-2\(^{-/-}\) | Decreased    |
| CG             | 26.40 / 3.30        | 85.40 / 18.40   | -87.60       | 89.60 / 21.10                    | 78.20 / 16.60                 | -78.80       | 2.10 / 0.70        | 65.50 / 10.70   | -59.50       |
| CHG            | 9.20 / 6.70         | 64.70 / 51.20   | -26.50       | 65.00 / 51.00                    | -21.50                       | 7.60 / 4.90  | 10.90 / 9.50       | -12.90        | 20.50 / 13.10  |
| CHH            | 54.10 / 43.80       | 59.30 / 48.80   | -18.70       | 61.20 / 55.30                    | -9.70                        | 4.60 / 2.50  | 3.30 / 4.30       | -29.10        | 23.90 / 7.90   |
|                |                     |                 |              |                                   |                              |              |                     |               |              |

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 \( \text{Percentage Decreased} = \text{Percentage}\) \( \frac{\text{OsMET1-2\(^{-/+}\)}}{\text{OsMET1-2\(^{-/-}\)}} \).
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 regulatory genomic regions (especially in the En/Spm TE regions) of transcriptionally upregulated lincRNAs in OsMET1-2/C0/C0.

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/C0/C0 relative to OsMET1-2/+/+ (Hu et al. 2014). Based on the DElncRNAs in the same sample set, it was possible to explore potential cis-regulatory effects of lncRNAs on the expression of their neighboring PCgenes. Specifically, we characterized the correlation between expression fold changes of DElncRNAs (including both common and mutant-specific lincRNAs and lncNATs) and those of their corresponding differentially expressed 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 lncRNAs 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 lncRNAs and lncNATs (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 DElncRNA expression in OsMET1-2/C0/C0 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 DElncRNA and those of their corresponding DEPCgenes (Figure 5, B and C). PCgenes paired with DElncRNA 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 lncRNA 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,
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 lncRNAs from various eukaryotic species (Kyriakou et al. 2016; Wang et al. 2017b; Akay et al. 2019). Major questions about lncRNA 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 lncRNA expression (Wang et al. 2017a; Xu et al. 2018; Chen et al. 2019). We therefore characterized and compared lncRNA expression (lincRNAs and lncNATs) 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 lncRNA expression, we also demonstrated the involvement of CHH methylation in the regulation of lncRNA 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 lncRNA expression by CG methylation: the heavily CG-methylated regions in OsMET1-2+/+ were induced to express novel mutant-specific lncRNAs 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 lncRNAs may have originated from TE-rich regions. To test this hypothesis, we investigated the composition of genomic regions that generated mutant-specific lncRNAs. A specific group of DNA transposons, the En/Spm DNA transposons, expressed more mutant-specific lncRNAs 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 lncRNA 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 lncRNAs in response to
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 lncRNA transcription does not correlate with their genomic abundance. This suggests that active lncRNA expression by En/Spm transposons is not determined by their intrinsic properties. Although both En/Spm transposons and MITEs are enriched in intergenic regions (Ouyang and Buell 2004), significant mutant-specific lncRNA 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 factor. 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 lncRNAs 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 lncRNA expression after the null mutation of OsMET1-2. In addition to the previously reported co-localization of TEs with expressed lncRNAs 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 lncRNA 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 En/Spm transposons (2.70% Table 1). The prima facie coincidence of lncRNA 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 lncRNA transcripts (Figure 4). Our observations suggest that together with CG methylation, CHH methylation mediated by the RdDM pathway is also involved in regulating lncRNA expression, especially for lncRNAs. However, in contrast to the clear negative effects of CG methylation on lncRNA 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 lncRNA regulatory regions could compensate the silence the TE transcription in the absence of inhibitory 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 lncRNA expression. Comparisons of lncRNA profiles from additional RdDM rice mutants will be necessary to determine whether intergenic lncRNAs 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 lncRNA expression will then be made clear.

LncRNA has been reported to regulate the expression of both neighboring (cis) and distant (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 lncRNAs exhibited positive correlations with their neighboring PCgenes in our rice materials. Given the abnormal phenotypes of OsMET1−/− (Hu et al. 2014), it will be interesting to construct lncRNA and/or PCgene mutants with which to characterize the specific functions of lncRNAs 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 lncRNAs in the regulation of gene expression at independent or distant loci were not explored in this study. Any potential trans-action of lncRNAs 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). LncRNA (lincRNA and lncNAT) 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.
Bakhhtiarizadeh 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 (lncRNAs) 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 JL, 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, Wurm-Rodrich J, Terrett OM, Lyczkowski 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 Magnaporthe 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. 172:399–407.

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. 2018. Systematic identification of novel lncRNAs in Ginkgo biloba. BMC Genomics. 18:511.

Decco D, Wang C, Shou H, Schultz MD, Chiarensa 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.

Usczenska-Ratajeczak B, Lagarde J, Grossi S, Guigo R, Johnson R. 2018. Towards a complete map of the human long non-coding RNA transcriptome. Nat Rev Genet. 19:535–548.

Verberne J, Markus SE, Haeger A, Verhoef R, Schols HA et al. 2009. Developmental complexity of arabian 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, Jiang 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 noncoding intergenic RNAs in zebrasfish. Sci Rep. 7:1250.

Wang L, Xia J, Jiang H, Lu Z, Cui J et al. 2018. Genome-wide identification and characterization of novel lncRNAs in Ginkgo biloba. Trees. 32:1429–1442.

Wierzbiicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IV/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.

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.

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.

Communicating editor: B. Andrews