Please cite this article as: Xiaoyu Zhang et al., Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*, CELL (2006), doi:10.1016/j.cell.2006.08.003

Resource

Cell

Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*

Xiaoyu Zhang,[1,5] Junshi Yazaki,[3,5] Ambika Sundaresan,[3,5] Shawn Cokus,[1,5] Simon W.-L. Chan,[1,6] Huaming Chen,[4] Ian R. Henderson,[1] Paul Shinn,[4] Matteo Pellegrini,[1] Steve E. Jacobsen,[1,2,*] and Joseph R. Ecker[3,4,*]

[1] Department of Molecular, Cell and Developmental Biology
[2] Howard Hughes Medical Institute
University of California, Los Angeles, Los Angeles, CA 90095, USA
[3] Plant Biology Laboratory
[4] Genomic Analysis Laboratory
The Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[5] These authors contributed equally to this work.
[6] Present address: Section of Plant Biology, University of California, Davis, Davis, CA 95616, USA.
*Contact: jacobsen@ucla.edu (S.E.J.); ecker@salk.edu (J.R.E.)
DOI 10.1016/j.cell.2006.08.003

SUMMARY

Cytosine methylation is important for transposon silencing and epigenetic regulation of endogenous genes, although the extent to which this DNA modification functions to regulate the genome is still unknown. Here we report the first comprehensive DNA methylation map of an entire genome, at 35 base pair resolution, using the flowering plant *Arabidopsis thaliana* as a model. We find that pericentromeric heterochromatin, repetitive sequences, and regions producing small interfering RNAs are heavily methylated. Unexpectedly, over one-third of expressed genes contain methylation within transcribed regions, whereas only ~5% of genes show methylation within promoter regions. Interestingly, genes methylated in transcribed regions are highly expressed and constitutively active, whereas promoter-methylated genes show a greater degree of tissue-specific expression. Whole-genome tiling-array transcriptional profiling of DNA methyltransferase null mutants identified hundreds of genes and intergenic noncoding RNAs with altered expression levels, many of which may be epigenetically controlled by DNA methylation.

INTRODUCTION

Cytosine DNA methylation is a conserved epigenetic silencing mechanism involved in many important biological processes, including defense against transposon proliferation, control of genomic imprinting, and regulation of gene expression (Bird, 2002; Goll and Bestor, 2005). In mammals, the de novo methyltransferases DNMT3a/b and the maintenance methyltransferase DNMT1 are responsible for DNA methylation, which occurs primarily at CG dinucleotides (Goll and Bestor, 2005). In plants, the DNMT1 homolog MET1 maintains CG methylation, while the DNMT3a/b homologs DRM1/2 and the plant-specific methyltransferase CMT3 are responsible for methylation at non-CG sites (Chan et al., 2005).

DNA methylation is critically important for normal development in both animals and plants; null mutations in mouse DNMT1 or DNMT3a/b result in embryonic lethality, and both *met1* and *drm1 drm2 cmt3* triple mutants exhibit developmental abnormalities in *Arabidopsis* (Chan et al., 2005, 2006; Goll and Bestor, 2005). An unbiased genome-wide identification and functional analysis of sites of DNA methylation should greatly broaden our understanding of how DNA methylation and gene expression are correlated on a global scale. However, although the detection of DNA methylation at individual loci (e.g., by bisulfite sequencing) is relatively straightforward, doing so in a high-throughput manner to explore the methylated fraction of a complete eukaryotic genome remains technically challenging.

A number of strategies for high-throughput detection of DNA methylation have been recently described and have led to important discoveries. They can be broadly divided into two types: sequencing-based and microarray-based. Sequencing-based analyses have included the direct sequencing of methylated DNA fragments isolated by affinity purification (hundreds of clones) (Selker et al., 2003), PCR amplification products from bisulfite-treated DNA (hundreds of sequences) (Rakyan et al., 2004), and genomic DNA fractionated by methylation-sensitive restriction-enzyme digestion (thousands of clones) (Rollins et al.,

2006). Recent advances in sequencing technology have made it possible to assay over 1,000 preselected CG sites in hundreds of genes (Bibikova et al., 2006). However, the cost and labor required make it difficult for sequencing-based methods to achieve whole-genome coverage. On the other hand, microarray-based approaches rely on the separation of methylated and unmethylated DNA by methods such as the differential digestion by methylation-sensitive restriction enzymes (Ching et al., 2005; Hatada et al., 2006; Lippman et al., 2004; Tran et al., 2005a, 2005b). As just one example, digestion by the enzyme McrBC (which cuts near methylcytosines) followed by hybridization to a PCR amplicon tiling array was performed to detect DNA methylation at 1 kb resolution in the repeat-rich heterochromatic knob region of *Arabidopsis* (1.5 Mb, or ~1.2% of the genome) (Lippman et al., 2004). However, these methods are limited by either sequence context (i.e., specific restriction-enzyme recognition sites) or the imprecise cutting of McrBC relative to the location of methylcytosine. Alternatively, methylated DNA can be enriched using immunoprecipitation and hybridized to microarrays to achieve a more unbiased detection (Keshet et al., 2006; Weber et al., 2005). In this case, the resolution and coverage are mostly determined by the microarray platform. For example, a BAC array has been recently used to determine the DNA methylation profile of the human genome at 80 kb resolution (Weber et al., 2005).

Previous methylation studies in *Arabidopsis* have identified a number of methylated transposons and indicated that the transposon-rich heterochromatic regions are heavily methylated (Hirochika et al., 2000; Lippman et al., 2004; Miura et al., 2001; Singer et al., 2001; The Arabidopsis Genome Initiative, 2000). Furthermore, a microarray study utilizing methylation-sensitive restriction enzymes identified CG methylation clusters (CG but not non-CG methylation) found predominantly within genes (Tran et al., 2005a). A similar approach was used to determine restriction sites that are differentially methylated in several *Arabidopsis* DNA methylation mutants (Tran et al., 2005b). However, the comprehensive identification of methylated regions in *Arabidopsis* has not been performed, and the extent and distribution of DNA methylation is unknown. Furthermore, despite the importance of DNA methylation in regulating gene expression, very few endogenous genes have been identified as being directly controlled by DNA methylation (Chan et al., 2005).

Here we describe the first genome-wide analysis of DNA methylation in the reference plant *Arabidopsis thaliana*. Using two biochemical methods in combination with whole-genome tiling microarrays, we mapped the methylated component of the *Arabidopsis* genome in wild-type as well as in the *met1* and *drm1 drm2 cmt3* triple mutant plants at 35 bp resolution. In addition, RNA expression profiles of wild-type and mutant plants were determined on both strands of the genome using the same tiling-microarray platform. We found that genes methylated in their promoters tended to be expressed in a tissue-specific manner, whereas genes methylated in their coding regions were constitutively expressed at higher levels. Analysis of the changes in DNA methylation and gene expression in the methyltransferase mutants allowed the genome-wide identification of hundreds of genes as well as many novel intergenic noncoding RNAs that appear to be regulated by DNA methylation. These results represent the first genome-wide high-resolution mapping of DNA methylation and the first systematic analysis of the role of DNA methylation in regulating gene expression for any organism.

RESULTS AND DISCUSSION

Identification of Genomic Regions Containing DNA Methylation

Methylated and unmethylated DNA was fractionated by methylcytosine immunoprecipitation (mCIP) using a monoclonal antibody that specifically recognizes methylcytosine (Keshet et al., 2006). The optimal antibody-to-DNA ratio was determined by applying varying amounts of antibody and assaying the amount of methylated DNA recovered in the elution fraction by real-time PCR. We found that 20 μg of antibody yielded maximal recovery of methylated DNA from 2 μg of input genomic DNA; all mCIP experiments described here were performed with this ratio. Under these conditions, effective recovery of methylated DNA was achieved for the *FWA* promoter region that contains extensive CG methylation (Figure 1A) (Soppe et al., 2000). Unlike vertebrate genomes, plants contain substantial amounts of non-CG DNA methylation. Importantly, the elution fraction was also enriched for the small transposon *AtSN1* (primarily methylated at non-CG sites), as well as for a region free of CG dinucleotides in *Ta3* with exclusively non-CG methylation (Figure 1A). DNA from the unbound and elution fractions (depleted and enriched for methylated DNA, respectively) was amplified and hybridized to whole-genome tiling microarrays. We utilized a single-chip tiling microarray that covers ~97% of one strand of the ~120 Mb *Arabidopsis* genomic sequence, with each of the 3.2 million 25 nt oligonucleotide probes spaced within a 35 bp window (see the Supplemental Data available with this article online).

Three independent statistical methods were used to identify sites of DNA methylation from oligonucleotide probe hybridization intensities: a two-state hidden Markov model (HMM) based on probe-level t statistics (Ji and Wong, 2005), a Wilcoxon signed-rank test (Hollander and Wolfe, 1999), and a nonparametric Kolmogorov-Smirnov test developed in this study (see Supplemental Data). Despite utilizing different statistical approaches, all three methods yielded similar results (Figure S1). The HMM method was used for all subsequent analyses, as it provided an easily interpretable bimodal distribution of posterior probability values and more defined boundaries between methylated and unmethylated regions (Figure S2).

To test the sensitivity of the mCIP-chip method, we used an independent biochemical approach to separate

Please cite this article as: Xiaoyu Zhang et al., Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*, CELL (2006), doi:10.1016/j.cell.2006.08.003

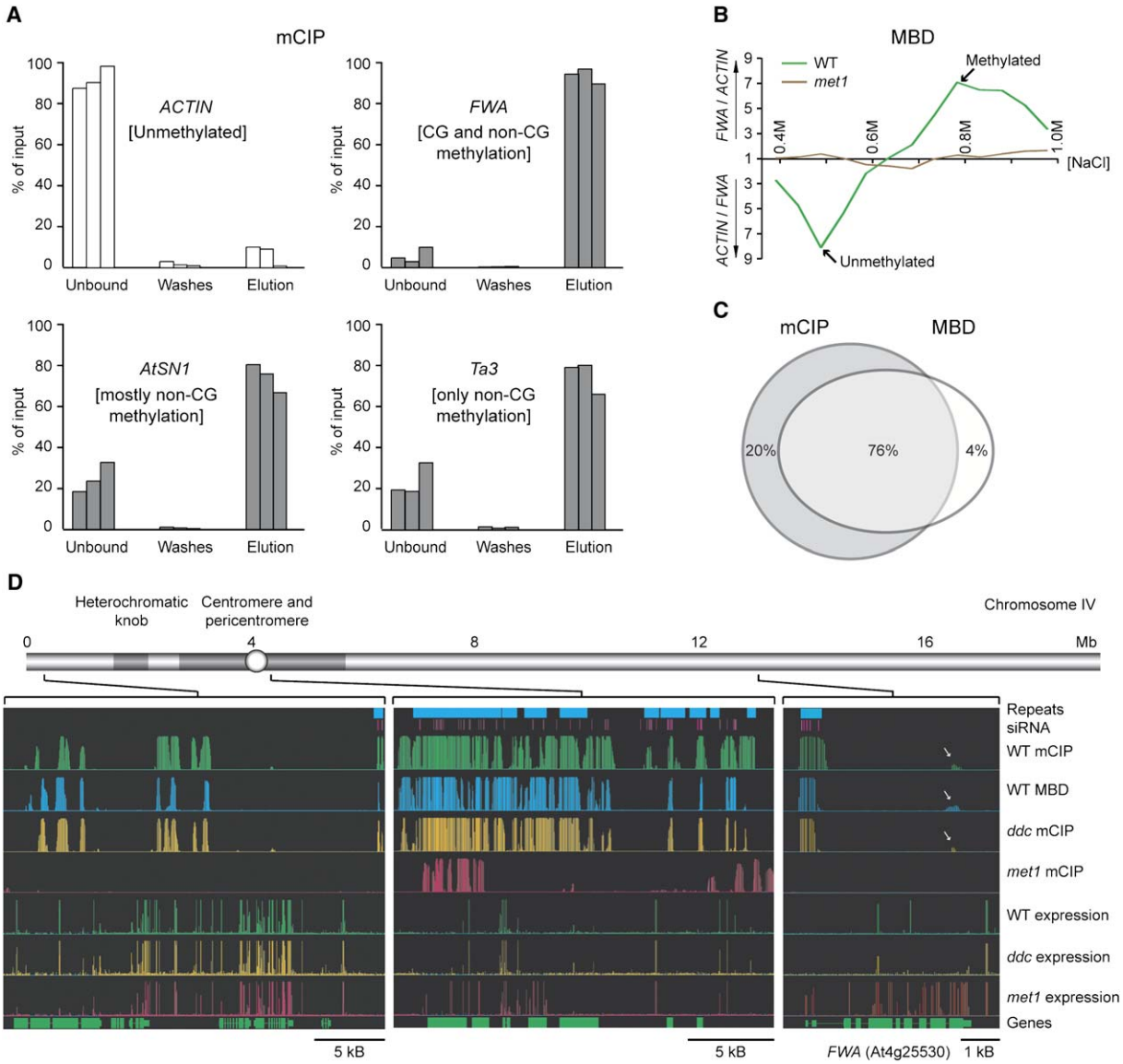


Figure 1. Biochemical Methods to Fractionate Methylated and Unmethylated DNA and Comparison of Results

(A) The mCIP method. The relative amount of the unmethylated *ACTIN-7* and the methylated *FWA*, *AtSN1*, and *Ta3* in the unbound, wash, and bound fractions as percentages of input was assayed by real-time PCR. Three biological replicates are shown.

(B) The MBD method. y axis = fold enrichment (upper half) or depletion (lower half) of *FWA* compared to *ACTIN-7* in wild-type and in the *met1* mutant that lacks CG methylation.

(C) Overlap of methylated regions detected by the mCIP (gray circle) and MBD methods (ellipse).

(D) Methylation and RNA expression patterns of a typical euchromatic region (bp 439,000–468,500; left), a repeat-rich region from pericentromeric heterochromatin (bp 4,827,000–4,849,500; middle), and the *FWA* gene (bp 13,038,000–13,043,000; right) on chromosome 4. White arrows indicate the detection of two methylated CG sites near the 3′ end of *FWA*. A schematic representation of chromosome 4 is shown on top (open circle = centromere, dark gray bars = pericentromeric heterochromatin and the heterochromatic knob, light gray bar = euchromatin). *ddc* = *drm1 drm2 cmt3* triple mutant.

methylated and unmethylated DNA. DNA methylated at CG dinucleotides was affinity purified using the methylcytosine binding domain (MBD) from the human protein MeCP2 (Cross et al., 1994) and hybridized to the same tiling array used for mCIP-chip (Figure 1B). The mCIP-chip method identified ~95% of all regions detected by the MBD-chip method and ~20% more genomic regions (Figure 1C). This difference likely reflects the fact that MBD binding requires a relatively high density of methylcytosine in CG dinucleotides (Figure S3) as well as the fact that only a subset of fractions eluted from the MBD column were hybridized to microarrays.

Using the mCIP-chip data, we constructed and annotated a set of methylation maps for each of the five chromosomes. Methylated genomic regions identified with this method were highly consistent with previous studies that have characterized individual methylated genes (Figure S4). Figure 1D shows results from three representative regions on chromosome 4, including a euchromatic region, a repeat-rich pericentromeric heterochromatic region, and the euchromatic methylated gene *FWA* (Chan et al., 2005). We performed targeted validation of a number of regions identified by mCIP-chip using genomic bisulfite sequencing (Jacobsen et al., 2000). All mCIP-chip-positive regions were confirmed to contain methylcytosines (in CG, CNG, or CNN contexts) (Figure S5). Regions previously known to be devoid of DNA methylation were also categorized as unmethylated by the mCIP-chip method (Figure S6). In addition, inclusion of the unmethylated chloroplast genome (Ngernprasirtsiri et al., 1988) on the microarray provided a large set of true negative probes to systematically estimate the level of false discovery for the mCIP-chip method and analysis processes. Within the 154,478 bp chloroplast genome, only two regions representing a mere 630 bp (0.4%) were detected as methylated. Thus, the mCIP-chip method and the analysis procedure yielded a low false-positive rate.

In the analyses of genome-wide DNA methylation patterns presented in this study, a "methylated region" was defined by combining adjacent probes with posterior probabilities over 0.5, allowing a maximal gap of 200 bp, and requiring a minimal run of 50 bp (see Supplemental Data). We purposely chose this relatively stringent cutoff to reduce falsely identified regions. Therefore, regions where methylated cytosines are very sparsely distributed may yield scores that are above background noise but do not meet this stringent threshold. As one example, we discovered that the last exon of the *FWA* gene contains two methylated CG sites, which was detected as a peak with a posterior probability of ~0.2 (Figure 1D). Thus, the analysis described below is likely a conservative estimate of the amount of genomic methylation. However, the full dynamic range of methylation patterning can be visualized in genome browsers (publicly available as described below) for any particular genomic region.

DNA Methylation Landscape of the *Arabidopsis* Genome

We detected 26,852 regions with a significant level of DNA methylation, covering 22,554,840 bp and representing ~18.9% of the entire sequenced nuclear genome (The Arabidopsis Genome Initiative, 2000). The chromosomal distribution of methylated DNA is shown in Figure 2A as the total length of methylated regions per 100 kb. As expected, we found extensive DNA methylation in the heterochromatic regions of each of the five chromosomes, including centromeres, pericentromeres, and the heterochromatic knob on chromosome 4 (Figure 2A) (Hirochika et al., 2000; Lippman et al., 2004; Miura et al., 2001; Singer et al., 2001; The Arabidopsis Genome Initiative, 2000). This distribution largely reflects the dense methylation of transposons and other repetitive sequences that are clustered in heterochromatin (Figures 2A and 2B).

A considerable amount of DNA methylation was also found in euchromatic regions, including nonrepetitive intergenic regions (~8% of total length) and genes. To perform a more detailed analysis of the level of DNA methylation in *Arabidopsis* protein-coding genes, we first classified the 30,334 annotated genes into four categories: 14,948 "known expressed" genes (genes with evidence of RNA expression and protein sequences with known or predicted function), 10,475 "unknown expressed" genes (expressed genes with sequences showing no homology to proteins of known or predicted function), 1,116 "nonexpressed genes" (computationally predicted genes with no experimental evidence for expression), and 3,811 "pseudogenes" (annotated as pseudogenes and very often containing homology to coding sequences of transposons). While numerous methylated genes were found in all four categories, the pseudogenes and nonexpressed genes showed a much higher level of methylation than known or unknown expressed genes (Figure 2C), which likely reflects an enrichment for transposons and other repeats within these two classes.

A High Level of DNA Methylation in *Arabidopsis* Genes

Mammalian genes frequently contain small transposons and repetitive DNA within their transcribed regions and are commonly DNA methylated. Although the transcribed regions of *Arabidopsis* genes are almost always transposon-free, we found an unexpectedly high level of DNA methylation in expressed genes (Figure 2C). Of the 25,423 expressed genes (excluding pseudogenes or nonexpressed genes), 15,627 (~61.5%) were entirely unmethylated ("unmethylated"), 1,331 (~5.2%) were methylated within their promoters (defined as the 200 bp region upstream of the transcription start site; "promoter-methylated"). Interestingly, a surprisingly large number of genes (8,465, ~33.3%) were methylated within transcribed regions but not within their promoter regions ("body-methylated"). Refined mapping of the position of DNA methylation within genes revealed a biased distribution toward the 3′ half, whereas promoters and the immediate 3′ flanking sequences of genes were hypomethylated (Figure 3A; see Figure S9 for similar results from an alternative mapping method) (Tran et al., 2005a). This bias against promoter and 3′-end methylation was not found in pseudogenes and nonexpressed genes (Figure 3B), which suggests that methylation at the 5′ and 3′ ends of expressed genes might be selected against, as methylation may interfere with important activities such as transcriptional initiation and termination.

DNA Methylation and Small RNAs

Small interfering RNAs (siRNAs) are known to cause RNA-directed DNA methylation (RdDM) (Chan et al., 2004, 2005; Dalmay et al., 2000; Mette et al., 2000). We carried

Please cite this article as: Xiaoyu Zhang et al., Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*, CELL (2006), doi:10.1016/j.cell.2006.08.003

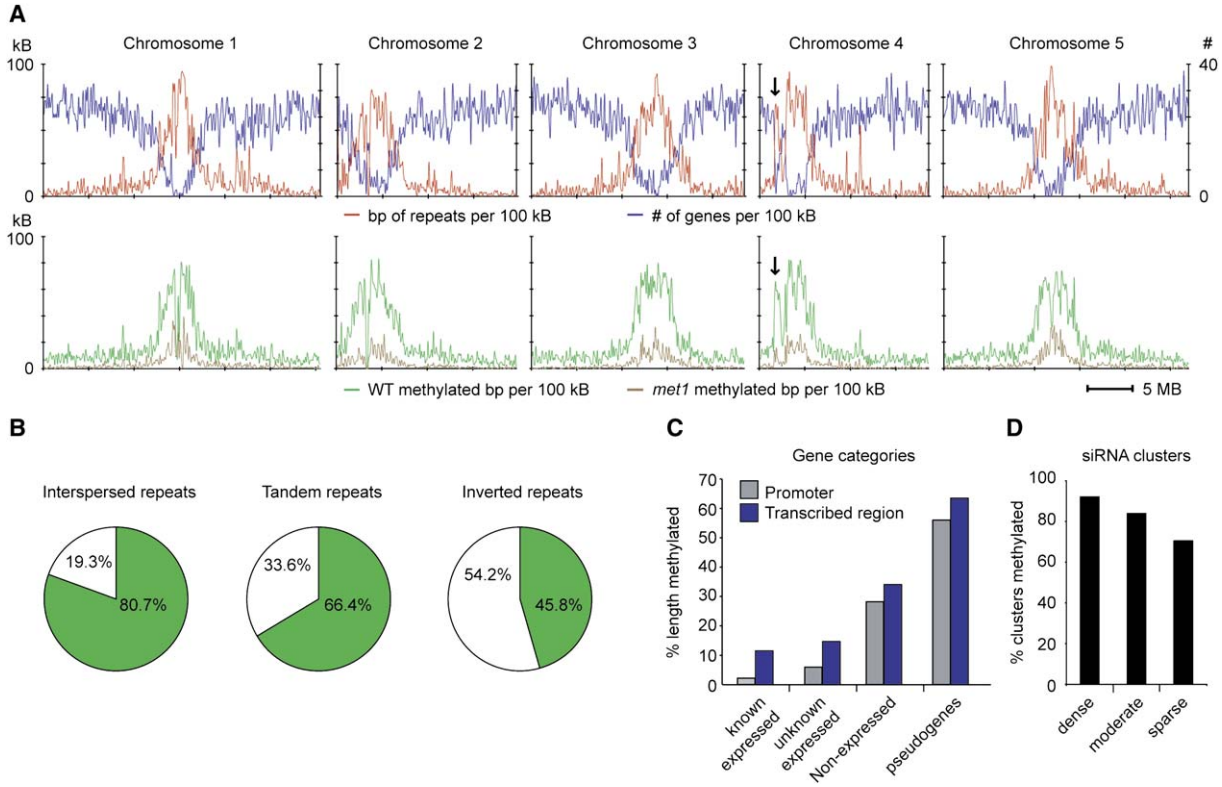


Figure 2. DNA Methylation Landscape in the *Arabidopsis* Genome

(A) Chromosomal distribution of cytosine-methylated DNA. Top panels show the total length of repeats (y axis, left-side scale) and number of genes (y axis, right-side scale) in a sliding 100 kb window. Bottom panels show the total length of methylated DNA in wild-type and in the *met1* mutant in 100 kb windows. Arrows indicate the heterochromatic knob on chromosome 4.

(B) Percentage of interspersed, tandem, and inverted repeats that are methylated (green).

(C) Methylation within promoters (200 bp upstream of the start of transcription) and transcribed regions of annotated genes of different categories (see text for definition). y axis = length of methylated sequences as a percentage of the total length of DNA in each category.

(D) Methylation of siRNA clusters (defined in Lu et al., 2005).

out global comparisons between methylated genomic regions and a large collection of *Arabidopsis* small RNA sequences determined by massively parallel signature sequencing (MPSS) (Lu et al., 2005). The majority of siRNA clusters (i.e., endogenous loci corresponding to high local concentrations of siRNAs) were heavily DNA methylated (Figure 2D). However, ~63% of methylated regions were not associated with siRNA clusters, suggesting that a large amount of DNA methylation is maintained without persistent targeting by siRNAs (Figure S7).

Interestingly, in the transcribed regions of genes, a much lower than average fraction of methylation was found to be associated with siRNAs (~9%, compared to the genome average of ~37%) (Figure 3C), suggesting that the maintenance of genic methylation is largely independent of targeting by siRNAs. This is consistent with the results from validation experiments that revealed that many of these regions contain MET1-dependent CG DNA methylation but not siRNA-targeted non-CG methylation (Figure S5) and also with the previous finding of CG methylation within many *Arabidopsis* genes (Tran et al., 2005a).

Previous evidence suggested that microRNAs might recruit DNA methylation enzymes to their target genes (Bao et al., 2004). However, we found that annealing sites in microRNA target genes were methylated at a level slightly below the genome average (22 of 136, ~16.2%) (see Figure S4 for *PHB* as an example). In addition, we found that only one (*MIR416a*) of the 103 microRNA precursor genes was methylated. For *trans*-acting siRNAs (tasiRNAs) (Allen et al., 2005; Peragine et al., 2004; Vazquez et al., 2004), we found that 2 of the 5 tasiRNA-generating loci (TAS1b and TAS3) and 7 of 9 tasiRNA target sites contained methylation. However, bisulfite sequencing of the methylated tasiRNA target sites in *ARF3* revealed CG methylation but an absence of non-CG methylation, which is a hallmark of RNA-directed DNA methylation (Figure S8) (Chan et al., 2005). Furthermore, DNA methylation persisted in the *dcl2 dcl3 dcl4* triple mutant that lacks detectable tasiRNAs (Henderson et al., 2006). Overall, these results do not support a general role for miRNAs or tasiRNAs in the active targeting of DNA methylation.

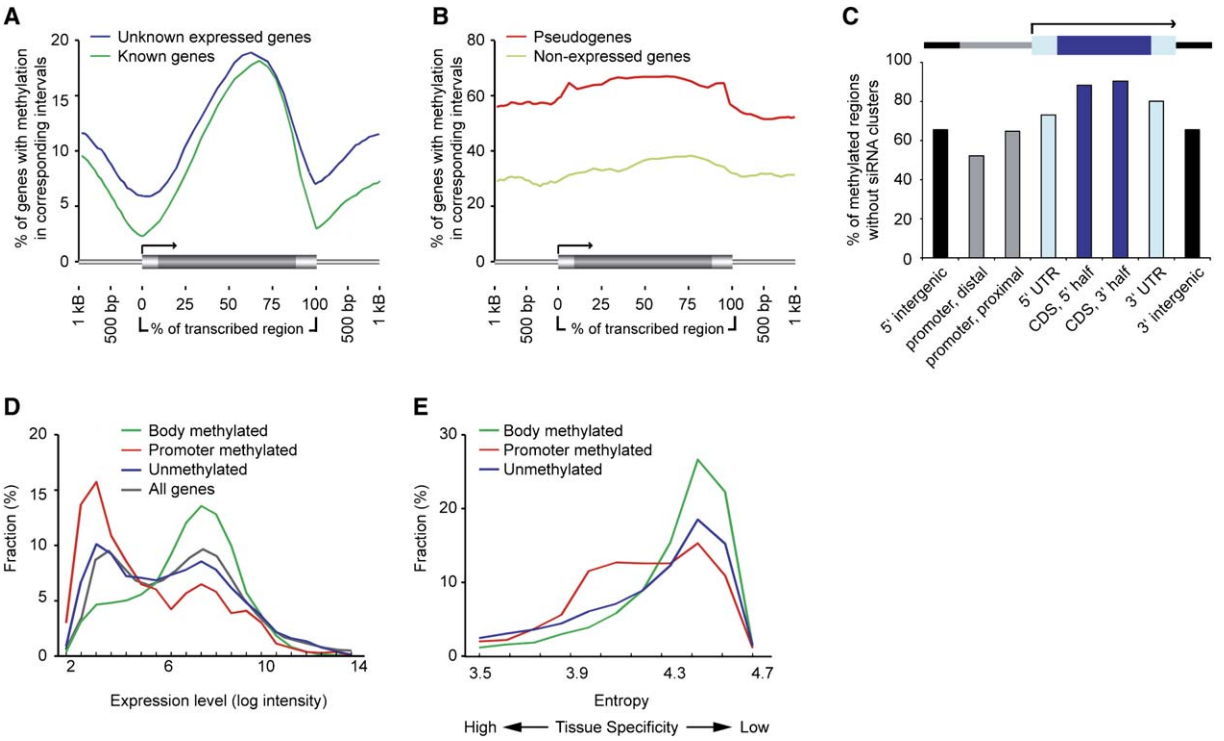


Figure 3. DNA Methylation of *Arabidopsis* Genes

(A) Distribution of DNA methylation within known genes and expressed genes with unknown functions. One kilobase regions upstream and downstream of each gene were divided into 50 bp intervals, each gene was divided into 20 intervals (5% each interval), and the percentage of genes with methylation in each interval was graphed. A schematic representation of a gene is shown as a thick horizontal bar (scaled to 2.5 kb, the average length of *Arabidopsis* genes).

(B) Distribution of DNA methylation within pseudogenes and nonexpressed genes.

(C) siRNA-independent methylation in different regions of genes. "Promoter, distal" and "promoter, proximal" refer to 200 bp–1 kb and 0–200 bp upstream of the transcription initiation site, respectively. The known and unknown expressed gene categories are included in this analysis.

(D) Expression levels of promoter-methylated, body-methylated, and unmethylated genes compared to all genes. x axis = expression level ($\log_2$ scale) averaged over 79 tissues and conditions (Schmid et al., 2005); vertical bars indicate the bins used. y axis = fraction of genes with given intensity level.

(E) Tissue specificity of promoter-methylated, body-methylated, and unmethylated genes measured by entropy level (Schug et al., 2005). Low entropy values indicate high tissue specificity. Vertical bars on x axis indicate the bins used. y axis = fraction of genes with the given entropy value.

Correlation between Genic DNA Methylation, Expression Level, and Tissue Specificity

In order to examine the relationship between DNA methylation and gene expression patterns, we compared the sites of DNA methylation with microarray expression data from 79 different tissues or conditions (Schmid et al., 2005). This analysis revealed a correlation between the position of DNA methylation within a gene (i.e., promoter or body), the level of gene expression, and its tissue specificity. As shown in Figure 3D, the expression level of body-methylated genes was significantly higher than that of unmethylated genes, whereas that of promoter-methylated genes was generally lower. In addition, a distinct fraction of promoter-methylated genes have a dramatically higher tissue specificity ($p < 10^{-37}$ by a Kolmogorov-Smirnov test) (Figure 3E). These results suggest that, in general, body-methylated genes are constitutively expressed at a higher level, whereas promoter-methylated genes tend to be expressed in a tissue-specific manner.

Comparison of the location of genic methylation with gene functional categories showed that the promoter-methylated group was highly enriched for genes involved in proteolysis, whereas the body-methylated group was enriched for catalytic enzymes (Table S1). Transcription factors were the most enriched category within the group of unmethylated genes (Table S1).

Alteration of DNA Methylation in the *met1* and *drm1 drm2 cmt3* Mutants

Using the mCIP-chip method, we analyzed genome-wide DNA methylation patterns in two DNA methylation-deficient mutant backgrounds (*met1* and a *drm1 drm2 cmt3* triple mutant) utilizing recently described null T-DNA insertion alleles (Chan et al., 2006). It has been previously shown by bisulfite genomic sequencing that, at all tested loci, the *drm1 drm2 cmt3* triple mutant eliminates the vast majority of non-CG methylation but not CG methylation, whereas the *met1* mutant lacks virtually all CG

methylation as well as a substantial amount of non-CG methylation (Cao and Jacobsen, 2002; Tariq et al., 2003). Genome-wide DNA methylation profiles detected by mCIP-chip in *drm1 drm2 cmt3* were roughly similar to that of wild-type; ~93% of the regions methylated in wild-type remained methylated in *drm1 drm2 cmt3* (see Figure 1D for examples). This suggests that the majority of non-CG methylation colocalizes with CG methylation and is consistent with previous findings that loss of non-CG methylation often does not disturb the remaining CG methylation (Cao and Jacobsen, 2002). Interestingly, the ~7% of regions that did lose methylation in *drm1 drm2 cmt3* showed a lower CG dinucleotide content relative to the genome average (Figure S3) and thus may represent regions that rely on non-CG methylation for the overall maintenance of methylation patterning.

In contrast, ~64% of the regions methylated in wild-type were no longer detected in *met1*. The greatest reduction was found in expressed genes, where over 97% of the methylation was eliminated. This is consistent with the fact that genic methylation often occurs only at CG sites and is less likely to be associated with siRNAs (Figure 3C and Figure S5) (Tran et al., 2005a). The residual methylated regions in *met1* were highly clustered in heterochromatin, consisted of mostly repetitive sequences (~92%), showed a much higher level of association with siRNA clusters, and had a significantly higher CNG content (Figure S10). It is therefore likely that, in *met1*, the residual methylation is maintained by CMT3 at CNG sites as well as by the siRNA-directed activity of DRM1/2 (Chan et al., 2005).

Loss of DNA Methylation in *met1* Results in Massive Transcriptional Reactivation of Pseudogenes and Transposons

To infer the function of DNA methylation in regulating gene expression, we determined the expression profiles of wild-type as well as the *met1* and *drm1 drm2 cmt3* DNA methyltransferase mutants using genome-wide tiling-array expression analyses (Yamada et al., 2003). Expression levels of sense and antisense transcription were determined separately by generating strand-specific probes and using two arrays, one covering the forward strands of the genome (the same array used for methylation analysis) and a second similar array covering the reverse genomic strands. Massive reactivation of previously silenced transposons and pseudogenes was observed in *met1*, resulting in drastically elevated transcriptional activity at the centromeric and pericentromeric heterochromatic regions (Figure 4A) (Lippman et al., 2004; Miura et al., 2001; Singer et al., 2001). In contrast, relatively few transposons or pseudogenes were activated in *drm1 drm2 cmt3*, and the transcriptional activity remained grossly similar to that of wild-type at the chromosomal level (Figure S11). These findings are consistent with our observation that CG methylation was largely unchanged in the absence of non-CG methylation and suggest that, in most cases, CG methylation alone is sufficient for the transcriptional silencing of transposons and pseudogenes.

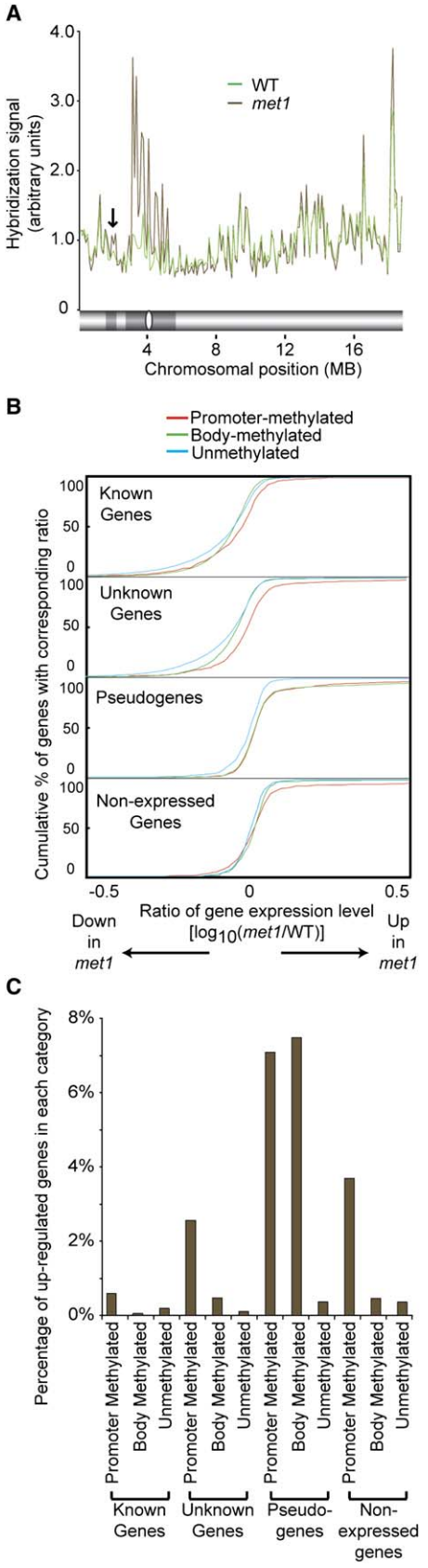
Promoter-Methylated Genes Are Preferentially Overexpressed in *met1*

The relative expression level of each gene in *met1* compared to wild-type was determined from the hybridization intensity of probes located within exons and UTRs (~35 probes per gene on average; see Supplemental Data for details). We then compared gene-level changes for promoter-methylated, body-methylated, and unmethylated genes within each category (Figure 4B). For the known, unknown, and nonexpressed categories, a significantly larger fraction of promoter-methylated genes were increased in steady-state RNA levels in *met1* compared to body-methylated or unmethylated genes ($p < 0.01$ by a Kolmogorov-Smirnov test). In contrast, expression of body-methylated genes did not appear to be systematically increased when compared to unmethylated genes on a global scale (Figure S12). For pseudogenes, the expression of both promoter- and body-methylated genes was significantly increased ($p < 10^{-5}$). The preferential increase in expression of promoter-methylated genes in *met1* was more obvious when only the most significantly upregulated genes were examined (Figure 4C). These findings suggest that methylation in promoter regions plays a more profound role in downregulating gene expression for expressed genes and that both promoter and body methylation are important for silencing transposons and pseudogenes. Because of its more subtle effects on methylation patterning, we did not see bulk changes in the expression of promoter-methylated genes in *drm1 drm2 cmt3* (Figure S11); however, the expression of many specific genes was altered (see below).

Antisense Gene Expression in DNA Methylation Mutants

Previous studies have revealed the presence of antisense expression corresponding to a large fraction of *Arabidopsis* genes (Yamada et al., 2003). Methylation in the body of genes has been proposed to suppress antisense transcription from cryptic promoters to ensure normal sense gene expression (Tran et al., 2005a). Although cases of antisense overexpression were found in *met1* (e.g., Figure 5A), such examples were relatively rare. Furthermore, we did not observe a systematic increase of antisense transcription from body-methylated genes in *met1*, and the change in antisense transcription in body-methylated genes was not significantly different from that in unmethylated genes (Figure 5B). Similarly, there was no significant difference between methylated genic regions or unmethylated control genic regions with respect to changes in the abundance of antisense transcripts (Figure 5B).

We also investigated whether changes in sense and antisense transcripts in *met1* were correlated (e.g., whether elevated levels of antisense RNA were generally correlated with decreased accumulation of sense RNAs). As shown in Figure 5C and Figure S13, changes in sense



and antisense expression were found to be largely independent for both body-methylated and unmethylated known genes. Taken together, these results are consistent with several recent studies of antisense transcription (Faghihi and Wahlestedt, 2006; Lu et al., 2005; Wang et al., 2005) and do not support a general role for DNA methylation within genes in suppressing antisense transcription or for antisense transcription generally interfering with sense transcription.

Profile of Genes Affected by DNA Methylation

Comparison of the genes that were most significantly increased in expression in *met1* and *drm1 drm2 cmt3* (listed in Tables S2 and S3) revealed an interesting distinction. Most genes that showed an increase in RNA abundance in *met1* were pseudogenes clustered in pericentromeric heterochromatic regions. In contrast, known genes distributed throughout euchromatin accounted for the largest fraction (~69%) of upregulated genes in *drm1 drm2 cmt3* (Figures 6A and 6B and Figure S14).

While it is difficult to determine the fraction of gene expression changes in *drm1 drm2 cmt3* that can be attributed to secondary effects, in many cases, the loss of DNA methylation of specific genes correlated with their overexpression, suggesting a direct role of non-CG methylation in regulating these genes. One such example is the F box protein encoded by the gene At2g17690. As shown in Figure 6C, the promoter region of At2g17690 contains a tandem repeat (seven copies of a 32-mer) that is heavily methylated and also associated with siRNAs, and the gene is normally not expressed (see RT-PCR results in Figure 6B). In *met1*, an incomplete reduction of promoter methylation was accompanied by a mild increase in At2g17690 gene expression. In contrast, promoter methylation was virtually eliminated in *drm1 drm2 cmt3*, resulting in a dramatic increase in steady-state RNA level for this gene (Figure 6B).

These results provide evidence that non-CG methylation is important in regulating gene expression on a genome-wide scale and identify a host of candidate genes that may be directly regulated by non-CG methylation. A possible explanation for why non-CG methylation is seemingly more important than CG methylation in regulating potentially functional genes may reside with the mechanisms by which these two types of methylation are inherited. In

Figure 4. Global Changes of Gene Expression in *met1* Compared to Wild-Type

(A) Chromosome-level RNA expression levels in wild-type and *met1*. Chromosome 4 is shown as an example. y axis = median probe-level hybridization intensity over 100 kb windows. A schematic representation of chromosome 4 is shown (bottom) and labeled as in Figure 1. Arrow indicates the increased expression level in the heterochromatic knob on chromosome 4 in *met1*.

(B) Global changes of gene expression in *met1* compared to wild-type, shown as cumulative distributions. x axis = gene-level expression fold change in *met1* compared to wild-type ($\log_{10}$ scale).

(C) Percentage of genes with increased expression in *met1* (listed in Table S2) within different gene categories.

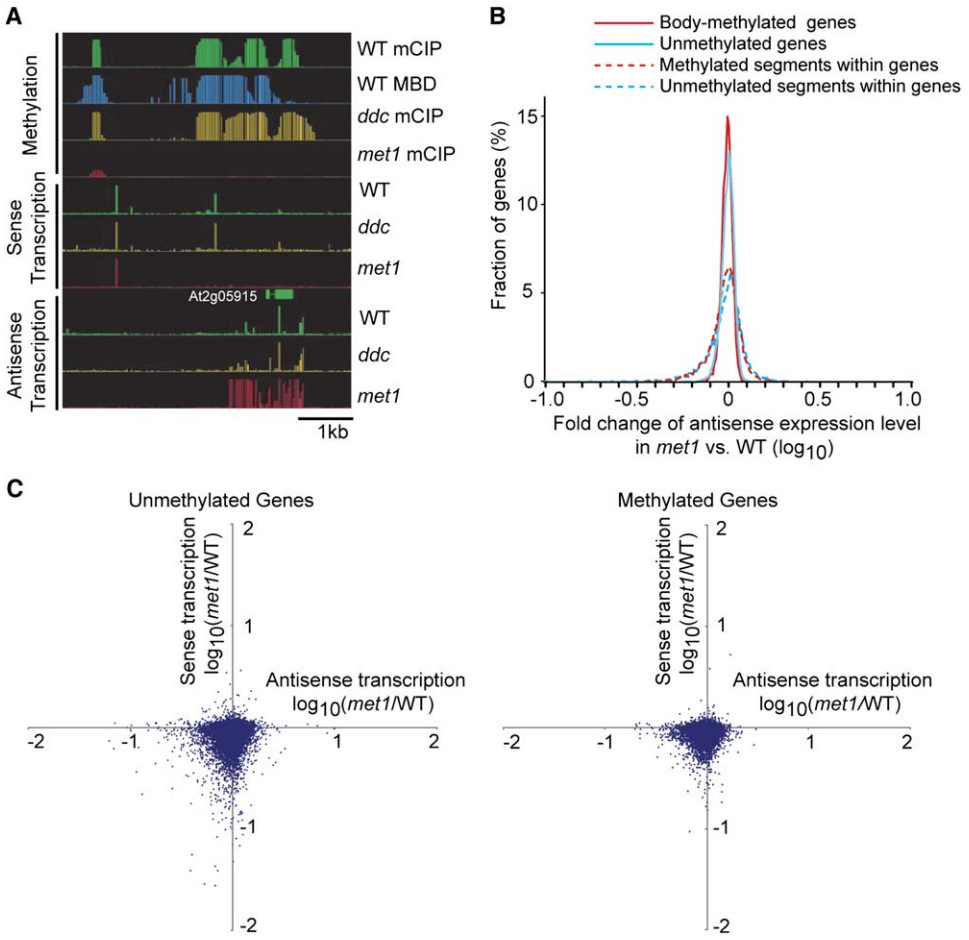


Figure 5. Global Changes of Antisense RNA Expression in *met1*

(A) An example (At2g05915) where loss of DNA methylation and increased antisense RNA accumulation coincide in *met1*.

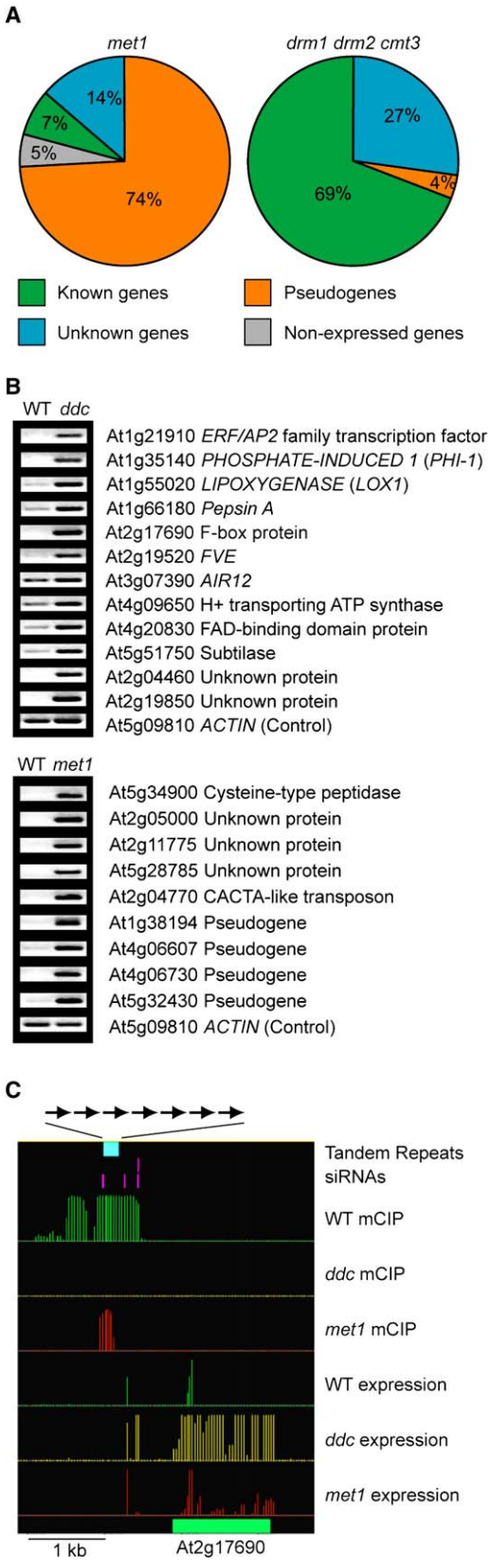
(B) Loss of methylation within genes does not systematically cause overexpression of antisense RNA in *met1*. Graph shows the distribution of expression fold change in *met1* versus wild-type for either the entire predicted transcribed region of body-methylated genes or the methylated segments within these genes. Unmethylated genes and unmethylated segments within genes are shown as controls. x axis = fold change in antisense RNA expression (log10 scale). y axis = percentage of regions or genes with a given fold change.

(C) Lack of correlation between changes in sense and antisense RNA expression for both unmethylated and body-methylated genes in *met1*. x axis = fold change in antisense RNA expression ($\log_{10}$ scale). y axis = fold change in sense RNA expression ($\log_{10}$ scale).

particular, the replication-coupled-maintenance methylation activity of MET1 at CG sites might mainly function to maintain the stable silencing of transposons and pseudogenes, whereas the siRNA-directed DRM activity and the histone methylation-directed CMT3 activity (Chan et al., 2005) are likely to be dynamic and may play more important roles in regulating the expression of endogenous genes. Consistent with this idea, profound differences exist in the siRNA populations accumulated in different tissues (Lu et al., 2005), which could account for differential targeting of genes by DNA methylation.

Intergenic Noncoding RNAs Controlled by DNA Methylation

Methylated regions account for ~8% of the total length of nonrepetitive intergenic regions, and we found that widespread changes in the accumulation of intergenic noncoding RNA (ncRNA) occurred in the DNA methylation mutants (Table S4). The most dramatic changes were observed in *met1*, where we detected 264 overexpressed ncRNAs compared to wild-type (Figure 7A). Of these, 214 (~80%) were hypomethylated in *met1* relative to wild-type (Figure 7B). Although many of these transcripts (~67%) were repetitive and could represent unannotated transposons, 34 were found to be single-copy in the genome (30 hypomethylated in *met1*), and an additional 54 had 2–10 copies (36 hypomethylated in *met1*) (Figure 7C). Strikingly, 87 of the 88 single- or low-copy ncRNAs did not exhibit significant homology with sequences from other organisms in GenBank (the NR, EST, or GSS databases). These results suggest that a large number of fast-evolving ncRNAs exist in



the *Arabidopsis* genome whose expression is controlled epigenetically by DNA methylation.

In addition to overexpressed ncRNAs, we also identified 60 intergenic regions that were expressed in wild-type plants but not in *met1* (Table S5). Consistent with this result, 59 of the 60 suppressed ncRNAs were also found in the GenBank database of expressed sequence tags (dbEST); none of these were annotated with known function. Nearly all suppressed ncRNAs (~97%) are present as single-copy sequences in the genome, and many (~17%) have homologous sequences in other plant genomes, suggesting that they might have a conserved biological function (or functions). Thus, it appears that in *Arabidopsis*, the normal expression of many potentially important ncRNAs may be regulated by DNA methylation.

Conclusion

In summary, we have combined optimized biochemical methods and high-density whole-genome tiling microarrays to enable the first high-resolution genome-wide characterization of DNA methylation for any organism. Consistent with previous studies, we found a significant enrichment of methylation in heterochromatin and siRNA clusters as well as an important role for methylation in silencing transposons and pseudogenes (Lippman et al., 2004; Miura et al., 2001; Singer et al., 2001). In addition, we discovered an unexpectedly large amount of genic methylation and uncovered intriguing distinctions among differentially methylated genes with regard to their functional classification, expression level, and tissue specificity. In parallel, genome-wide transcription profiling allowed the identification of hundreds of genes and ncRNAs whose expression is affected by changes in DNA methylation, and these analyses have uncovered an unexpectedly important role for non-CG methylation in the regulation of functional genes.

The entire set of whole-genome DNA methylation and gene expression data can be downloaded from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/projects/geo/, accession numbers GSE5094 and GSE5074). The data can also be viewed along with additional information including annotations of repeats and small RNAs at http://epigenomics.mcdb.ucla.edu/DNAmeth/ and at http://signal.salk.edu/cgi-bin/methylome. These new tools and genome-wide resources should serve as the raw material for future studies to elucidate the global control of DNA methylation patterning by mechanisms that include

Figure 6. Genes Overexpressed in *met1* and *drm1 drm2 cmt3* DNA Methyltransferase Mutants

(A) Composition of overexpressed genes within different gene categories.

(B) Examples of RT-PCR validation of overexpressed genes identified by tiling-array analysis in *met1* and *drm1 drm2 cmt3*.

(C) The F box-containing gene At2g17690 as an example where loss of promoter methylation in *drm1 drm2 cmt3* is associated with ectopic overexpression. Tracks are labeled as in Figure 1. Arrows at the top represent the tandem repeats located in the promoter region. RT-PCR validation result for this gene is shown in (B).

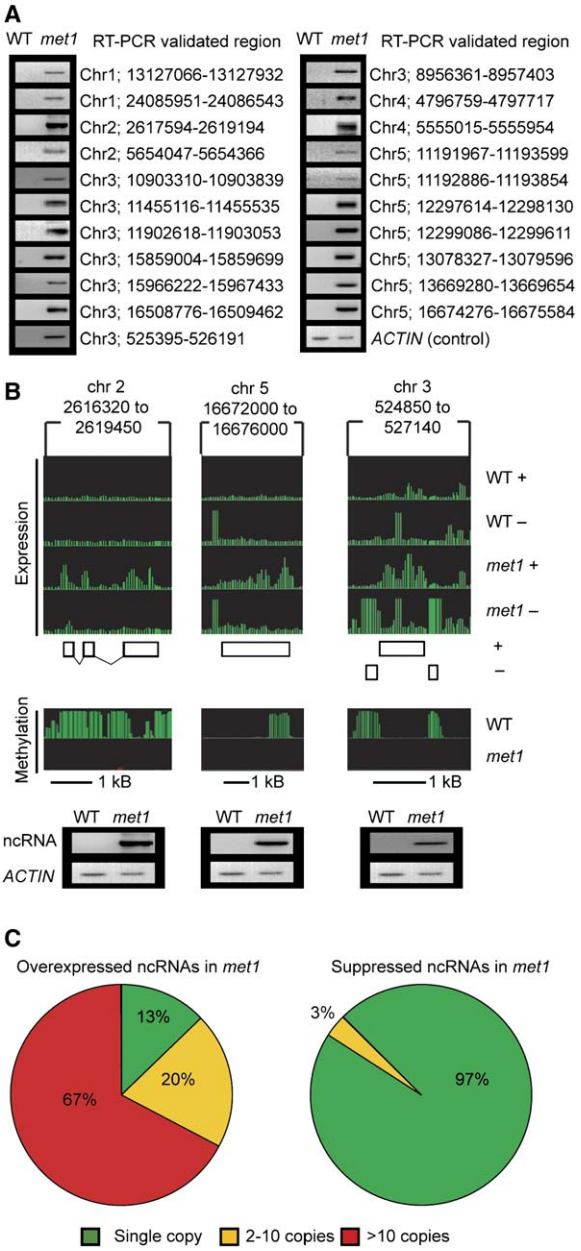


Figure 7. Expression of Intergenic Noncoding RNAs (ncRNAs) Regulated by DNA Methylation

(A) RT-PCR validation of overexpressed ncRNAs in *met1*.

(B) Examples of hypomethylation correlated with ncRNA overexpression. Expression levels are shown for the forward (+) and reverse (–) strands. RT-PCR validation results are shown below. The structures of the overexpressed ncRNAs deduced from cloned sequences are diagrammed as open rectangles. DNA methylation was determined by the mCIP-chip method.

(C) Copy number of ncRNAs that are overexpressed or suppressed in *met1*.

small RNAs, histone modifications, and chromatin remodeling (Chan et al., 2004, 2005; Dalmay et al., 2000; Jackson et al., 2002; Malagnac et al., 2002; Mette et al., 2000; Tamaru and Selker, 2001) and for investigations into the biological functions of genes and noncoding transcripts whose expression is controlled by DNA methylation. Finally, whole-genome tiling arrays are now available for many other organisms, and the approaches developed should assist in future studies of larger and more complex genomes, including those that pertain to epigenetic regulation in human disease (Mockler et al., 2005).

EXPERIMENTAL PROCEDURES

Plant Materials and DNA Extraction

All plants used in this study were of the *Arabidopsis thaliana* Columbia (Col-0) accession. The *met1* and *drm1 drm2 cmt3* mutants were previously published (Chan et al., 2006; Saze et al., 2003). The mutant alleles were *met1-3* and *drm1-2*, *drm2-2*, *cmt3-11*. To avoid variability caused by inbreeding, all homozygous mutant plant material used for DNA methylation and RNA expression experiments was prepared using F1 plants from segregating populations. Plants were grown under continuous light. DNA was extracted from 5-week-old plants using the Plant DNeasy Maxi Kit (QIAGEN) and sonicated to ~350 bp. The entire aerial parts of two or three plants were pooled for each biological replicate.

Methylcytosine Immunoprecipitation

The methylcytosine immunoprecipitation (mCIP) method was adapted from a previous study (Keshet et al., 2006). Immunoprecipitation was performed by incubating 2 μg of sonicated genomic DNA with 20 μg mouse anti-methylcytosine monoclonal antibody (Calbiochem) in 600 μl of buffer FB (10 mM Tris-HCl [pH 7.5], 50 mM NaCl, 1 mM EDTA) at 4°C for 12 hr. One hundred microliters each of Dynabeads Protein G and Protein A (Dynal) were added to the mix and incubated at 4°C for 6 hr. Dynabeads were washed six times by gentle mixing at 4°C for 10 min with 600 μl of buffer FB. Elution was performed three times by vortexing in 200 μl TE containing 1.5%, 0.5%, and 0.1% SDS, respectively. DNA was recovered from the unbound fraction, each of the six washes, and the elution fraction by phenol-chloroform extraction and ethanol precipitation. A fraction of the recovered DNA was used for real-time PCR to determine the amount of the methylated *FWA* promoter, *AtSN1*, and *Ta3* as well as the unmethylated *ACTIN* promoter in each fraction (Figure 1A). The remaining DNA from the unbound and elution fractions was amplified (see below) and used for microarray hybridization. Three biological replicates were performed for each genotype and yielded consistent results (see Supplemental Data).

Affinity Purification of Methylated DNA Using the Methylcytosine Binding Domain of the Human MeCP2 Protein

The expression and purification of the methylcytosine binding domain (MBD) were performed as previously described (Cross et al., 1994; Selker et al., 2003). Briefly, 10 mg of purified His-tagged MBD was affixed to a Ni-agarose gel column, and 10 μg of sonicated genomic DNA was subjected to binding by MBD in 10 ml loading buffer (20 mM HEPES [pH 7.9], 100 mM NaCl, 10% glycerol, 0.1% Triton X-100, 0.5 mM PMSF) at a flow rate of 1 ml/min. After an initial wash with 10 ml loading buffer containing 0.4 M NaCl, 26 ml of loading buffer containing a linear gradient of NaCl from 0.4 M to 1.0 M was used to elute DNA with flow-through collected in 2 ml fractions, followed by a final wash with 10 ml loading buffer containing 1.0 M NaCl. DNA was recovered from each elution. A fraction of the DNA was used to determine the amount of the methylated *FWA* promoter and the unmethylated *ACTIN* promoter by real-time PCR, which showed that the 0.5 M NaCl fraction had the highest *ACTIN*/*FWA* ratio and that the 0.8 M NaCl fraction had the highest *FWA*/*ACTIN* ratio (Figure 1B). The remaining DNA from these two fractions was amplified and used for

microarray hybridization. Three biological replicates were performed for each genotype.

Microarray Design, Experimental Procedures, and Data Analyses

Detailed descriptions are included in the Supplemental Data.

Reverse Transcription-Polymerase Chain Reaction and Cloning

One microgram of total RNA was denatured at 65°C for 10 min, followed by first-strand cDNA synthesis in a 50 μl reaction mix using the Transcript First-Strand cDNA Synthesis Kit (Roche), and incubated at 85°C for 5 min. One microliter of the mix was used with gene-specific primers for reverse transcription-polymerase chain reaction (RT-PCR). The cycling parameters for ncRNAs were 95°C for 10 min, 21 cycles of 1 min at 95°C, 1 min at 58°C, 1 min at 72°C, and a final elongation step at 72°C for 10 min. The cycling parameters for overexpressed genes were 95°C for 30 s, 30 cycles of 10 s at 95°C, 30 s at 60°C, 30 s at 72°C, and a final elongation step at 72°C for 1 min. PCR primers are listed in Tables S6 and S7, respectively. For ncRNAs, PCR products were purified with the QIAquick PCR Purification Kit (QIAGEN) and cloned using the TOPO TA Cloning Kit (Invitrogen).

Bisulfite Sequencing

Genomic bisulfite sequencing was performed as previously described (Cao and Jacobsen, 2002). The regions sequenced and the primers used are listed in Table S8.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, Supplemental References, 14 figures, and 9 tables and can be found with this article online at http://www.cell.com/cgi/content/full/126/6/■■■/DC1/.

ACKNOWLEDGMENTS

We thank Howard Cedar for the initial mCIP protocol, Jim Carrington for miRNA and tasiRNA target gene information, Zhirong Bao and Sean R. Eddy for providing the RECON repeat library, and T. Gingeras and colleagues at Affymetrix Laboratories for assistance in tiling-array development. S.W.-L.C., X.Z., J.Y., S.E.J., and J.R.E. designed the experiments. X.Z. generated the mCIP and MBD methylation data. J.Y. and A.S. generated the expression data. X.Z., M.P., S.C., and S.W.-L.C. analyzed the methylation data. S.C., M.P., and H.C. developed analysis methods for the expression data. J.Y., S.C., M.P., X.Z., and H.C. analyzed the expression data. H.C. and J.R.E. designed the tiling arrays. S.W.-L.C., J.Y., I.R.H., A.S., and P.S. generated the validation data. X.Z. wrote the manuscript. S.E.J. and J.R.E. edited the manuscript. This work was supported by grants from the NIH ENCODE Program HG003523 (S.E.J. and J.R.E.), NIH grant GM60398 (S.E.J.), and NSF 2010 Project DBI0520253 (J.R.E.). A.S. was supported by the NIH/NIGMS-funded UCSD Genetics Training Program (T32 GM08666). S.W.-L.C. is a DOE Energy Biosciences Fellow of the Life Science Research Foundation. I.R.H. is supported by an EMBO Long-Term Postdoctoral Fellowship. S.E.J. is an investigator of the Howard Hughes Medical Institute.

Received: June 30, 2006
Revised: August 1, 2006
Accepted: August 7, 2006
Published online: August 31, 2006

REFERENCES

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207–221.

Bao, N., Lye, K.W., and Barton, M.K. (2004). MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev. Cell *7*, 653–662.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E.W., Wu, B., Doucet, D., Thomas, N.J., Wang, Y., Vollmer, E., et al. (2006). High-throughput DNA methylation profiling using universal bead arrays. Genome Res. *16*, 383–393.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Cao, X., and Jacobsen, S.E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc. Natl. Acad. Sci. USA *99* (*Suppl 4*), 16491–16498.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat. Rev. Genet. *6*, 351–360.

Chan, S.W., Henderson, I.R., Zhang, X., Shah, G., Chien, J.S.-C., and Jacobsen, S.E. (2006). RNAi, DRD1, and Histone Methylation Actively Target Developmentally Important Non-CG DNA Methylation in Arabidopsis. PLoS Genet. *2*, e83.

Ching, T.T., Maunakea, A.K., Jun, P., Hong, C., Zardo, G., Pinkel, D., Albertson, D.G., Fridlyand, J., Mao, J.H., Shchors, K., et al. (2005). Epigenome analyses using BAC microarrays identify evolutionary conservation of tissue-specific methylation of SHANK3. Nat. Genet. *37*, 645–651.

Cross, S.H., Charlton, J.A., Nan, X., and Bird, A.P. (1994). Purification of CpG islands using a methylated DNA binding column. Nat. Genet. *6*, 236–244.

Dalmay, T., Hamilton, A., Mueller, E., and Baulcombe, D.C. (2000). Potato virus X amplicons in arabidopsis mediate genetic and epigenetic gene silencing. Plant Cell *12*, 369–379.

Faghihi, M.A., and Wahlestedt, C. (2006). RNA interference is not involved in natural antisense mediated regulation of gene expression in mammals. Genome Biol. *7*, R38.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Hatada, I., Fukasawa, M., Kimura, M., Morita, S., Yamada, K., Yoshikawa, T., Yamanaka, S., Endo, C., Sakurada, A., Sato, M., et al. (2006). Genome-wide profiling of promoter methylation in human. Oncogene *25*, 3059–3064.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat. Genet. *38*, 721–725.

Hirochika, H., Okamoto, H., and Kakutani, T. (2000). Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation. Plant Cell *12*, 357–369.

Hollander, M., and Wolfe, D.A. (1999). Nonparametric Statistical Methods, Second Edition (New York: John Wiley and Sons, Inc.).

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556–560.

Jacobsen, S.E., Sakai, H., Finnegan, E.J., Cao, X., and Meyerowitz, E.M. (2000). Ectopic hypermethylation of flower-specific genes in Arabidopsis. Curr. Biol. *10*, 179–186.

Ji, H., and Wong, W.H. (2005). TileMap: create chromosomal map of tiling array hybridizations. Bioinformatics *21*, 3629–3636.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H., and Simon, I. (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat. Genet. *38*, 149–153.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature *430*, 471–476.

Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., and Green, P.J. (2005). Elucidation of the small RNA component of the transcriptome. Science *309*, 1567–1569.

Malagnac, F., Bartee, L., and Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J. *21*, 6842–6852.

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. *19*, 5194–5201.

Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature *411*, 212–214.

Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E., and Ecker, J.R. (2005). Applications of DNA tiling arrays for whole-genome analysis. Genomics *85*, 1–15.

Ngernprasirtsiri, J., Kobayashi, H., and Akazawa, T. (1988). DNA methylation as a mechanism of transcriptional regulation in nonphotosynthetic plastids in plant cells. Proc. Natl. Acad. Sci. USA *85*, 4750–4754.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev. *18*, 2368–2379.

Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Andrews, T.D., Howe, K.L., Otto, T., Olek, A., et al. (2004). DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. PLoS Biol. *2*, e405.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., and Bestor, T.H. (2006). Large-scale structure of genomic methylation patterns. Genome Res. *16*, 157–163.

Saze, H., Scheid, O.M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat. Genet. *34*, 65–69.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. Nat. Genet. *37*, 501–506.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J., Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol. *6*, R33.

Selker, E.U., Tountas, N.A., Cross, S.H., Margolin, B.S., Murphy, J.G., Bird, A.P., and Freitag, M. (2003). The methylated component of the Neurospora crassa genome. Nature *422*, 893–897.

Singer, T., Yordan, C., and Martienssen, R.A. (2001). Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev. *15*, 591–602.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of- function epigenetic alleles of a homeodomain gene. Mol. Cell *6*, 791–802.

Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature *414*, 277–283.

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc. Natl. Acad. Sci. USA *100*, 8823–8827.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796–815.

Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E., and Henikoff, S. (2005a). DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. Curr. Biol. *15*, 154–159.

Tran, R.K., Zilberman, D., de Bustos, C., Ditt, R.F., Henikoff, J.G., Lindroth, A.M., Delrow, J., Boyle, T., Kwong, S., Bryson, T.D., et al. (2005b). Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in Arabidopsis. Genome Biol. *6*, R90.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol. Cell *16*, 69–79.

Wang, X.J., Gaasterland, T., and Chua, N.H. (2005). Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. Genome Biol. *6*, R30.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. *37*, 853–862.

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., et al. (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome. Science *302*, 842–846.

Accession Numbers

All raw microarray data (CEL files) for DNA methylation and expression analyses have been deposited in GEO under the accession numbers GSE5094 and GSE5074, respectively. The ncRNA sequences have been deposited in GenBank under the accession numbers BT025666–BT025694.