


RESEARCH

Open Access



# siRNAs regulate DNA methylation and interfere with gene and lncRNA expression in the heterozygous polyploid switchgrass

Haidong Yan<sup>1,2</sup>, Aureliano Bombarely<sup>2</sup>, Bin Xu<sup>3</sup>, Taylor P. Frazier<sup>4</sup>, Chengran Wang<sup>1</sup>, Peilin Chen<sup>1</sup>, Jing Chen<sup>1</sup>, Tomas Hasing<sup>2</sup>, Chenming Cui<sup>5</sup>, Xinquan Zhang<sup>1</sup>, Bingyu Zhao<sup>2\*</sup> and Linkai Huang<sup>1\*</sup> 

## Abstract

**Background:** Understanding the DNA methylome and its relationship with non-coding RNAs, including microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), is essential for elucidating the molecular mechanisms underlying key biological processes in plants. Few studies have examined the functional roles of the DNA methylome in grass species with highly heterozygous polyploid genomes.

**Results:** We performed genome-wide DNA methylation profiling in the tetraploid switchgrass (*Panicum virgatum* L.) cultivar 'Alamo' using bisulfite sequencing. Single-base-resolution methylation patterns were observed in switchgrass leaf and root tissues, which allowed for characterization of the relationship between DNA methylation and mRNA, miRNA, and lncRNA populations. The results of this study revealed that siRNAs positively regulate DNA methylation of the mCHH sites surrounding genes, and that DNA methylation interferes with gene and lncRNA expression in switchgrass. Ninety-six genes covered by differentially methylated regions (DMRs) were annotated by GO analysis as being involved in stimulus-related processes. Functionally, 82% (79/96) of these genes were found to be hypomethylated in switchgrass root tissue. Sequencing analysis of lncRNAs identified two lncRNAs that are potential precursors of miRNAs, which are predicted to target genes that function in cellulose biosynthesis, stress regulation, and stem and root development.

**Conclusions:** This study characterized the DNA methylome in switchgrass and elucidated its relevance to gene and non-coding RNAs. These results provide valuable genomic resources and references that will aid further epigenetic research in this important biofuel crop.

**Keywords:** *Panicum virgatum* L., DNA methylation, Gene expression, Non-coding RNAs, RNA-dependent DNA methylation, Differentially methylated regions

## Background

Switchgrass (*Panicum virgatum* L.) is a warm season, perennial, C<sub>4</sub> grass native to the North American prairie [1]. Switchgrass has been identified as one of the most promising feedstock grass species for biofuel production [2]. There are two different ecotypes of switchgrass,

upland and lowland, which are distinguished according to their natural habitats. The majority of upland ecotypes are octoploid (2n = 8x = 72) and most lowland ecotypes are tetraploid (2n = 4x = 36) [3, 4]. The tetraploid switchgrass genome is comprised of two sets of closely related homeologous chromosomes, and it is highly heterozygous due to the outcrossing nature of the species [5, 6].

DNA methylation is an epigenetic mechanism that influences molecular processes pertaining to plant growth and development [7], including immunity to virus infection [8], gene imprinting [9], transposon silencing [10, 11], and embryogenesis [12]. Methylation on the

\*Correspondence: bzha07@vt.edu; huanglinkai@sicau.edu.cn

<sup>1</sup> Department of Grassland Science, Animal Science and Technology College, Sichuan Agricultural University, Chengdu 611130, China

<sup>2</sup> Department of Horticulture, Virginia Tech, Blacksburg, VA 24061, USA

Full list of author information is available at the end of the article



cytosine of three different nucleotide patterns is common in plants: CG, CHG, and CHH (H=A, T, or C) contexts [13]. DNA methylation can induce repression or activation of genes and transposable elements (TEs) in response to environmental and developmental signals. Therefore, characterization of cytosine methylation patterns can provide valuable information that allows for a greater understanding of the regulation of gene expression, and the dynamics of TEs in the shaping of the plant genome architecture.

Non-coding RNAs are RNA molecules that are not translated into proteins. There are two major classes of non-coding RNAs: microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). MiRNAs are small sequences between 20 and 24 nt in length. These molecules often regulate gene expression post-transcriptionally by either mRNA cleavage or translation repression [14, 15]. lncRNAs are transcripts of more than 200 nucleotides that have numerous functions within the plant cell. They have been shown to target promoters upstream of genes, function as precursors of small RNAs, generate alternatively spliced transcripts, produce endogenous small interfering RNAs, and form RNA–protein complexes [16]. Non-coding RNAs have been shown to be closely related to DNA methylation. It has been reported that siRNAs can increase CHH methylation in many plant species through the RNA-dependent DNA methylation (RdDM) pathway [17–20]. More recently, lncRNAs have also been demonstrated to have key roles in the regulation of DNA methylation in mammals [21, 22]. For example, the long non-coding ecCEBPA transcript, encoded within the CEBPA gene, binds to DNMT1 and limits methylation of the CEBPA gene [21]. Alternatively, the H19 lncRNA alters genome-wide DNA methylation patterns by regulating S-adenosylhomocysteine hydrolase in mammals [22]. To date, the relationship between lncRNAs and DNA methylation in plants, especially in C4 grasses, has not been extensively studied [23, 24].

C4 grasses are the most prominent in the grass family and include some of the most economically important crop species such as maize (*Zea mays*), great millet (*Sorghum bicolor*), sugarcane (*Saccharum officinarum*), and switchgrass. Primary production of these crops annually accounts for roughly 25% of arable land [25]. Currently, DNA methylation patterns have been characterized in several C3 grasses such as rice (*Oryza sativa*) [26], *Brachypodium distachyon* [27], and wheat [28]; however, few DNA methylation studies have been performed on C4 grasses, such as switchgrass, which usually contain highly heterozygous and polyploid genomes. Thus far, DNA methylation studies in model plants, such as *Arabidopsis thaliana*, rice, and *B. distachyon*, have generated several conserved rules: (1) mCG is dominating among

the three methylated contexts; (2) methylation may act as a mechanism for controlling gene expression, but the methylation variation depends on types of contexts and gene region; and (3) RdDM pathways are highly associated with mCHH [17, 26, 27, 29, 30]. Since switchgrass is a polyploid, it would be interesting to investigate if it has unique methylation patterns that are different from the general rules identified in other model plant species.

In this study, we performed global DNA methylation sequencing on leaf and root tissues of tetraploid switchgrass (cv. Alamo) and profiled expression of mRNAs, miRNAs, and lncRNAs in the leaf tissue. Our results revealed that siRNAs positively regulated DNA methylation at the mCHH sites surrounding genes and that DNA methylation may interfere with both gene and lncRNA expression in the polyploid switchgrass genome. In addition, we identified a small subset of genes hypomethylated in the root tissue that were characterized by differentially methylated regions (DMRs), which could potentially be involved in stimulus-related GO processes. Finally, we predicted two precursors (lncRNAs) of miRNAs that might function in cellulose biosynthesis, stress regulation, and stem and root development. Overall, this study described the DNA methylome of switchgrass and its relation to gene and non-coding RNAs. These results provide a platform for future epigenetic studies in biofuel crops.

## Results

### Single-base resolution landscapes of DNA methylation in switchgrass

Whole-genome bisulfite sequencing was applied to genomic DNA of switchgrass extracted from leaves and roots. In total, 411 million reads were generated for switchgrass leaf tissue and 392 million reads were generated for switchgrass root tissue. Of the reads generated in this study, 334 (81%) million reads from the leaf sample and 304 (77%) million reads from the root sample could be aligned to the reference genome of switchgrass (V4.1) (Additional file 1: Table S1). For both tissues, the effective coverage rate of cytosine was analyzed at the chromosome level, as well as in genic and repetitive regions, to evaluate the quality of the methylation data (Additional file 2: Data S1). The effective coverage rates of chromosomes ranged between 75 and 85%, whereas the coverage rates of gene regions ranged between 82 and 87%. For the repetitive regions, low complexity repeats (cryptic, tandem, and interspersed repeats) had coverage rates between 56 and 64%, and the remaining repetitive elements had coverage rates ranging from 69 to 86% (Additional file 2: Data S1). In addition, the sequencing depths of the leaf and root samples were calculated to be 29X and 26X, respectively. These numbers suggest that the

amount of detected cytosines reached saturation (Additional file 1: Table S1 and Additional file 3: Figure S1). Overall, the methylation data generated in this experiment had a highly effective coverage rate (~80%) of cytosine. The sequencing depths indicated a sufficient level for further methylation analysis.

A total of 66 and 77 million methylcytosines (mCs) were identified in the genomic DNA isolated from switchgrass leaves and roots, respectively. For the leaf tissue, the cytosine methylation levels were higher at the mCG sites (60%) than those at mCHG (40%) and mCHH (4%) sites (Table 1). The data from the root tissue displayed similar patterns of mCG (59%) and mCHG sites (39%), and increased mCHH sites (7%) (Table 1). Most of the mCs identified in this study were located in unannotated intergenic and TE regions (Fig. 1a). Most mCG and mCHG sites were either highly methylated (methylation levels >80%) or lowly methylated (methylation levels <10%). These sites shared similar patterns in both tissues. Most mCHH sites, however, only displayed low methylation levels (<15%) (Fig. 1b). Interestingly, the methylation levels of the intronic regions were higher than the levels of other regions (Fig. 1c).

A chromosome-scale view of the methylation sites and their levels was generated from the sequencing data (Fig. 1d and Additional file 4: Figure S2). It was discovered that the terminal regions of the chromosomes had a lower level of methylation at mCG and mCHG sites while methylation at mCHH sites exhibited a more flat distribution pattern.

Methylation across the three types was also found to vary for different plant species (Fig. 1e and Additional file 5: Table S2). For each species analyzed, mCG and mCHG sites had higher methylation levels across the genomes than mCHH sites, which were universally the lowest (<6%). Specifically, maize and *Arabidopsis* had the highest and lowest proportion of mCG (85%) and mCHG

(10%), respectively. For the four Poaceae plants examined in this study, the levels of mCG and mCHG were higher in switchgrass than *B. distachyon* and rice, but lower than similar sites in maize (Fig. 1e).

#### DNA methylation profiles in genic and transposable element rich regions

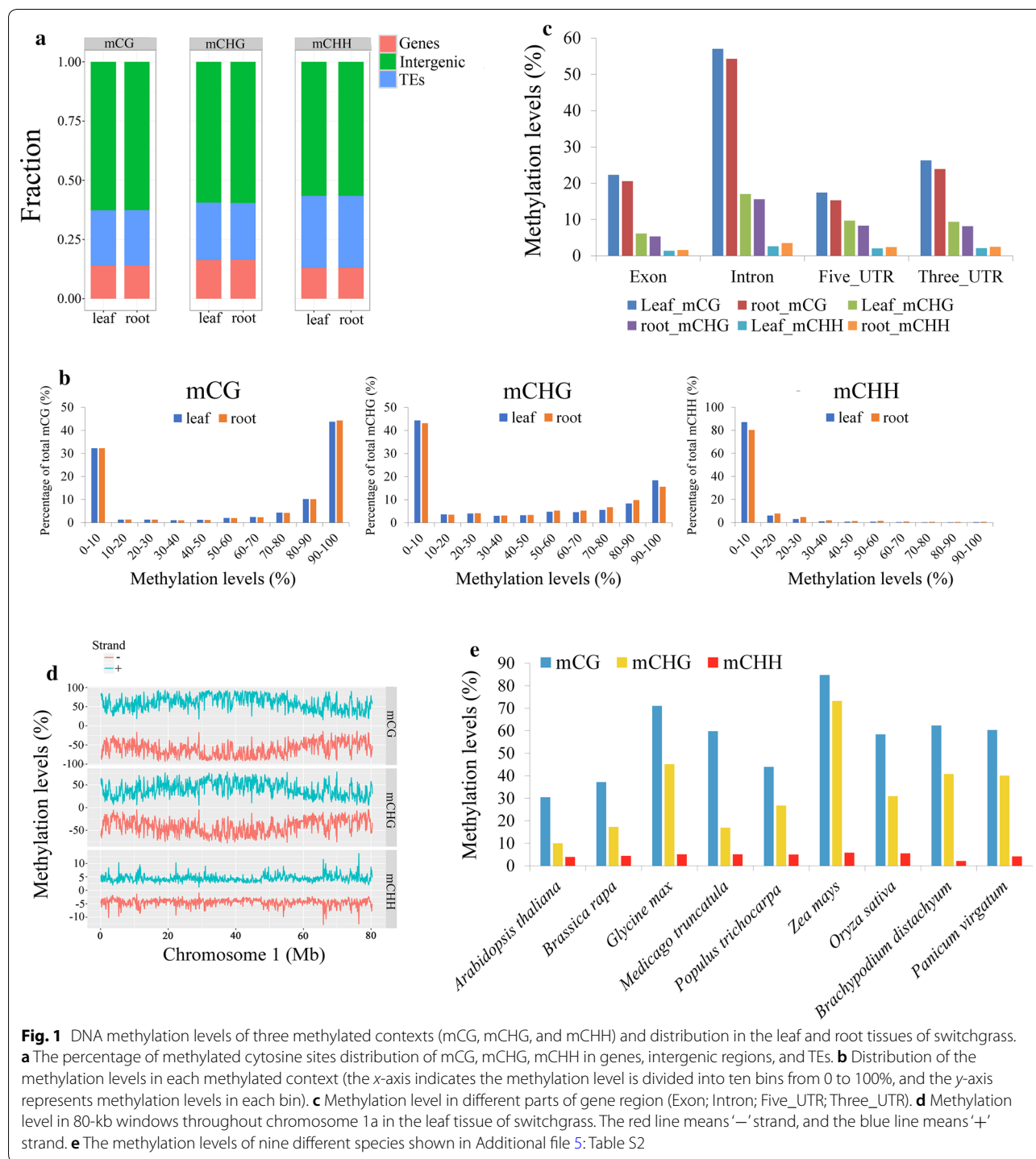
Methylation patterns in the protein-coding genes were examined by analyzing the three common methylation contexts across the gene body regions and the 2-kb regions flanking the genes (Fig. 2a, b). Relatively high methylation levels were discovered on the CG sites, followed by the CHG and CHH sites in the genic regions (Fig. 2a). Interestingly, the 5' flanking region upstream of the genes exhibited low levels of methylation that increased at the mCG and mCHG sites near the start of the coding sequence but subsequently decreased across the gene body. Another region of highly methylated region for these mCG and mCHG sites was observed near the point of transcription termination; however, lower levels of methylation were then found in the 3' flanking region (Fig. 2a, c).

The methylation patterns of TEs were also examined similarly to the genic regions and included both up- and downstream regions (Fig. 2b). The average methylation level of TEs was higher than that of protein-coding genes (Fig. 2a–c). The mCG and mCHG sites showed a similar trend where the methylation levels were much higher at the start and end point of transcription than in both the TE gene body and flanking regions. The methylation levels of mCHH sites were similar at the start and end points, as well as across the body and flanking regions, for all examined TEs (Fig. 2b). Previous reports have suggested that methylation levels of CG and CHG are positively correlated with genome size in plants [17], which was re-confirmed in this study using different set of plant genomes including switchgrass (Additional file 6: Figure S3). In addition, it has been shown that proliferation of TEs contributes to the genome size [31]. Several reports, along with the data generated in this study, have discovered that the average methylation levels in TEs are higher than in the genic regions (Fig. 2a–c) [19, 20, 32]. Therefore, TE methylation levels contribute to the overall genome methylation levels in plants and regulation of these levels may positively correlate with plant genome sizes.

To further characterize the methylation patterns within TEs, the methylation trends of ten major families of TEs were compared. These TEs were grouped into two classes: class I (retrotransposons including Copia, Gypsy, LTR-Other, LINE, and SINE), and class II (DNA transposons including DNA-Other, hAT, MULE-MuDR, EnSpm, and Stowaway) (Fig. 3a–d). While the methylation

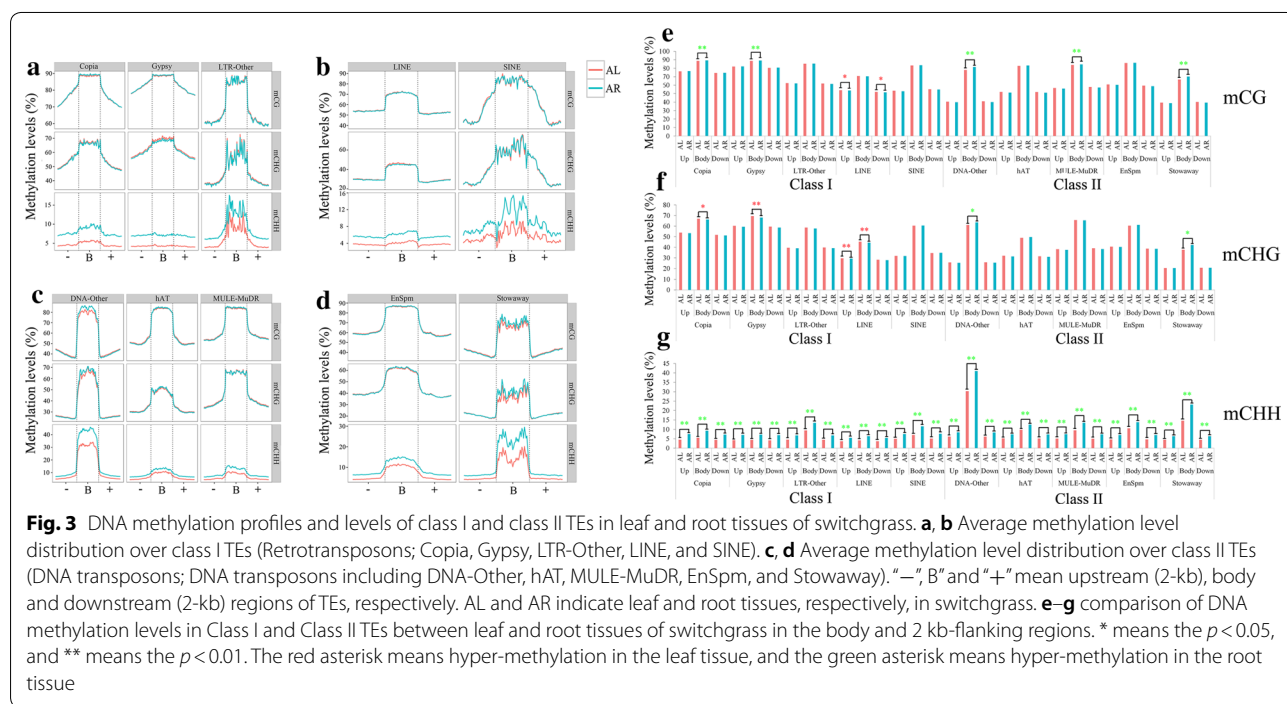
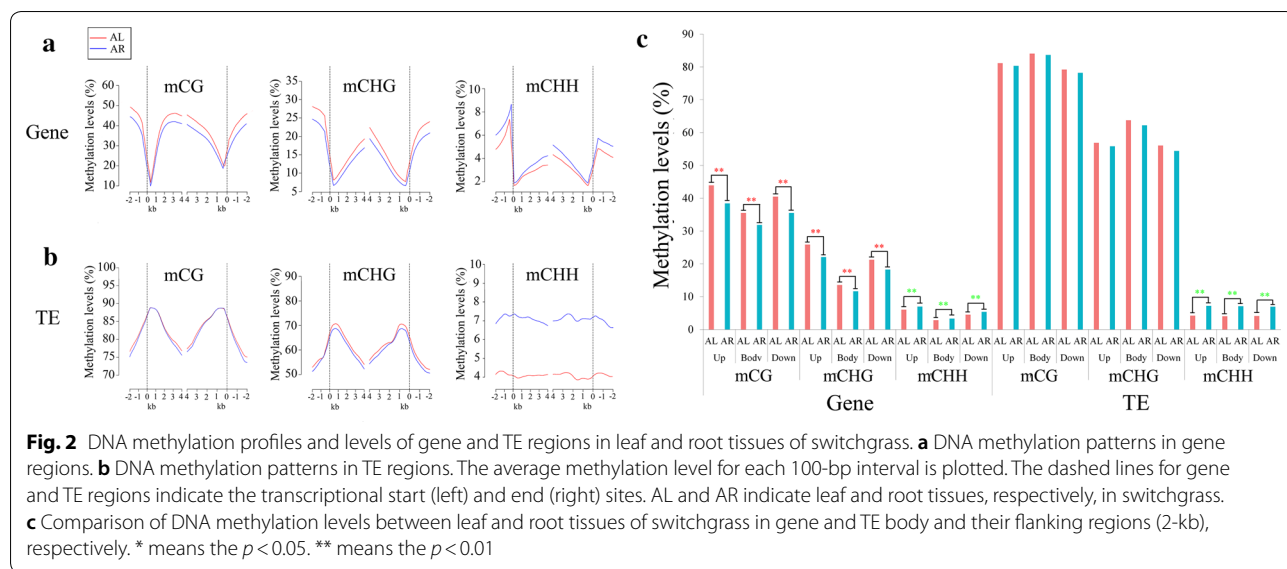
**Table 1 The number of epigenetically modified sites and the estimated proportion of the genome methylated according to three methylation contexts in leaf and root tissues of switchgrass**

	mCG	mCHG	mCHH
Leaf			
Number of modified sites	29,369,883	23,637,970	13,739,344
Proportion of the genome (%)	44.00	35.41	20.58
Methylation level (%)	60.34	40.06	4.16
Root			
Number of modified sites	30,097,757	23,937,098	22,509,824
Proportion of the genome (%)	39.32	31.27	29.41
Methylation level (%)	59.35	38.95	6.67



distributions of the ten TE families were different, they all displayed a similar pattern with obviously higher methylation levels in body regions than flanking regions. For the class I TEs, significantly higher methylation levels were detected on mCG and mCHG sites than on mCHH sites (Fig. 3a, b). In addition, the mCG and mCHG positions were hyper-methylated in the body regions of all

class I TEs, while the mCHH levels were equally methylated to those in the up- and downstream regions in Copia, Gypsy, and LINE. The mCHH sites were found to be hyper-methylated in the body regions of Copia and Gypsy. For the class II TEs, the levels of mCG and mCHG were also higher than mCHH (Fig. 3c, d). In comparison with the class I TEs, the class II TEs methylation patterns



showed a significant increase in methylation levels at all three methylation contexts across the DNA transposon gene body rather than in the flanking regions, despite mCHH being slightly elevated in the body regions of hAT and MULE-MuDR (Fig. 3c, d). The levels of mCG and mCHG were overall higher in class I than class II TEs (Additional file 7: Figure S4; Additional file 8: Table S3). Transposition of non-LTR retrotransposons (LINE and SINE) is rarely observed in plants, indicating that these

retroelements are inactive [33]. This could be attributed to the high methylation levels of mCG and mCHG sites observed in both LINE and SINE (Fig. 3b).

**DNA methylation differences between leaf and root tissues in switchgrass**

To identify global differentially methylated DNA between leaf and root tissues, the methylation levels of both genic and TE regions were compared. The analysis of these two

genomic components was further divided into upstream, body, and downstream regions (Fig. 2c; Additional file 9: Table S4). In the genic regions, the mCG and mCHG levels were significantly higher in leaves compared to roots, whereas the mCHH levels were significantly lower in leaves. In the TE regions, mCG and mCHG displayed no significant methylation differences between the leaf and root tissue. However, hypomethylated mCHH sites were observed in the TE body and flanking regions in the leaf tissue (Fig. 2c; Additional file 9: Table S4). Therefore, there is a tissue-specific regulation pattern of DNA methylation in the genic regions in switchgrass. These results are consistent with those found in *Arabidopsis*, Norway spruce (*Picea abies*), and castor bean (*Ricinus communis*) [17, 20, 34].

Differential methylation of the two TE classes was also observed between the leaf and root tissues in the TE body and flanking regions (Fig. 3e–g; Additional file 10: Table S5). For the mCG and mCHG sites in both TE classes, no significant methylation difference was identified between the leaf and root tissues (Fig. 2c); however, significant differences were detected when the methylated TEs were annotated into 15 regions (five regions per upstream, body, and downstream regions) for the two classes. For the class I TEs, four regions displayed significant mCG changes in leaves compared to roots, including two hypomethylated and two hyper-methylated regions. In addition, four regions were CHG hyper-methylated and all regions were CHH hypomethylated (Fig. 3e–g; Additional file 10: Table S5). For the class II TEs, three regions displayed significant CG hypomethylation and two regions were significantly CHG hypomethylated. All regions were CHH hypomethylated for this class in leaf tissue (Fig. 3e–g; Additional file 10: Table S5).

Differential methylation regions (DMRs) play an important role for epigenetic methylation modifications across plant species [35]. In this study, the DMRs were analyzed for switchgrass through a sliding window by subtracting the leaf methylation loci from the root methylation loci. A total number of 1,480,569 DMRs were discovered, and the number of identified CHH DMRs (574,126) was larger than that for CG DMRs (490,017) and CHG DMRs (416,426) (Fig. 4a). Most mCG DMRs (80.29%) were hyper-methylated, and nearly half of mCHG DMRs (55.39%) were hyper-methylated in the leaf tissue. Surprisingly, almost all mCHH DMRs (99.59%) were hypomethylated in the leaf tissue (Fig. 4b).

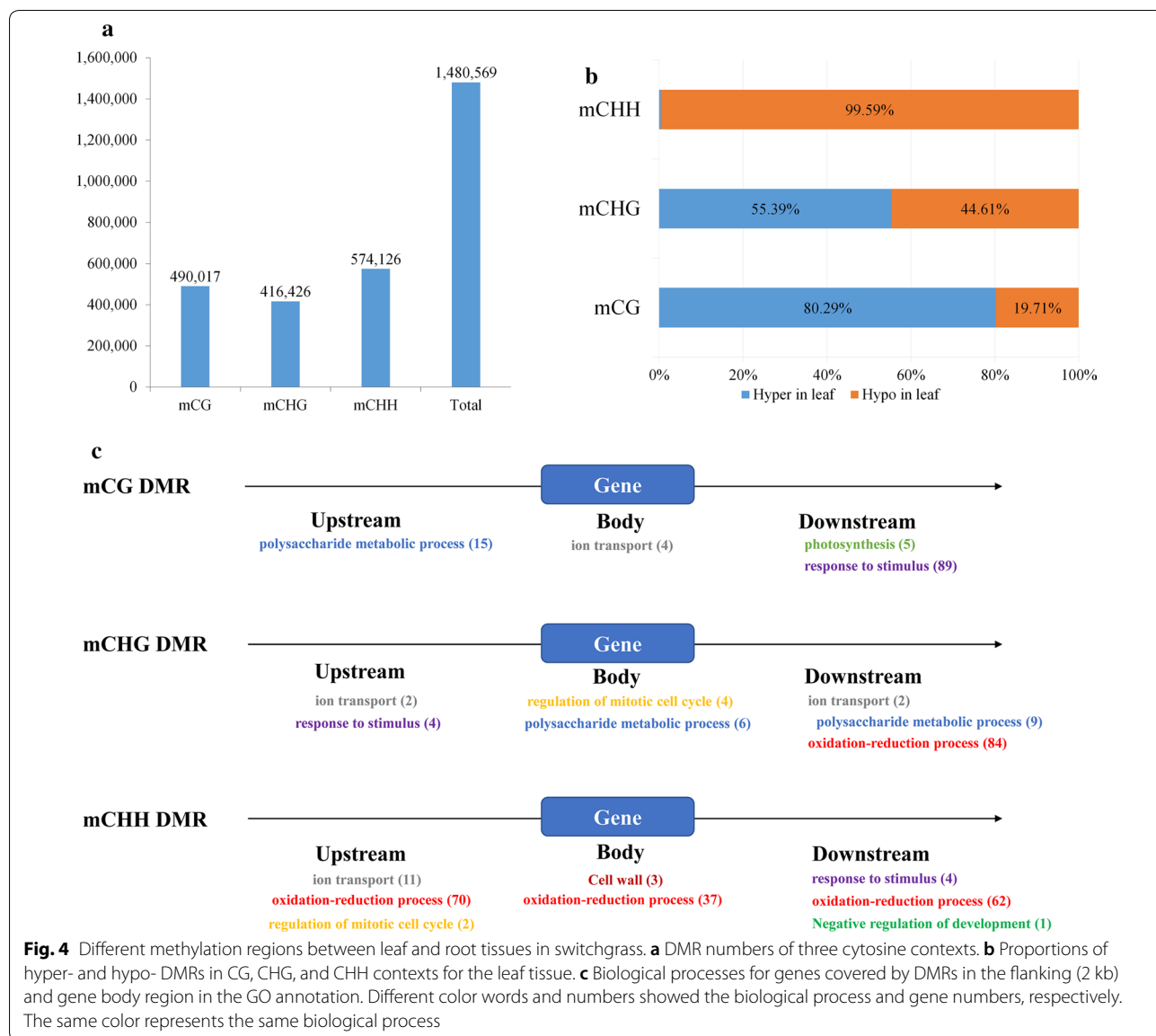
Plant roots are essential for responding to abiotic stress stimuli and must quickly react to environmental changes by controlling key stress-related genes [36, 37]. Previous reports have showed that root and leaf tissues were differently methylated [38–41]. To detect differential methylation modifications between leaf and root tissues in

switchgrass, we applied GO annotation towards genes covered by the DMRs ( $p < 0.05$ ; Additional file 11: Data S2). The enriched molecular processes were different among all three methylation contexts and the locations of the DMRs within genes (Fig. 4c; Additional file 11: Data S2). We identified 96 genes covered by the DMRs that were involved in various response-related processes including response to stimulus (GO:0050896), response to abiotic stimulus (GO:0009628), regulation of response to stimulus (GO:0048583), response to temperature stimulus (GO:0009266), and response to cold and freezing (GO:0009409 and GO:0050826) (Additional file 12: Data S3; Additional file 11: Data S2; Fig. 4c). Eighty-nine genes were covered by downstream DMRs on the mCG sites and one gene was covered by mCHG and mCHH DMRs in the upstream and downstream regions, respectively. For the other six genes, three were covered by mCHG DMRs in the upstream region and three were covered by mCHH DMRs in the downstream region. Majority of these genes (82%, 79/96) were hypomethylated in the root tissue, suggesting this subset of 79 genes may function in stress-related pathways by methylation modification. Additionally, 148 genes were identified that function related to oxidation–reduction processes (highlighted in red and yellow in Additional file 13: Figure S5). These genes were covered by mCHH DMRs in both the flanking and body regions (Fig. 4c and Additional file 13: Figure S5), and all were found to have CHH hyper-methylation in the root (Additional file 14: Data S4).

#### Effects of DNA methylation on gene expression

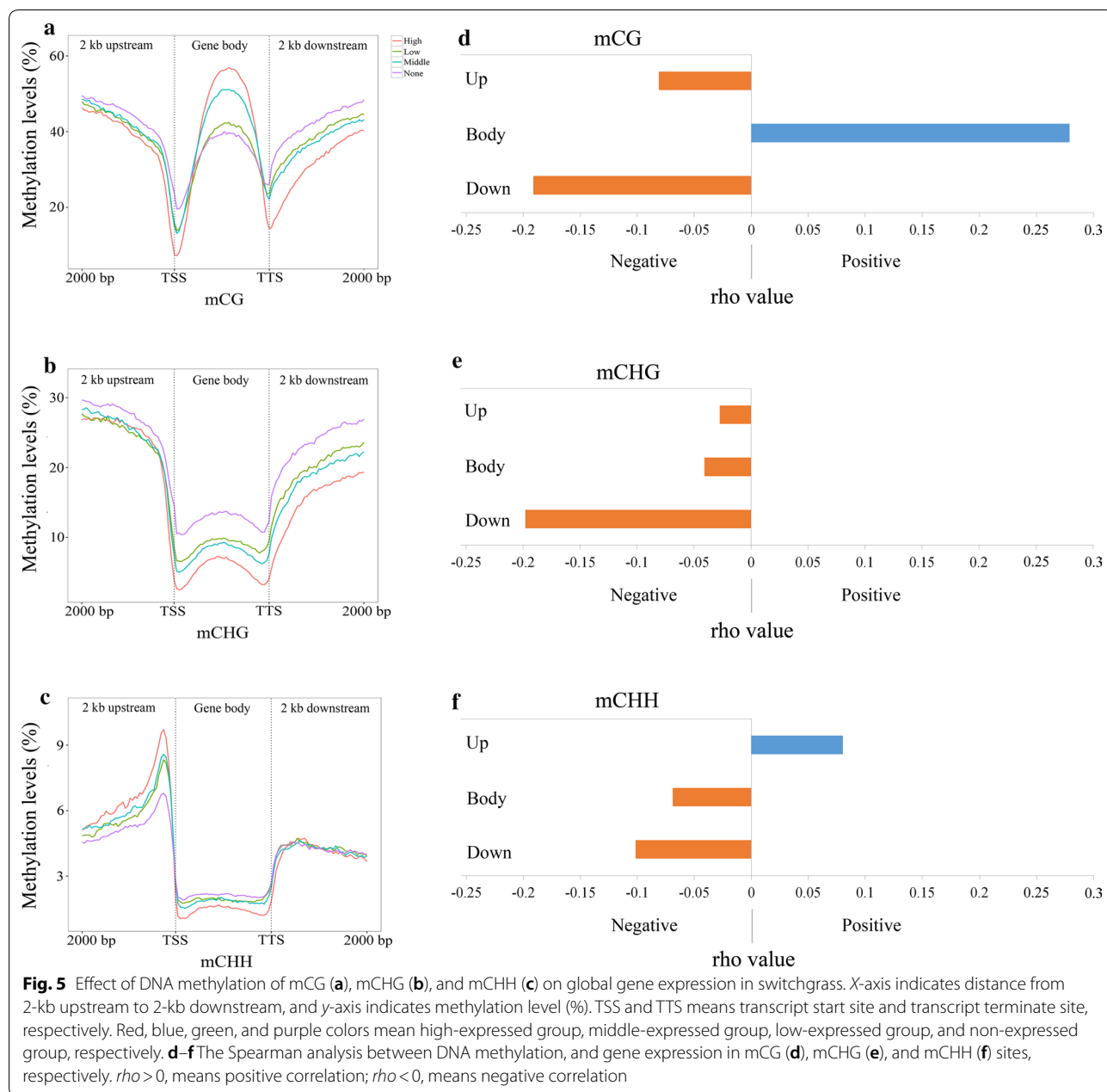
To further determine whether DNA methylation influences gene expression in switchgrass, total RNAs were extracted from the same sets of leaf samples and subjected to RNA-sequencing. A total of 150,545 transcripts were assembled from the RNA-seq data and these transcripts were grouped into four categories according to the criteria of Lu et al. [42]. These categories included a non-expressed group (reads per kilobase per million reads mapped [RPKM]  $\leq 1$ ), a low-expressed group ( $1 < \text{RPKM} \leq 10$ ), a middle-expressed group ( $10 < \text{RPKM} \leq 100$ ), and a high-expressed group ( $\text{RPKM} > 100$ ) (Additional file 15: Table S6 and Fig. 5a–c).

In the 2-kb regions upstream and downstream of genes, the category of non-expressed gene group contained the highest levels of mCG. Alternatively, the methylation levels at this site were the lowest for the high-expressed gene group (Fig. 5a). This indicates a negative association between mCG levels and gene expression in the flanking regions (Fig. 5). In contrast, the mCG levels found to occur in the gene body regions were positively correlated to the gene's expression level (Fig. 5a). For mCHG, the methylation levels in the gene body and downstream



regions were found to be negatively correlated with gene expression. Correspondingly, the non-expressed gene category contained the highest mCHG levels in the upstream, gene body, and downstream regions (Fig. 5b). For the mCHH sites, methylation levels in the upstream region close to transcription starting sites (TSS) were positively correlated with gene expression (Fig. 5c). There was a negative association, however, between gene expression and mCHH levels in the gene body regions, although the changes in methylation levels were minimal across the different categories for this region (Fig. 5c). No obvious relation was observed between mCHH levels in the downstream region (2-kb or less) of genes or in the proximity close to transcription termination sites (TTS) (Fig. 5c).

Spearman correlation analysis was performed to discern statistically the relationships between DNA methylation and gene expression within the 2-kb flanking regions of the protein-coding genes [43] (Fig. 5d–f; Additional file 16: Table S7). The results from that analysis found similar associations between DNA methylation levels and gene expression patterns (Fig. 4a–c). For the mCG sites, there was a significantly different positive correlation of methylation across the gene body region ( $\rho = 0.279$ ,  $p = 2.25E-267$ ), whereas mCG in the upstream ( $\rho = -0.080$ ,  $p = 2.42E-51$ ) and downstream ( $\rho = -0.191$ ,  $p = 2.02E-289$ ) regions exhibited a significantly different negative correlation (Fig. 4d; Additional file 16: Table S7). The mCHG sites were found to be



negatively and significantly correlated with gene expression in the body and flanking regions; however, the absolute  $\rho$  of the upstream ( $\rho = -0.027$ ,  $p = 5.25E-07$ ) and body regions ( $\rho = -0.045$ ,  $p = 1.14E-17$ ) were lower than that of the downstream regions ( $\rho = -0.198$ ,  $p = 2.71E-311$ ). This suggests a weaker correlation between CHG methylation and gene expression in the upstream and body regions than in the downstream regions (Fig. 5e; Additional file 16: Table S7). For mCHH sites, the upstream region exhibited a positive and significant correlation ( $\rho = 0.078$ ,  $p = 5.05E-47$ ), whereas the

gene body and downstream regions displayed negative and significant correlations ( $\rho = -0.068$ ,  $p = 7.6E-39$  in the body;  $\rho = -0.107$ ,  $p = 6.84E-92$  in the downstream) (Fig. 5f). Regression analysis, based on a zero hurdle model, was conducted to verify the associations between DNA methylation and gene expression [44]. Those results found that the mCG sites of the gene body and the mCHH sites of upstream regions were significantly and positively associated with gene expression (model coefficient = 0.01255,  $p = 2.83E-244$  for the mCG sites; model coefficient = 0.02431,  $p = 1.63E-41$  for the



mCHH sites). Other methylated regions, which had zero model coefficients less than 0, had negative associations with gene expression. The predictivity of the regression model was tested using a tenfold cross-validation method [45] and generated small positive  $Q^2$  values, suggesting a limited ability of the model to predict gene expression (Additional file 17: Table S8). This is expected as gene expression is regulated by a number of different genetic and environmental factors interacting with each other.

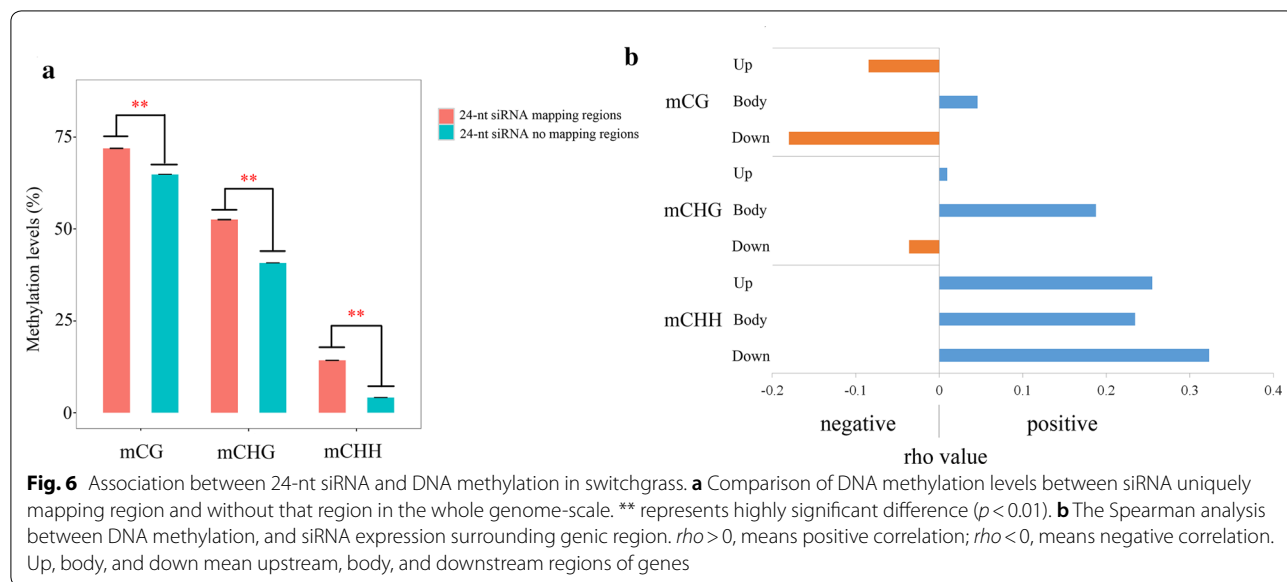
Collectively, the Spearman and regression analyses determined that the mCG and mCHH sites were positively correlated with gene expression in the gene body and upstream regions, respectively. The other regions analyzed exhibited negative correlations. Despite these results, most regions had very weak correlations according to previously established criteria ([www.statstutor.ac.uk/resources/uploaded/spearman.pdf](http://www.statstutor.ac.uk/resources/uploaded/spearman.pdf); the criteria ‘very weak’ was defined as the absolute value of  $\rho < 0.2$ ) in the Spearman analysis. Regardless, these weak correlations supported all of the relationships predicted through RNA-seq analysis with the exception that no association was found in the downstream region for mCHH sites (Fig. 5a–c).

**mCHH is positively correlated with siRNA expression in genic regions**

siRNAs can direct the de novo methylation of cytosine of complementary DNA sequences [46, 47]. In this study, switchgrass miRNAs were profiled using high-throughput sequencing of leaf tissue to elucidate the relationship between DNA methylation and siRNA expression. Switchgrass siRNAs were extracted from the miRNA data. The 24-nt class was the most abundant

group (Additional file 18: Figure S6). The 24-nt siRNAs were mapped to the switchgrass reference genome, and the average methylation levels in the regions mapped with or without a siRNA were calculated (Fisher’s exact test,  $p < 0.05$ ; Additional file 19: Table S9). The levels of mCG, mCHG, and mCHH in the siRNA-covered regions were significantly ( $p < 0.01$ ) higher than in the regions not targeted by any siRNAs (Fig. 6a). Therefore, these three methylation sites might positively correlate to siRNAs on a whole-genome scale.

Spearman correlation analysis was employed to further understand if the relationship between siRNA and DNA methylation occurs within the gene body and the 2-kb flanking regions surrounding the genes (Fig. 6b; Additional file 20: Table S10). The mCG sites exhibited a positive and significant correlation in the gene body region ( $\rho = 0.044$ ,  $p = 1.68E-07$ ), but negative correlations in the upstream ( $\rho = -0.085$ ,  $p = 2.77E-46$ ) and downstream regions ( $\rho = -0.184$ ,  $p = 3.26E-192$ ). Similarly, the mCHG sites were found to have positive and significant correlations between siRNA and DNA methylation in the gene body region ( $\rho = 0.185$ ,  $p = 1.39E-110$ ), and a negative correlation in the downstream flanking region ( $\rho = -0.035$ ,  $p = 2.56E-08$ ). No correlation between the two elements was determined for this site for the upstream gene region ( $p > 0.05$ ). In contrast, the gene body ( $\rho = 0.228$ ,  $p = 3.42E-214$ ), as well as the flanking regions (upstream:  $\rho = 0.253$ ,  $p = 1.02E-171$ ; downstream:  $\rho = 0.317$ ,  $p = 2.02E-115$ ), were found to have positive and significant correlations at the mCHH sites. Regression analysis was conducted based on a negative binomial model [44] to validate these relationships. For mCG sites, regression analysis identified positive and



significant correlations in the gene body (model coefficient = 0.00845,  $p = 1.84E-69$ ) and upstream regions (model coefficient = 0.00177,  $p = 6.52E-11$ ), and a negative and significant correlation in the downstream region (model coefficient = -0.00412,  $p = 1.27E-46$ ). The mCHG sites displayed positive correlations in the upstream and body regions (model coefficient = 0.00558,  $p = 2.07E-60$  for the upstream region; model coefficient = 0.00835,  $p = 1.98E-33$  for the body region); however, there was no significant correlation in the downstream region ( $p > 0.05$ ). In contrast, mCHH sites in all regions exhibited positive and significant correlations (model coefficient  $> 0$ ,  $p < 0.05$ ). Through the tenfold cross-validation analysis, the model showed a limited power (small positive  $Q^2$  values) to predict the siRNA expression (Additional file 21: Table S11).

Collectively, the levels of mCHH were positively and significantly correlated with siRNA expression in both the gene body and flanking regions. The mCHH sites had overall higher correlation  $\rho$  than the mCG and mCHG sites in the Spearman analysis (Fig. 6b). Taken together, these results suggest that RdDM plays a critical role in directing formation of mCHH in switchgrass.

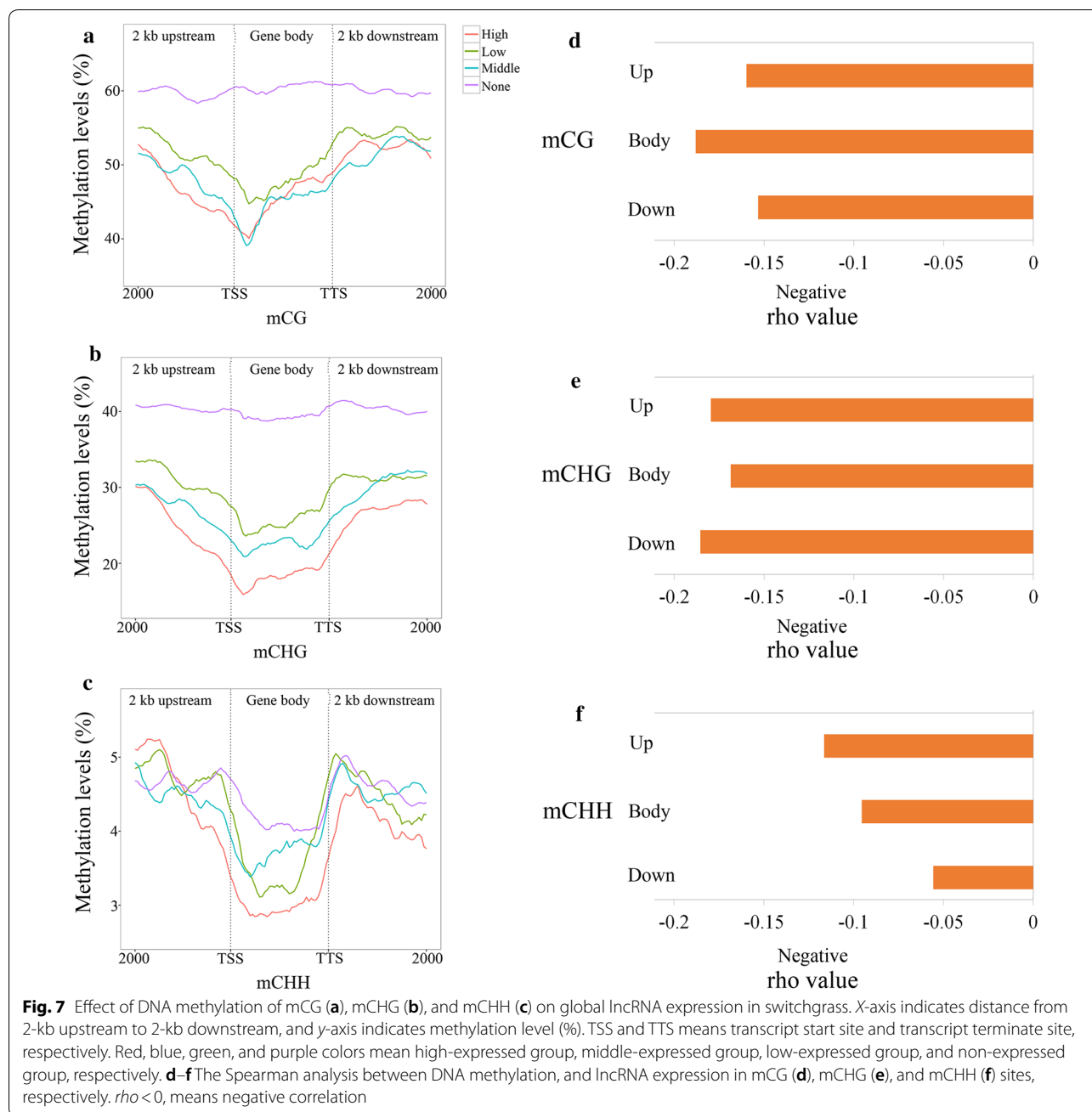
#### Identification of long non-coding RNAs and their negative relation to DNA methylation

lncRNAs have been shown to be closely associated with DNA methylation [21]. A total of 9244 lncRNAs were identified in the leaf tissue. These lncRNAs ranged in length from 200 base pair (bp) to 3023 bp with an average of 399 bp. The lncRNAs were found to be close to adjacent genes and had an average distance of 12,632 bp away from their flanking genes. This number is higher than that found in similar studies for rice (871 bp) and maize (6761 bp) [48]. The switchgrass lncRNAs could be classified into six groups, the majority of which were classified as the intergenic group (8217; 88.89%). This was followed by the bidirectional (470; 5.08%) and antisense groups (381; 4.12%). The rest of the lncRNAs ( $< 100$ ) were classified into three additional small groups: intronic sense (1.07%), sense (0.77%), and intronic antisense (0.06%) (Additional file 22: Figure S7). To determine if the switchgrass lncRNAs have homologous genes in rice and maize, the identified 9244 lncRNAs were BLAST against 8594 rice and 4403 maize lncRNAs in CANTATAdb (<http://cantata.amu.edu.pl/index.html#about>) [49]. Surprisingly, only 63 (0.68%) and 308 (3.33%) switchgrass lncRNAs were found to have homologous genes in rice and maize, respectively. Therefore, most of the lncRNAs identified in this study are unique to switchgrass. The identified lncRNAs were then BLAST against the Rfam family database (<http://rfam.xfam.org/>) [50]. Seven lncRNAs (Additional file 23: Table S12) were found to belong to the HAR1A

[51], Xist\_exon1 [52], SOX2OT\_exon2 [53], H19\_3 [54], ZEB2\_AS1\_2 [55], and RFPL3 [56] non-coding RNA families, which have been characterized in animal and human systems [57]. The function of these seven lncRNAs in plants, however, still needs to be evaluated. The lncRNAs were aligned to miRbase (v21). Two lncRNAs, TCONS\_00103604 and TCONS\_00155383, were identified that could potentially be the precursors of miR169 and miR171, respectively (Additional file 24: Table S13). According to the gene annotations of these miRNA targets, three pectinesterase inhibitor genes are targeted by miR169 and an additional two genes are believed to be involved in responses to stressful conditions. The genes targeted by miR171 were primarily predicted to be Scarecrow (SCR) genes (Additional file 24: Table S13).

DNA methylation allows for an organism to quickly respond to environmental stimuli by altering levels of mRNA expression [58]. lncRNAs have structural and functional regulations similar to mRNAs [59–61]. Thus, the expression of lncRNAs might also be regulated by DNA methylation. The 9244 switchgrass lncRNAs were classified into four groups based on their expression levels (high-expressed group, middle-expressed group, low-expressed group, and non-expressed group) using the same criteria as shown in Fig. 4a–c [42]. DNA methylation association patterns at mCG and mCHH sites were found to be different between mRNA and lncRNA genes. In general, the non-expressed lncRNA group was hypermethylated on the mCG and mCHG sites, which was not found for the mRNA genes (Figs. 5a, b, 7a, b). In the upstream flanking region, the non-expressed lncRNAs displayed the highest mCG levels. In contrast, the high-expressed lncRNAs were hypomethylated at this site (Fig. 7). These results suggest that mCG in the upstream flanking region might be negatively associated with gene expression of lncRNAs (Fig. 7a). In the gene body and downstream flanking regions, the non-expressed group exhibited the highest methylation levels, and the high- and middle-expressed lncRNAs displayed similar methylation patterns (Fig. 7a). Additionally, the mCHG sites of lncRNAs had comparable patterns with mRNAs, and the mCHG levels in both the flanking and body regions were found to be negatively associated with gene expression (Figs. 5b, 7b). In the gene body region, mCHH levels were determined to be negatively associated with gene expression; however, this was not found to be true in the flanking regions (Fig. 7c).

We also performed Spearman and regression analysis (based on a zero hurdle model) to verify the predicted relationships of lncRNAs within the 2-kb flanking regions (Fig. 7d–f; Additional file 25: Table S14; Additional file 26: Table S15). For all three methylation contexts, these two different statistical approaches both



showed that methylation levels had significant negative correlations to lncRNA expression in the 2-kb flanking regions as well as the gene body region, except for the mCHH sites of downstream region ( $p > 0.05$ ) in the regression analysis. This predictivity of regression model was also tested via the tenfold cross-validation method. The yielded small positive  $Q^2$  values from the model exhibited a restrictive ability to predict the lncRNA expression. (Additional file 25: Table S14; Additional file 26: Table S15). Overall, these results

supported the findings of a negative relationship between lncRNA expression and mCHG in the flanking and body regions, mCG in upstream region, and mCHH in the gene body region (Fig. 7a–c).

### Discussion

Switchgrass is a largely self-incompatible species [62, 63]. The relatively complex genome of switchgrass has hampered the development of whole-genome sequencing on this important biofuel crop. Recently, the latest version

of the switchgrass genome (v4.1) was released, which contains high-chromosome-scale contiguity that can allow for researchers to apply modern omics techniques to this plant species. A recent report generated a profile of DNA methylation on genomic features of two switchgrass genotypes, AP13 (Alamo clone) and VS16 [64]. In general, this study found similar methylation patterns for switchgrass as the ones identified in the current study: methylation levels of genic flanking regions were higher than in the gene body region and TEs themselves were more highly methylated than the upstream and downstream regions flanking them. However, this previous study did not examine the association between the DNA methylome and the transcriptome of switchgrass. To resolve relationships between these two omics, we conducted genome-wide single-base resolution methylome, lncRNA, miRNA, and mRNA sequencing and aligned the data to the latest switchgrass genome (v4.1) with a relatively high assembly quality.

#### Switchgrass has unique DNA methylation features but also shares general DNA methylation profiles with other plant species

In this study, the methylation distribution across the gene body was analyzed. It was discovered that the methylation levels of intron regions were higher than other regions in the switchgrass genome (Fig. 1c). These results are not consistent with those found in *C<sub>3</sub>* model plants, such as *Arabidopsis*, rice, and poplar [65]. In addition, the mCG sites were the most dominant in switchgrass, followed by mCHG and mCHH. The levels of mCG and mCHG were the highest in the gene body region (Fig. 2a), which is consistent with observations in most plants [34, 36, 42, 66, 67]. The gene flanking regions in switchgrass, however, displayed higher mCG levels than in the gene body region (Fig. 2a, c). This is a stark contrast to methylation patterns in most other plant species [32, 68, 69] and suggests that more methylation modifications, and subsequently more epigenetic mechanisms, may occur in the gene flanking regions of switchgrass. The average methylation level of TEs was higher than that of the genic regions in switchgrass, which is consistent with results found in most other plant species [17, 20, 32]. Additionally, we discovered that the terminal chromosome regions were less methylated at the mCG and mCHG sites, which is similar to rice [17, 20, 32]. It has been reported that the largest number of repeats with high methylation levels exists in telomeric heterochromatin regions [70]. Therefore, the lower methylation levels detected in the terminal chromosome regions in this study may indicate that these terminal regions do not cover telomeric heterochromatin, which could be caused by sequence limitations of

the current switchgrass genome. Overall, despite having a polyploid genome and several methylation differences, switchgrass does indeed share some DNA methylation profiles with the majority of plant species.

#### Identification of stress-related genes covered by DMRs

Switchgrass has a relatively large and strong root system that allows it to be more tolerant to various abiotic stresses [71]. In addition, identification of DMRs between roots and leaves might help find stress-related genes controlled by DNA methylation.

A total of 96 genes covered by DMRs were found to be involved in stimulus-related GO processes. For these genes, 82% were mCG hypomethylated in the downstream region and mCHG hypomethylated in the upstream region in the root tissues (Additional file 12: Data S3). Since the levels of mCG in the downstream regions and mCHG in the upstream regions were negatively correlated with gene expression, these 79 genes may possibly be upregulated by hypomethylation in the root tissue (Fig. 4c). Additionally, nine of the 79 hypomethylated genes were annotated to be auxin response factors and Small Auxin Upregulated RNA (SAUR) genes, which have been shown to be involved in stress defense responses [72]. These data suggest that switchgrass roots can quickly respond to environmental stresses, as well as maintain proper growth and development, by changing methylation on some auxin and stress-responsive genes. Some of these 79 genes also contained specific functional domains. For example, two genes contained Mildew resistance Locus O (MLO) domains, which are known to function as negative regulators of broad spectrum disease resistance [73]. In addition, one gene had a flavin adenine dinucleotide (FAD)-binding domain that is related to DNA photolyase [74]. Finally, one gene possessed a DOS2-like protein (BSD) domain, which has been reported in a rice *OsBSD* gene and has been shown to have a crucial role in plant growth [75] (Additional file 12: Data S3).

The mCHH DMRs covered a different subset of 148 switchgrass genes, which were identified and annotated to function in oxidation–reduction processes. All of these genes were hyper-methylated in the root tissues (Additional file 14: Data S4). The reactive oxygen intermediates (ROIs) produced as signaling molecules by these gene products, along with each gene's role in oxidation–reduction reactions, could control switchgrass response to various processes including pathogen defense, programmed cell death, abiotic stress response, and systemic signaling [76]. Unfortunately, the mRNA expression profiles of these genes were not examined in the switchgrass root tissue. Therefore, additional studies are needed to determine if all or most of these 148 DMR genes have higher

or lower expression levels in roots compared to other plant tissues. It would be interesting to further characterize if these genes do indeed play roles in switchgrass tolerance to abiotic stress.

#### **DNA methylation can either positively or negatively regulate gene expression depending on the methylation sites in different genic regions**

DNA methylation is not simply an inhibitor of gene expression [17, 26, 27, 29, 30]. In the gene promoter regions, the presence of DNA methylation usually suppresses transcription initiation, although low methylation levels occurring in these regions may promote gene expression. Our results found a negative relationship between DNA methylation of CG and CHG sites in the gene promoter regions and gene expression, suggesting that methylation of promoters represses gene expression in switchgrass (Fig. 5a, b). This observation was consistent with studies in *Arabidopsis* [26, 30, 67] and rice [26, 30, 67]. These results support previous studies that found that promoter symmetrical methylation is a common epigenetic mechanism controlling gene expression in eukaryotes. Interestingly, the mCHH levels in the upstream gene regions were positively correlated with gene expression (Fig. 5c). This could be attributed to TEs located close to or within the promoters of nearby genes [77, 78]. It is possible that RNA polymerase II- or IV-mediated transcription, which is initiated by TEs, can spread to the nearby gene regions and subsequently increase expression of the nearby genes [46, 79]. We observed that a higher number of 24-nt siRNAs target to TEs that are closer to genes than those TEs that are located at the intergenic regions (Additional file 27: Figure S8). Studies have shown that siRNAs can induce mCHH methylation through the RdDM pathway [80]. The abundance of siRNAs in these areas can result in a higher level of mCHH. This would explain the positive association between mCHH sites and gene expression in the upstream promoter region (Fig. 5c).

In the gene body region, the mCG sites were positively correlated with gene expression (Fig. 5a). This phenomenon could be attributed either to DNA methylation directly/indirectly preventing the initiation of intragenic promoters or to DNA methylation interfering with the activities of repetitive DNAs within the transcriptional unit [81]. In contrast, a negative association was found to exist between gene expression and methylation levels of mCHG and mCHH in the gene body regions, suggesting that the mCHG and mCHH in the actively transcribed genes may not inhibit gene transcription (Fig. 5b, c) [82, 83].

The data from this study also revealed that mCG and mCHG sites in the transcriptional termination regions

(TTR) were negatively correlated with gene expression (Fig. 5a, b). The differences in methylation levels among four categories (high, low, middle, and non-expressed genes) near the TTR regions were substantially greater than in the promoter regions. These results imply that mCG and mCHG in the TTR region might have a more significant role in the regulation of gene expression than their corresponding sites in the promoter region (Fig. 5a, b).

Overall, the mCG and mCHH were found to be positively related to gene expression in the gene body and upstream flanking regions, respectively. These results differed from other regions that displayed a negative correlation (Fig. 5). Taken together, these findings suggest that DNA methylation can either positively or negatively regulate gene expression and control of these processes depends on the methylation sites in different genic regions.

#### **The 24-nt siRNAs induce increased mCHH levels in the genic regions of switchgrass**

RdDM is a process by which 24-nt siRNAs direct the de novo methylation of cytosine of complementary DNA sequences. The de novo methylation of DNA at all mCG, mCHG, and mCHH contexts is usually performed by domains rearranged methylase 2 (DRM2) through the RdDM pathway [84]. In this study, mCHH levels, but not mCG and mCHG levels, were found to have a positive correlation with siRNA expression (Fig. 6b). After the initial methylation of DNA, mCG and mCHG sites could be sustained by copying the information from the parental strand after DNA replication. In comparison, most mCHH sites need to be maintained de novo after each round of DNA replication through the RdDM pathway [80]. Therefore, only the production of mCHH sites was positively correlated with siRNA expression in the switchgrass genic region, suggesting that 24-nt siRNAs can significantly increase mCHH levels in the canonical RdDM pathway. However, on a whole-genome scale, these three methylation contexts might all be positively correlated with siRNA expression (Fig. 6a), suggesting that siRNAs might activate mCG and mCHG in the intergenic and TE regions rather than the genic regions.

#### **Identification of two functional precursory lncRNAs of miRNAs**

Previous studies suggest that lncRNAs could be precursors of miRNAs [85, 86]. By aligning the lncRNAs identified in this study to miRbase (v 21), two lncRNAs TCONS\_00103604 and TCONS\_00155383, were found that could be the precursors of miR169 and miR171, respectively (Additional file 24: Table S13). Based on the annotations of the genes targeted by these miRNAs,

we found three genes targeted by miR169 identified as pectinesterase inhibitors (Additional file 24: Table S13). Pectinesterase can convert components of the plant cell wall to pectic acid and may have a role in the cellulose biosynthetic pathway [87, 88]. Correspondingly, the chemical and structural features of switchgrass cell walls can have a significant effect on biofuel yields [89, 90]. Therefore, the regulation of miR169 and its precursor sequence may contribute to biofuel yields in switchgrass. miR169 might also target two genes that code for stress-responsive proteins, indicating that miR169 and its precursor might also regulate stress tolerance in switchgrass. The targets of miR171 were primarily predicted to be the SCR genes (Additional file 24: Table S13), which are expressed in hypocotyl, inflorescence, and stem tip tissues [91]. Recent studies have revealed that mutations in SCR genes can cause proliferation of bundle sheath cells and abnormal differentiation of bundle sheath chloroplasts in maize [92]. In addition, suppression of *LaSCR1* was shown to decrease root numbers in transformed roots of white lupin (*Lupinus albus*) and *Medicago truncatula* [93]. Therefore, miR171 and its precursor (TCONS\_00155383) are likely to be involved in stem and root development in switchgrass.

## Conclusion

This study utilized a combination of omics techniques to conduct genome-wide single-base-resolution methylome, lncRNA, miRNA, and mRNA sequencing in switchgrass. The results presented here support that siRNAs positively regulate DNA methylation at mCHH sites and that DNA methylation may interfere with gene and lncRNA expression in the polyploid switchgrass genome. In addition, a total of 96 genes were identified that are covered by DMRs in leaf and root tissues. These genes are believed to be involved in stimulus-related GO processes and 79 of them were hypomethylated in the root tissue. We also identified 9244 novel lncRNAs in switchgrass and predicted two lncRNA precursors of miRNAs that may function in cellulose biosynthesis, stress regulation, and stem and root development. Overall, the successful DNA methylome and transcriptome sequencing of switchgrass presented in this study provides a reference for other highly heterozygous Poaceae grasses with similar characteristics. These results could also serve as genomic resources for identifying further methylation patterns and additional non-coding RNAs analysis in switchgrass.

## Experimental procedures

### Plant materials

The switchgrass cv. Alamo was propagated through tiller-splitting and planted in pots (0.25 m diameter × 0.4 m

tall) containing 1500 g soil (pH 5.37, 1.26% organic qualitative content, 98.38 mg/kg N, 4.48 mg/kg P, and 328.22 mg/kg K). The plants were maintained in a greenhouse (Wenjiang, Sichuan, China) at 28 °C/20 °C (day/night) with a photoperiod of 16 h/8 h (day/night). Four months after transplanting and approximately 1 week before the emergence of flower primordium (E5 stage), the flag leaves and roots of six individual plants were collected [94]. The leaf and root tissues were frozen in liquid nitrogen and stored at −80 °C before total DNAs and RNAs were extracted.

### Methylation data analysis

#### DNA extraction

The flag leaves and roots were pooled from six random individuals with equal masses for each one and were ground to a fine powder in liquid nitrogen. Genomic DNA was extracted using a plant genomic DNA kit (Tiangen, China) following the manufacturer's instructions. The DNA integrity and concentration were measured by agarose gel electrophoresis and NanoDrop spectrophotometer, respectively.

#### Library construction and sequencing

Bisulfite sequencing libraries were prepared using the TruSeq Nano DNA LT kit (Illumina, San Diego, CA, USA) as described in Du's study [95]. The genomic DNAs were then fragmented into 100–300 bp by sonication (Covaris, USA) and purified using a MiniElute PCR Purification Kit (QIAGEN, Silicon Valley Redwood City, CA, USA). The fragmented DNAs were end repaired and a single 'A' nucleotide was appended to the 3' end of each fragment. After ligating the DNAs to the sequencing adapters, the genomic fragments were bisulfite converted via a Methylation-Gold kit (ZYMO, Murphy Ave. Irvine, CA, USA). The converted DNA fragments were PCR amplified and sequenced as paired-end reads using the Illumina HiSeq™ 4000 platform by the Gene Denovo Biotechnology Co. (Guangzhou, China).

#### Data filtering

The raw reads generated from the Illumina HiSeq™ 4000 were filtered to get high-quality reads using fastq-mcf (v1.04) tool [96] according to the following principles: (1) reads with more than 10% of unknown nucleotides (N) were removed, and (2) reads with more than 40% of low-quality ( $Q$ -value  $\leq 20$ ) bases were removed.

#### Methylation level analysis

We mapped the clean reads to the switchgrass reference genome ([https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Pvirgatum\\_er;V4.1](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvirgatum_er;V4.1)) using BSMAP software (v 2.90) [97] with the default

parameters. The effective coverage of cytosine was defined as the number of coverage reads on cytosine  $\geq 1$ . The effective coverage rate of cytosine was calculated based on the ratio of the number of effective coverage of cytosine per all cytosines in one specific region. The mC was called from these effective coverage cytosines, and the methylation level was calculated using a custom script as part of the BSMAP package based on the following ratio:  $(mC)/(mC + non-mC)$ . This was calculated for the whole switchgrass genome, every chromosome, and for all genomic regions for each of the three methylation contexts (CG, CHG, and CHH). The methylation patterns for the gene body, transposable elements, and flanking upstream and downstream 2-kb regions were plotted using R project software (<http://miyoviqo.tha.im/>) according to the average methylation levels for each 100-bp interval. ANOVA analysis was applied to test the significance of the average methylation levels of both genic and TE regions between leaf and root tissues [98].

#### **Differentially methylated regions analysis**

Differentially methylated regions (DMRs) between leaf and root tissues of switchgrass for CG, CHG, and CHH were identified based on the following criteria:

(1) The length of each DMR region was between 40 bp and 10 kb; (2) the distance between adjacent methylated sites was  $< 200$ -bp; (3) more than ten reads needed to be present for each cytosine, and more than four reads were need for coverage for each methylated cytosine; (4) more than five methylated cytosines must be present in at least one sample; (5) the fold change of the average methylation level was  $> 2$ ; (6) Pearson's chi-square test ( $\chi^2$ ) value was  $p \leq 0.05$ .

#### **Enrichment analysis of DMR-related genes**

We sorted the putative DMRs covering the gene body regions, the 2-kb flanking regions of genes, and the TEs and conducted GO enrichment analysis via topGO package in R [99] for the DMR-related genes using a hypergeometric test with a corrected  $p \leq 0.05$ .

#### **RNA-seq and data analysis**

The six individual plants were pooled into two independent biological replicates, and each replicate contained three individuals. Total RNAs were isolated from the flag leaves of the two replicates using TRIzol reagent (Invitrogen, Carlsbad, NM, USA) according to the manufacturer's instructions. rRNAs were removed from the samples, retaining only mRNAs and non-coding RNAs. The sequencing libraries were built following Illumina's standard protocol for RNA-seq library construction as previously described [20], and the libraries were sequenced on an Illumina HiSeq™ 4000 RNA-sequencer

at Gene Denovo Biotechnology Co. (Guangzhou, China). To obtain quality reads, the raw data were filtered by removing sequences containing adaptors, low-quality reads ( $Q$ -value  $\leq 20$ ), and reads containing more than 10% of unknown nucleotides (N) using fastq-mcf (v1.04) tool [96]. Bowtie2 [100] was used to map the reads against the ribosome RNA (rRNA) database to remove rRNA mapped reads. The remaining reads were used for further transcriptome analysis and were mapped to the switchgrass reference genome ([https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Pvirgatum\\_er;V4.1](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvirgatum_er;V4.1)) using TopHat2 (v 2.0.3.12) with default parameters [101]. Reconstruction of the transcripts was conducted using both Cufflinks [102] and TopHat2. The reconstructed transcripts were re-aligned to the switchgrass reference genome and clustered into 12 categories using Cuffcompare [102]. The abundance of each transcript was quantified using RSEM (v 1.2.19) [103] and transcript expression levels were normalized using RPKM (Reads per kb per Million reads).

#### **LncRNA data analysis**

##### **LncRNA prediction**

Two programs: CNCI (v2) [104] and CPC [105] (<http://cpc.cbi.pku.edu.cn/>) were used to evaluate the protein-coding potential of new transcripts with default parameters. Transcripts that had both CPC and CNCI scores less than 0 were chosen as long non-coding RNAs.

##### **LncRNA family analysis**

Rfam divides non-coding RNAs into families based on their evolution from a common ancestor. Producing multiple sequence alignments of these families can provide insight into their structure and function, similar to the case of protein families. To better annotate lncRNAs at the evolution level, the software Infernal [106] (<http://eddylab.org/infernal/>) was used in the sequence alignment and lncRNAs were classified by sequence consensus and secondary structures.

##### **miRNA precursor prediction**

To find potential miRNA precursors, lncRNAs were aligned to miRBase (v 21) using BLAST (v 2.2.25), and those with greater than 90% identity were selected. Additionally, the software miRPara (v 6.2) [107] was employed to predict miRNA precursors using default parameters.

##### **Functional group analysis**

GO analysis and KEGG analysis were applied to determine the biological roles of the targets genes of identified lncRNAs in switchgrass. GO terms were assigned based on the Gene Ontology (GO) database [108] using Goseq

R package [109] and the latest KEGG (Kyoto Encyclopedia of Genes and Genomes) database (<http://www.genome.jp/kegg/>) using KOBAS [110]. The  $p$  value (Hypergeometric-P value) cut-off was set to 0.05.

### Small RNA-seq data analysis

RNAs extracted from the flag leaves of the two biological replicates were used for small RNA-sequencing. Small RNA molecules with a size range of 18–30 nt were enriched by polyacrylamide gel electrophoresis (PAGE) and libraries were built using the NEBNext multiplex small RNA library prep set (NEB#E7300, Ipswich, MA, England) based on the manufacturer's protocol. The small RNAs were then sequenced using an Illumina HiSeq 2500 by Gene Denovo Biotechnology Co. (Guangzhou, China). Raw reads were filtered by removing the following: reads without 3' adapters, reads containing 5' adapters, reads containing 3' and 5' adapters but no small RNA fragment between them, reads containing poly A sequences in the small RNA fragment, and reads shorter than 18-nt (not including adapters). The high-quality sequences were then BLAST (v 2.2.25) [111] against the small RNAs in the GenBank database (Release 209.0) and Rfam database (v 11.0) to remove rRNAs, snRNAs, snoRNAs, scRNAs, and tRNAs. High-quality, clean small RNA sequences were then aligned with the switchgrass reference genome using Bowtie (v 1.1.2) [112] to remove those that mapped to exons or introns. The known switchgrass miRNAs were identified by searching the clean sequences against the miRBase database (Release 21). In addition, the 24-nt reads that did not match any miRNAs were retained and used as siRNAs for subsequent analyses. The siRNA sequences were normalized by calculating the siRNA reads per million based on the total abundance of genome-matched small RNA reads.

### Statistics analysis

The relationships between methylation and mRNAs, siRNAs, and lncRNAs were all analyzed using Spearman and regression analysis. For Spearman analysis, the evaluation of correlation was based on the following criteria ([www.statstutor.ac.uk/resources/uploaded/spearmans.pdf](http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf)): (1)  $\rho > 0$  equated to a positive correlation, (2)  $\rho < 0$  signified a negative correlation, (3) very weak was indicated by the absolute value of  $\rho < 0.2$ , (4) weak was indicated by  $0.2 \leq$  the absolute value of  $\rho < 0.4$ , (5) moderate was determined by  $0.4 \leq$  the absolute value of  $\rho < 0.6$ , (6) strong was determined by  $0.6 \leq$  the absolute value of  $\rho < 0.79$ , and (7) very strong was indicated by  $0.8 \leq$  the absolute value of  $\rho < 1.0$ . The relationships of mRNA, siRNA and lncRNA with methylation were evaluated by modeling gene expression as a function of the number of

methylated cytosines in the body of the gene and both flanking regions [44]. For mRNA and lncRNA we used hurdle regression analysis. These two-part models handle read counts with excessive amounts of zeroes (i.e. unexpressed genes) by specifying process for zeroes. This process was modeled with the logistic distribution. The relationship between siRNA and methylation was modeled solely with the negative binomial distribution because zeroes were not overrepresented in this case. The prediction power of all regression models was then tested using a tenfold cross-validation method [45]. The input data were randomly and equally divided into ten subsets (tenfold) for the validation. The cross-validate  $R^2$  (also known as  $Q^2$ ) values were calculated by a formula:  $Q^2 = 1 - \text{PRESS (predictive error sum of squares)}/\text{TSS (total sum of squares)}$ .

### Additional files

**Additional file 1: Table S1.** Information on bisulfite sequencing data in switchgrass.

**Additional file 2: Data S1.** Effective coverage rate of cytosine in the chromosome, gene, and repeat regions in both leaf and root tissues of switchgrass.

**Additional file 3: Figure S1.** Sequencing depth and saturation analysis of in leaf (a) and root (b) tissues of switchgrass. The x-axis represents the sequencing depth. The y axis represents the percentage for cytosine with a sequencing depth over a specific depth of the whole genome cytosine.

**Additional file 4: Figure S2** Chromosome distribution for root (a) and leaf (b). Methylation level in 80-kb windows throughout chromosomes in the leaf tissue of switchgrass. The red line means '−' strand, and the blue line means '+' strand.

**Additional file 5: Table S2.** The levels of mCG, mCHG, and mCHH and genome size in genomes of different species.

**Additional file 6: Figure S3.** Correlation between genome sizes and methylation levels for different species as *Arabidopsis*, *B. rapa*, *G. max*, *M. truncatula*, *P. trichocarpa*, *Z. mays*, *O. sativa*, *B. distachyum*, and *P. virgatum* (leaf tissue) used in Figure 1e.

**Additional file 7: Figure S4.** Methylation levels of Class I and Class II TEs in the leaf (a) and root (b) of switchgrass.

**Additional file 8: Table S3.** Comparison of methylation levels between class I and class II transposons in leaf and root tissues.

**Additional file 9: Table S4.** Comparison of methylation levels in genic and TE regions between switchgrass leaf and root tissues.

**Additional file 10: Table S5.** Comparison of methylation levels in different TE types between switchgrass leaf and root tissues.

**Additional file 11: Data S2.** GO annotation of genes covered by the DMRs between switchgrass leaf and root tissues.

**Additional file 12: Data S3.** The methylation levels of 96 genes covered by DMRs and involved in stimulus related GO processes in both leaf and root tissues of switchgrass.

**Additional file 13: Figure S5.** Molecular function for genes covered by the CHH DMRs in the GO annotation in gene upstream (a), body (b), and downstream (c). The colored boxes mean the adjust  $p < 0.05$ , and the more deeper color, the less adjust  $p$  values for the boxes.

**Additional file 14: Data S4.** The CHH methylation of DMR covered genes annotated in the oxidation-reduction process and oxidoreductase in both leaf and root tissues.



**Additional file 15: Table S6.** The number of mRNA and lncRNA in the four expression levels.

**Additional file 16: Table S7.** Correlation analysis between methylation levels and gene expression in different genic regions.

**Additional file 17: Table S8.** Regression analysis and 10-fold cross-validation between DNA methylation and gene expression based on zero hurdle model.

**Additional file 18: Table S8.** Length distribution of siRNAs (from 20 to 25 nucleotides) in switchgrass for two biological repeats including Alamo\_1 (a) and Alamo\_2 (b).

**Additional file 19: Table S9.** Comparison of methylation levels between siRNA uniquely mapping region and without that region.

**Additional file 20: Table S10.** The correlation analysis between gene methylation and siRNA expression.

**Additional file 21: Table S11.** Regression analysis and 10-fold cross-validation between DNA methylation and siRNA expression based on negative binomial model.

**Additional file 22: Figure S7.** Proportion of six different types of lncRNA in switchgrass.

**Additional file 23: Table S12.** Prediction of lncRNA families in switchgrass.

**Additional file 24: Table S13.** Annotation of targets of miR169 and miR171.

**Additional file 25: Table S14.** The correlation analysis between methylation levels and lncRNA expression.

**Additional file 26: Table S15.** Regression analysis and 10-fold cross-validation between DNA methylation and lncRNA expression based on zero hurdle model.

**Additional file 27: Figure S8.** siRNA levels (per bp per million reads) in TEs relative to the distance from the nearest gene.

## Abbreviations

TEs: transposable elements; MiRNAs: microRNAs; lncRNAs: long non-coding RNAs; RdDM: RNA-dependent DNA methylation; DMRs: differentially methylated regions; mCs: methylcytosines; TSS: transcription starting sites; ROIs: reactive oxygen intermediates; DRM2: domains rearranged methylase 2; SAUR: Auxin upregulated RNA; MLO: mildew resistance locus O (MLO); FAD: flavin adenine dinucleotide; BSD: DOS2-like protein.

## Authors' contributions

HY and LH designed research studies; HY conducted experiments, acquired data, and analyzed data; HY, BZ, AB, and LH wrote the manuscript; BX, TPF, CW, PC, JC, TH, CC, and XZ revised the manuscript. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Department of Grassland Science, Animal Science and Technology College, Sichuan Agricultural University, Chengdu 611130, China. <sup>2</sup> Department of Horticulture, Virginia Tech, Blacksburg, VA 24061, USA. <sup>3</sup> College of Grassland Science, Nanjing Agricultural University, Nanjing 210095, China. <sup>4</sup> Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996, USA. <sup>5</sup> Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Blacksburg, VA 24061, USA.

## Acknowledgements

The authors would like to thank the Guangzhou Gene denovo Biotechnology Co., Ltd for assisting in the sequencing and data analysis.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The raw data of mRNA, microRNA, lncRNA, and methylation will be uploaded to NCBI: SRP137891.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Funding

This work was supported by the Modern Agro-industry Technology Research System (CARS-34), the Sichuan International Cooperation Projects (2017HH0071) and the Sichuan Province Breeding Research grant (2016NYZ0039), the project was also partially supported by USDA-NIFA Grant Number 2011-67009-30133 to BZ, and the Virginia Agricultural Experiment Station (VA135872), and the National High-Technology Research and Development Program (863 Program) of China (No. 2012AA101801-02).

## Publisher's Note

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

Received: 20 March 2018 Accepted: 10 July 2018

Published online: 24 July 2018

## References

- Mclaughlin SB, Kszos LA. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenergy*. 2005;28:515–35.
- Karp A, Shield I. Bioenergy from plants and the sustainable yield challenge. *New Phytol*. 2008;179:15–32.
- Hopkins AA, Taliaferro CM, Murphy CD, Christian D. Chromosome number and nuclear DNA content of several switchgrass populations. *Crop Sci*. 1996;36:1192–5.
- Parrish DJ, Fike JH. The biology and agronomy of switchgrass for biofuels. *Crit Rev Plant Sci*. 2005;24:423–59.
- Casler MD. Switchgrass breeding, genetics, and genomics. Berlin: Springer; 2012.
- Okada M, Lanzatella C, Saha MC, Bouton J, Wu RL, Tobias CM. Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. *Genetics*. 2010;185:745–60.
- Finnegan EJ, Peacock WJ, Dennis ES. DNA methylation, a key regulator of plant development and other processes. *Curr Opin Genet Dev*. 2000;10:217–23.
- Pooggin MM. How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? *Int J Mol Sci*. 2013;14:15233–59.
- Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. *Nature*. 1993;366:362–5.
- Ibarra CA, Feng XQ, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*. 2012;337:1360–4.
- Saze H, Tsugane K, Kanno T, Nishimura T. DNA methylation in plants: relationship to small RNAs and histone modifications, and functions in transposon inactivation. *Plant Cell Physiol*. 2012;53:766–84.
- Rival A, Ilbert P, Labeyrie A, Torres E, Doubeau S, Personne A, Dussert S, Beulé T, Durand-Gasselin T, Tregear JW. Variations in genomic DNA methylation during the long-term in vitro proliferation of oil palm embryogenic suspension cultures. *Plant Cell Rep*. 2013;32:359–68.
- Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 2010;11:204–20.
- Bartel DP. MicroRNAs genomics, biogenesis, mechanism, and function. *Cell*. 2004;116:281–97.

15. Zhang BH, Pan XP, Cobb GP, Anderson TA. Plant microRNA: a small regulatory molecule with big impact. *Dev Biol.* 2006;289:3–16.
16. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 2009;23:1494–504.
17. Ausin I, Feng S, Yu C, Liu W, Kuo HY, Jacobsen EL, Zhai J, Gallegobartolome J, Wang L, Egertsdotter UDNA. methylome of the 20-gigabase Norway spruce genome. *P Natl Acad Sci USA.* 2016;113:E8106–13.
18. Gent JJ, Ellis NA, Guo L, Harkess A, Yao YY, Zhang XY, Dawe RK. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res.* 2013;23:628–37.
19. Song Q, Guan X, Chen ZJ. Dynamic roles for small RNAs and DNA methylation during ovule and fiber development in allotetraploid cotton. *PLoS Genet.* 2015;11:e1005724.
20. Xu W, Yang T, Dong X, Li DZ, Liu A. Genomic DNA methylation analyses reveal the distinct profiles in castor bean seeds with persistent endosperms. *Plant Physiol.* 2016;171:1242–58.
21. Ruscio AD, Ebralidze AK, Benoukrat T, Amabile G, Goff LA, Terragni J, Figueroa ME, Pontes LDF, Alberichjorda M, Zhang P. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature.* 2013;503:371–6.
22. Zhou JC, Yang LH, Zhong TY, Mueller M, Yi M, Na Z, Xie JK, Giang K, Chung H, Sun XG. H19 lncRNA alters DNA methylation genome wide by regulating S-adenosylhomocysteine hydrolase. *Nat Commun.* 2015;6:10221.
23. Pck A, Dennis ES, Wang MB. Analysis of argonaute 4-associated long non-coding RNA in *Arabidopsis thaliana* sheds novel insights into gene regulation through RNA-directed DNA methylation. *Genes.* 2017;8:198.
24. Song YP, Dong C, Tian M, Zhang DQ. Stable methylation of a non-coding RNA gene regulates gene expression in response to abiotic stress in *Populus simonii*. *J Exp Bot.* 2015;67:1477–92.
25. Still CJ, Berry JA, Collatz GJ, Defries RS. Global distribution of C3 and C4 vegetation: carbon cycle implications. *Global Biogeochem Cycles.* 2003;17:6–16.
26. Xiu L, Zhu JD, Hu FY, Ge S, Ye MZ, Xiang H, Zhang GJ, Zheng XM, Zhang HY, Zhang SL. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics.* 2012;13:300.
27. Eichten SR, Stuart T, Srivastava A, Lister R, Borevitz JO. DNA methylation profiles of diverse *Brachypodium distachyon* aligns with underlying genetic diversity. *Genome Res.* 2016;26:1520–31.
28. Gardiner LJ, Quintontulloch M, Olohan L, Price J, Hall N, Hall A. A genome-wide survey of DNA methylation in hexaploid wheat. *Genome Biol.* 2015;16:273.
29. Yang H, Chang F, You C, Cui J, Zhu G, Wang L, Zheng Y, Qi J, Ma H. Whole-genome DNA methylation patterns and complex associations with gene structure and expression during flower development in *Arabidopsis*. *Plant J.* 2015;81:268–81.
30. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell.* 2006;126:1189–201.
31. Schnable PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, Liang CZ, Zhang JW, Fulton L, Graves T. The B73 maize genome: complexity, diversity, and dynamics. *Science.* 2009;326:1112–5.
32. Zhang J, Liu Y, Xia EH, Yao QY, Liu XD, Gao LZ. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proc Natl Acad Sci USA.* 2015;112:E7022.
33. Schmidt T. LINES, SINES and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Mol Biol.* 1999;40:903–10.
34. Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshedwilliams L, Fischer RL, Zilberman D. Genome-wide demethylation of *Arabidopsis* endosperm. *Science.* 2009;324:1451–4.
35. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature.* 2011;480:245–9.
36. Song QX, Lu X, Li QT, Chen H, Hu XY, Ma B, Zhang WK, Chen SY, Zhang JS. Genome-wide analysis of DNA methylation in soybean. *Mol Plant.* 2013;6:1961–74.
37. Wang QS, Dong C, Li T, Li PW, Song YP, Chen JH, Quan MY, Zhou DL, Zhang DQ. The role of DNA methylation in xylogenesis in different tissues of poplar. *Front Plant Sci.* 2016;7:1003.
38. Cao DH, Gao X, Liu J, Wang XP, Geng SJ, Yang CW, Liu B, Shi DC. Root-specific DNA methylation in *Chloris virgata*, a natural alkaline-resistant halophyte, in response to salt and alkaline stresses. *Plant Mol Biol Rep.* 2012;30:1102–9.
39. Ekanayake IJ. Inheritance of root characters and their relations to drought resistance in rice. *Crop Sci.* 1985;25:927–33.
40. Hurley MB, Rowarth JS. Resistance to root growth and changes in the concentrations of ABA within the root and xylem sap during root-restriction stress. *J Exp Bot.* 1999;50:799–804.
41. Yan SP, Tang ZC, Su WA, Sun WN. Proteomic analysis of salt stress-responsive proteins in rice root. *Proteomics.* 2005;5:235–44.
42. Lu XD, Wang WX, Ren W, Chai ZG, Guo WZ, Chen RM, Wang L, Zhao J, Lang ZH, Fan YL. Genome-wide epigenetic regulation of gene transcription in maize seeds. *PLoS ONE.* 2015;10:e0139582.
43. Myers L, Sirois MJ. Spearman correlation coefficients, differences between. Hoboken: Wiley; 2006.
44. Hu MC, Pavlicova M, Nunes EV. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *Am J Drug Alcohol Abuse.* 2011;37:367–75.
45. Bengio Y, Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation. *J Mach Learn Res.* 2004;5:1089–105.
46. Haag JR, Pikaard CS. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol.* 2011;12:483–92.
47. Matzke M, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet.* 2014;15:394–408.
48. Wang H, Niu QW, Wu HW, Liu J, Ye J, Yu N, Chua NH. Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J.* 2015;84:404–16.
49. Szcześniak MW, Roskiewicz W, Makalowska I. CANTATAdb: A collection of plant long non-coding RNAs. *Plant Cell Physiol.* 2016;57:e8.
50. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* 2005;33:121–4.
51. Tolosa A, Sanjuán J, Leal C, Costas J, Moltó MD, Frutos RD. Rapid evolving RNA gene HAR1A and schizophrenia. *Schizophr Res.* 2008;99:370–2.
52. Herzog LBK, Romer JT, Horn JM, Ashworth A. Xist has properties of the X-chromosome inactivation centre. *Nature.* 1997;386:272–5.
53. Fantes J, Ragge NK, Lynch SA, Mcgill NI, Collin JR, Howardpeebles PN, Hayward C, Vivian AJ, Williamson K, Van HV. Mutations in SOX2 cause anophthalmia. *Nat Genet.* 2003;33:462–3.
54. Gabory A, Ripoche MA, Le DA, Watrin F, Ziyat A, Forné T, Jammes H, Ainscough JF, Surani MA, Journot L. H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development.* 2009;136:3413–21.
55. Lan T, Chang L, Wu L, Yuan Y. Downregulation of ZEB2-AS1 decreased tumor growth and metastasis in hepatocellular carcinoma. *Mol Med Rep.* 2016;14:4606–12.
56. Qin Y, Chen WB, Xiao Y, Yu WD, Cai X, Dai M, Xu TT, Huang WL, Guo W, Deng WG. RFLP3 and CBP synergistically upregulate hTERT activity and promote lung cancer growth. *Oncotarget.* 2015;6:27130–45.
57. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 2015;43:D130.
58. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet.* 2003;33:245–54.
59. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C. The transcriptional landscape of the mammalian genome. *Science.* 2005;309:1559–63.
60. Kampa D, Cheng J, Kapranov P, Yamanaka M, Brubaker S, Cawley S, Drenkow J, Piccolboni A, Bekiranov S, Helt G. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* 2004;14:331–42.
61. Zhang Y, Chen Y. Long noncoding RNAs: new regulators in plant development. *Biochem Biophys Res Commun.* 2013;436:111–4.
62. Martinezreyna JM, Vogel KP. Incompatibility systems in switchgrass. *Crop Sci.* 2002;42:1800–5.

63. Sharma MK, Sharma R, Cao PJ, Jenkins J, Bartley LE, Qualls M, Grimwood J, Schmutz J, Rokhsar DS, Ronald PC. A genome-wide survey of switchgrass genome structure and organization. *PLoS ONE*. 2012;7:e33892.
64. Dworkin M, Xie S, Saha M, Thimmapuram J, Kalavacharla VK. Analyses of methylomes of upland and lowland switchgrass (*Panicum virgatum*) ecotypes using MeDIP-seq and BS-seq. *BMC Genomics*. 2017;18:851.
65. Feng SH, Cokus SJ, Zhang XY, Chen P, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA*. 2010;107:8689–94.
66. Wang H, Beyene G, Zhai JX, Feng SH, Fahlgren N, Taylor N, Bart R, Carrington JC, Jacobsen SE, Ausin I. CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proc Natl Acad Sci USA*. 2015;112:13729–34.
67. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*. 2010;328:916–9.
68. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol*. 2016;17:194.
69. Zakrzewski F, Schmidt M, Van Lijsebettens M, Schmidt T. DNA methylation of retrotransposons, DNA transposons and genes in sugar beet (*Beta vulgaris* L.). *Plant J*. 2017;90:1156–75.
70. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*. 2008;9:465–76.
71. Mann JJ, Barney JN, Kyser GB, Ditomaso JM. Root system dynamics of *Miscanthus x giganteus* and *Panicum virgatum* in response to rainfed and irrigated conditions in California. *Bioenergy Res*. 2013;6:678–87.
72. Wang SK, Bai YH, Shen CJ, Wu YR, Zhang SN, Jiang DA, Guilfoyle TJ, Ming C, Qi YH. Auxin-related gene families in abiotic stress response in *Sorghum bicolor*. *Funct Integr Genomics*. 2010;10:533–46.
73. Kim MC, Panstruga R, Elliott C, Müller J, Devoto A, Yoon HW, Park HC, Cho MJ, Schulze-Lefert P. Calmodulin interacts with MLO protein to regulate defence against mildew in barley. *Nature*. 2002;416:447–51.
74. Müller M, Carell T. Structural biology of DNA photolyases and cryptochromes. *Curr Opin Struct Biol*. 2009;19:277–85.
75. Nutan KK, Kushwaha HR, Singlapareek SL, Pareek A. Transcription dynamics of Saltol QTL localized genes encoding transcription factors, reveals their differential regulation in contrasting genotypes of rice. *Funct Integr Genom*. 2017;17:69–83.
76. Mittler R. Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci*. 2002;7:405–10.
77. Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*. 2009;461:1130–4.
78. Wang YH, Warren JT. Mutations in retrotransposon AtCOPIA4 compromises resistance to *Hyaloperonospora parasitica* in *Arabidopsis thaliana*. *Genet Mol Biol*. 2010;33:135–40.
79. Zheng BL, Wang ZM, Li SB, Yu B, Liu JY, Chen XM. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in *Arabidopsis*. *Gene Dev*. 2009;23:2850–60.
80. Böhmendorfer G, Rowley MJ, Zhu Y, Amies I, Wierzbicki AT. RNA directed DNA methylation requires stepwise binding of silencing factors to long noncoding RNA. *Plant J*. 2014;79:181–91.
81. Yang XJ, Han H, Carvalho DDD, Lay FD, Jones PA, Liang GN. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell*. 2014;26:577–90.
82. Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE, Henikoff SDNA. Methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr Biol*. 2005;15:154–9.
83. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet*. 2007;39:61–9.
84. Miguel C, Marum L. An epigenetic view of plant cells cultured in vitro: somaclonal variation and beyond. *J of Exp Bot*. 2011;62:3713–25.
85. Cai XZ, Cullen BR. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA*. 2007;13:313–6.
86. Uva P, Sacco LD, Corno MD, Baldassarre A, Sestili P, Orsini M, Palma A, Gessani S, Masotti A. Rat mir-155 generated from the lncRNA Bic is 'hidden' in the alternate genomic assembly and reveals the existence of novel mammalian miRNAs and clusters. *RNA*. 2013;19:365–79.
87. Wu MF, Tseng KC, Huang T, Chang HM. Pectinesterase inhibitor in rubbery banana (*Musa sapientum* L.). *J Food Sci*. 2002;67:1337–40.
88. Zykwska AW, Ralet MC, Garnier CD, Thibault JF. Evidence for in vitro binding of pectin side chains to cellulose. *Plant Physiol*. 2005;139:397–407.
89. Demartini JD, Pattathil S, Miller JS, Li HJ, Hahn MG, Wyman CE. Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy Environ Sci*. 2013;6:898–909.
90. Sarath G, Dien BS, Saathoff AJ, Vogel KP, Mitchell RB, Chen H. Ethanol yields and cell wall properties in divergently bred switchgrass genotypes. *Bioresour Technol*. 2011;102:9579–85.
91. Wysocka-Diller JW, Helariutta Y, Fukaki H, Malamy JE, Benfey PN. Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot. *Development*. 2000;127:595–603.
92. Slewinski TL, Anderson AA, Zhang CT, Turgeon R. Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol*. 2012;53:2030–7.
93. Sbabou L, Bucciarelli B, Miller S, Liu J, Berhada F, Filalimaltoug A, Allan D, Vance C. Molecular analysis of SCARECROW genes expressed in white lupin cluster roots. *J Exp Bot*. 2010;61:1351–63.
94. Hardin CF, Fu CX, Hisano H, Xiao XR, Shen H, Stewart CN, Parrott W, Dixon RA, Wang ZY. Standardization of switchgrass sample collection for cell wall and biomass trait analysis. *Bioenergy Res*. 2013;6:755–62.
95. Du JM, Johnson LM, Groth M, Feng SH, Hale CJ, Li SS, Vashisht AA, Gallego-bartolome J, Wohlschlegel JA, Patel DJ. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol Cell*. 2014;55:495–504.
96. Aronesty E. Comparison of sequencing utility programs. *Op Bioinform J*. 2013;7:1–8.
97. Xi YX, Li W. BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinform*. 2009;10:232.
98. Cuevas A, Febrero M, Fraiman R. An anova test for functional data. *Comput Stat Data Anal*. 2007;47:111–22.
99. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package version 2.26.0. 2016. <http://www.bioconductor.org/packages/release/bioc/html/topGO.html>. Accessed 5 Apr 2017
100. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
101. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
102. Trapnell C, Roberts A, Goff LA, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc*. 2012;7:562–78.
103. Li B, Dewey CN. RSEM accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform*. 2011;12:323.
104. Sun L, Luo HT, Bu DC, Zhao GG, Yu KT, Zhang CH, Liu YN, Chen RS, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41:e166.
105. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei LP, Gao G. CPC assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35:W345–9.
106. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29:2933–5.
107. Wu YG, Wei B, Liu HZ, Li TX, Rayner S. MirPara: a SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. *BMC Bioinform*. 2011;12:107.
108. Harris MA, Clark JI, Ireland A, Lomax J, Ashburner M, Foulger RE, Eilbeck K, Lewis SE, Marshall B, Mungall CJ. The gene ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32:D258–61.
109. Young MD, Wakefield MJ, Smyth GK, Alicia O. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. 2010;11:1–12.
110. Mao XZ, Cai T, Olyarchuk JG, Wei LP. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*. 2005;21:3787–93.
111. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
112. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.