Subtle Perturbations of the Maize Methylome Reveal Genes and Transposons Silenced by Chromomethylase or RNA-Directed DNA Methylation Pathways

Sarah N. Anderson,* Gregory J. Zynda,† Jawon Song,‡ Zhaoxue Han,† Matthew W. Vaughn,† Qing Li,§ and Nathan M. Springer*,1

*Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108, †Texas Advanced Computing Center, University of Texas, Austin, TX 78758, ‡State Key Laboratory of Crop Stress Biology for Arid Areas, Northwest A&F University, Yangling 712100, Shaanxi, China, and §National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, 430070, China

ORCID IDs: 0000-0002-1671-2286 (S.N.A.); 0000-0003-4992-9614 (G.J.Z.); 0000-0002-1384-4283 (M.W.V.); 0000-0003-3232-9479 (Q.L.); 0000-0002-7301-4759 (N.M.S.)

ABSTRACT DNA methylation is a chromatin modification that can provide epigenetic regulation of gene and transposon expression. Plants utilize several pathways to establish and maintain DNA methylation in specific sequence contexts. The chromomethylase (CMT) genes maintain CHG (where H = A, C or T) methylation. The RNA-directed DNA methylation (RdDM) pathway is important for CHH methylation. Transcriptome analysis was performed in a collection of Zea mays lines carrying mutant alleles for CMT or RdDM-associated genes. While the majority of the transcriptome was not affected, we identified sets of genes and transposon families sensitive to context-specific decreases in DNA methylation in mutant lines. Many of the genes that are up-regulated in CMT mutant lines have high levels of CHG methylation, while genes that are differentially expressed in RdDM mutants are enriched for having nearby mCHH islands, implicating context-specific DNA methylation in the regulation of expression for a small number of genes. Many genes regulated by CMTs exhibit natural variation for DNA methylation and transcript abundance in a panel of diverse inbred lines. Transposon families with differential expression in the mutant genotypes show few defining features, though several families up-regulated in RdDM mutants show enriched expression in endosperm tissue, highlighting the potential importance for this pathway during reproduction. Taken together, our findings suggest that while the number of genes and transposon families whose expression is reproducibly affected by mild perturbations in context-specific methylation is small, there are distinct patterns for loci impacted by RdDM and CMT mutants.

KEYWORDS DNA methylation chromomethylase transposable elements RNA directed DNA methylation maize

The epigenome describes the potential for additional heritable information that can be passed on through mitosis or meiosis (Hofmeister et al. 2017). DNA methylation is one molecular mechanism that can provide epigenetic information. There is interest in the potential forcryptic information in genomes which is normally silenced by epigenetic mechanisms but could be activated through epigenetic changes without requiring any genetic change. This cryptic information may represent an untapped source of variation that could be used for crop improvement.

Much is known about the mechanisms that control DNA methylation and the functional roles of DNA methylation in regulating transposon and gene expression in the model plant Arabidopsis thaliana (Law and Jacobsen 2010; Matzke and Mosher 2014). However, our knowledge of the regulating mechanisms and function of DNA methylation is much more limited in crop plants. Evidence in rice and maize suggests that major perturbations of DNA methylation disrupt development and the seeds/plants are not viable (Yamauchi et al. 2014; Hu et al. 2014; Li et al. 2014). Forward genetic screens for factors involved in epigenetic phenomena such as paramutation (Dorweiler et al. 2000; Hollick et al. 2005; Alleman et al. 2006; Hale et al. 2007; Erhard et al. 2009) or transgene silencing (McGinnis et al. 2006) have identified several genes that are associated with DNA methylation or chromatin in maize (Hollick 2017). In addition, reverse genetic approaches have been utilized in attempts to document
the function of putative methyltransferase genes or other genes associated with DNA methylation (Papa et al. 2001; Li et al. 2014). To date, these mutants have provided partial reductions in DNA methylation in specific sequence contexts but no mutants with drastic reductions in genomic DNA methylation have been recovered in maize.

Surveys of natural variation for DNA methylation among diverse lines of maize have revealed many examples of differentially methylated regions (DMRs) (Eichten et al. 2011, 2013; Regulski et al. 2013; Li et al. 2013b). A subset of the genes located near DMRs exhibit a negative correlation between DNA methylation and gene expression (Eichten et al. 2013; Li et al. 2015b). This is primarily found at genes that have CG or CHG methylation in regions surrounding the transcriptional start site (TSS) and show qualitative (on/off) expression variation among genotypes (Li et al. 2015b). This suggests the potential for cryptic information in the maize genome that is epigenetically silenced in some lines but can be active due to epigenetic changes in other genotypes.

Several maize mutant lines with context-specific or locus-specific perturbations of DNA methylation have been previously identified (Li et al. 2014). The mutants include mop1-1 and mop3-1, two mutants recovered in screens for factors required for maintenance of the paramutated state at the B’ locus (Dorweiler et al. 2000). The mop1 locus encodes an RNA-dependent RNA polymerase related to RDR2 in Arabidopsis (Alleman et al. 2006) while the mop3 locus encodes the largest subunit of RNA Pol IV (Sloan et al. 2014) and this same locus has also been defined by the rmr6-1 mutant allele and has been named as the rmr6 (required to maintain repression 6) / rpd1 (RNA polymerase d1) locus (Erhard et al. 2009). Mutant alleles for the two chromomethylase genes present in the maize genome, Zmet2 and Zmet5 also influence context-specific DNA methylation patterns (Papa et al. 2001; Makarevitch et al. 2007; Li et al. 2014). These genes are likely paralogs resulting from a whole genome duplication event and are orthologous to CMT3 from Arabidopsis thaliana (Bewick et al. 2017). Previous research has found that mop1-1 and mop3-1 genotypes have lost CHH methylation at many genomic regions with elevated CHH, and there are changes in CG and CHG at these sites as well (Li et al. 2014). However, as these types of regions are quite rare in the maize genome these mutants have minimal effects on genome-wide levels of CG and CHG methylation. The zmet2-m1 mutant and, to a lesser extent, the zmet5-m1 mutant, result in a reduction of CHG methylation. These mutants also cause reductions of CWA methylation (where W is A or T) in genomic regions with low, but detectable, CG and CHG methylation. The extent, the rmr6-1 mutant and, to a lesser extent, the rmr6/rpd1 locus, provided evidence for a potential role in stress response (Forestan et al. 2016). The rmr6-1 mutation appears to increase the proportion of the genome that is transcribed but has subtle effects at most loci with relatively few genes with significant changes in expression level (Forestan et al. 2017). However, there have not been comprehensive studies on the overlap of genes or transposons that are sensitive to mutations in different CMT or RdDM genes in maize.

Each of the mutant genotypes used for this study has subtle effects on genomic methylation levels and can produce viable plants. There are several phenotypic abnormalities observed in mop1-1 and mop3-1 stocks (Dorweiler et al. 2000; Parkinson et al. 2007; Erhard et al. 2009; Barber et al. 2012; Sloan et al. 2014) although the penetrance in multiple backgrounds has not been well characterized. We sought to determine if the subtle changes in DNA methylation in these mutants would reveal genes or transposons that are sensitive to these shifts in DNA methylation or chromatin. A limited number of genes and transposon families have altered transcript abundance in these genotypes, with changes representing both direct effects of the mutated genes and indirect effects of direct targets. A subset of differentially expressed genes have high levels of DNA methylation in wild-type that are reduced in the mutant genotypes. Many of these genes exhibit natural variation for DNA methylation and transcript abundance. This provides evidence that the natural variation at these genes is due to epigenetic rather than genetic variation and highlights cryptic information present in the maize genome that could be accessible through alterations to the epigenome.

Plants with loss-of-function alleles for mop1, mop3/rmr6/rpd1, zmet2 and zmet5 are viable but can exhibit some morphological abnormalities (Dorweiler et al. 2000; Papa et al. 2001; Makarevitch et al. 2007; Parkinson et al. 2007; Erhard et al. 2009). The mop1-1 and/or mop3-1 mutations have been shown to play important roles in the regulation of specific maize loci (Dorweiler et al. 2000; Allen et al. 2006; Sloan et al. 2014), transgenes (McGinnis et al. 2006) or transposable elements (Lisch et al. 2002; Woodhouse et al. 2006). Microarray profiling of gene expression has revealed evidence for altered expression of small sets of genes in studies of mop1-1 (Madzima et al. 2014) and zmet2-m1 (Makarevitch et al. 2007). There is also evidence from a RNA-seq experiment for altered regulation of transposable element expression in apical meristem tissue (Jia et al. 2009). Transcriptome analysis of rmr6-1, representing a distinct mutant allele at the mop3/ rmr6/rpd1 locus, provided evidence for a potential role in stress response (Forestan et al. 2016). The rmr6-1 mutation appears to increase the proportion of the genome that is transcribed but has subtle effects at most loci with relatively few genes with significant changes in expression level (Forestan et al. 2017). However, there have not been comprehensive studies on the overlap of genes or transposons that are sensitive to mutations in different CMT or RdDM genes in maize.

MATERIALS AND METHODS

Biological materials

All the mutant and control samples used in this study are listed in Table S1 and SRA accession numbers are listed for each dataset. The zmet2-m1, zmet2-m2, and zmet5-m1 mutant alleles were backcrossed into B73 for at least 7 generations. The mop1-1 and mop3-1 plant materials have been previously described (Gent et al. 2014; Li et al. 2015a); the mop1-1 tissues were from plants that have been backcrossed into a B73 genetic background while mop3-1 plants represent an uncharacterized genetic background. In all cases, “mutant” samples are homozygous for the mutant allele and “control” samples are homozygous for the wild-type allele. Tissue for RNA and DNA isolations was collected from three biological replicates. Plants were grown in standard greenhouse conditions for 20 days to reach the V3 stage. The 2nd and 3rd leaves were collected individually for each seedling. The 2nd leaf was used to isolate DNA for genotyping, and the 3rd leaf was used for RNA isolation and sequencing. For each biological replicate, 4-6 seedlings were pooled.
Library preparation and sequencing

Total RNA was isolated using the TRIZol reagent following the manufacturer’s protocol. RNA was quantified using RiboGreen and 3 µg total RNA was used to construct libraries using TruSeq strand-specific kit (Illumina) following manufacturer’s suggestions. The final library was quantified using PicoGreen and twelve libraries were pooled per Illumina lane. Library quality was checked using Agilent Bioanalyzer. Sequencing was performed on HiSeq2500 using 2 × 50 bp mode.

Gene expression analysis

Trim_glore was used to trim low-quality base from the 3’ end of the reads, as well as to remove adapters. Reads that passed quality control were mapped to B73 version 4 genome (Jiao et al. 2017) using Tophat2 (Kim et al. 2013), allowing at most 1 mismatch (-N 1) and the expected inner distance between mate pairs of 200 bp (-r 200). Reads that are properly paired and uniquely mapped were filtered out using samtools (-f 0x0002 –q 50). HTSeq (Anders et al. 2015) was used to summarize the number of reads mapped to each V4 gene model with the union mode, generating a matrix of count values for each gene in each genotype.

Raw read counts were input into DEseq2 (Love et al. 2014) to perform differential expression analysis. Pair-wise comparisons were made between each mutant and the appropriate wild type. Genes with a FDR value of < 0.05 and log(FoldChange) > 1 were called differentially expressed (DE) genes. Detailed analysis was restricted to genes with consistent DE calls in at least two mutant contrasts in the same pathway (RdDM or CMT). Genes were considered expressed if at least 3 replicates in the libraries described had an RPM (reads per million) value >1.

TE expression analysis

B73v4 (Jiao et al. 2017) TE annotation was modified to remove helitrons and the file was resolved using RTrackLayer in R so that each base of the genome was assigned to only a single TE. Exon regions were masked from the TE file using Bedtools (Quinlan and Hall 2010) subtract. Gene annotations were added to this modified TE annotation file, and mapped reads were assigned to features using HTSeq (Anders et al. 2015). A custom script was used to read through the HTSeq sam output, assigning unique-mapping reads to individual TE elements and multi-mapped reads to TE families if mapped positions hit only a single TE family. Unique and multi-mapped reads were combined for per-family expression counts, and RPM values were calculated by normalizing to the number of gene reads plus TE family reads in each library. All reads mapped to gene annotations plus TE annotations were excluded from the TE expression analysis. Differentially expressed TE families were determined using DESeq2 (Love et al. 2014) using a log(FoldChange) cutoff of 1 and FDR adjusted p-value cutoff of 0.05. Detailed analysis was restricted to TE families with consistent DE calls in at least two mutant contrasts in the same pathway (RdDM or CMT).

WGBS data analysis

The WGBS datasets used in this study are detailed in Table S2 and SRA accession numbers for each sample are provided. One µg DNA was used to prepare libraries for whole-genome bisulftite sequencing using the KAPA library preparation kit. DNA was sheared to a peak between 200-250 bp. End repair was performed to make blunt-ended fragments, followed by adding base A to the 3’ end, and adapter ligation. Size selection was performed to enrich library with a size between 250-450 bp. Bisulfite conversion was then carried out using Zymo DNA methylation lightning kit according to user’s manual. Finally, the libraries were enriched using PCR amplification. Library quality was checked using the Agilent Bioanalyzer. Library quantification was performed with qPCR before sequencing. Sequencing was performed on HiSeq2000 with paired end 100 cycles.

Analysis was performed as previously described (Li et al. 2015b; Song et al. 2016). Read quality was checked with FASTQC, adapters and low-quality bases at the 3’ end of each read were trimmed using Trim_glore. The high quality reads were mapped to B73 V4 genome (Jiao et al. 2017) using BSMAp (Xi and Li 2009) allowing at most 5 mismatches. Only properly paired reads with unique mappings were kept and used for calling DNA methylation. Methylation calls were performed using the methratio.py script from BSMAP. Finally, DNA methylation in each context (CG, CHG, CHH) was summarized for each 100-bp non-overlapping tile of the 10 maize chromosomes.

DMR calling

DMRs were called using previously described criteria (Li et al. 2015b). Briefly, each 100-bp tile with > 6 CG/CHG sites, > 2X coverage and > 60% difference for CG/CHG were called as CG and/or CHG DMRs. CHH DMRs were called using the same coverage and site number criteria, but with a requirement for <5% CHH in one genotype and >25% CHH in another genotype, reflecting the low level of CHH methylation in the maize genome.

mCHH Islands

High CHH tiles were called genome-wide by requiring CHH methylation over 25% with at least 10 informative read counts per tile. Genes and TE families were considered to be associated with a mCHH island if within at least one high methylation tile was identified within the gene/TE or in the 2 kb region surrounding the gene/TE.

Data and reagent availability

Mutant lines are available upon request. All data used in this study are deposited at the NCBI Sequence Read Archive (SRA), and accession numbers for all libraries are listed in Table S1. Expression values, differential expression calls, list assignments, and DMR calls for all genes and TE families are listed in Tables S2 and S3, respectively. Unique mapping read counts and descriptors for each transposable element are listed in Table S4. Supplemental material available at Figshare: https://doi.org/10.25387/g3.6071000.

RESULTS

Alteration of gene and transposon expression in maize mutants with perturbed methylomes

RNA-seq was used to perform transcriptome profiling for maize lines carrying mutations in genes encoding CMT (this study) or RdDM components (Gent et al. 2014; Li et al. 2015a). Together these factors are expected to be responsible for the majority of CHG and CHH methylation in the maize genome. For CMT genes, three biological replicates of seedling leaf tissue were profiled for mutations in two different genes, with multiple alleles utilized for one of the genes (Table S1). In addition, for the RdDM genes we analyzed seedling leaf tissue for mop1-1 and mop3-1 (Li et al. 2015a), along with immature ear tissue for mop1-1 (Gent et al. 2014). Homozygous wild-type siblings were used as controls for RdDM mutants. The genetic background, read number and accession information for each sample is provided in Table S1.
The transcript abundance of individual genes was estimated from the RNA-seq data for each sample. Differentially expressed (DE) genes in each mutant line (relative to the appropriate control) were identified using DESeq2 followed by a requirement for a minimum of 2 fold-change and an FDR value of less than 0.05 (Table S2). The observed differences in gene expression in the mutant lines could be direct effects of the mutation on transcription abundance, indirect effects caused by direct targets, or could be the result of introgressions of linked loci that contain cis-regulatory variation. The number of genes in each 2 MB bin with differential expression was assessed throughout the genome (Figure S1). For mutations that were identified in one background and then backcrossed into B73 (zmet2-m1, zmet2-m2, zmet5-m1, mop1-1), there were often a cluster of DE genes surrounding the locus of the mutation itself. For the other mutation (mop3-1) that was not backcrossed into another genetic background, there is less evidence for expression changes at linked genes (Figure S1). Based on these results we omitted DE genes located within 40 MB of the mutations in subsequent analyses.

A principal component analysis was performed using all DE genes to cluster samples used in this study (Figure 1). When combining all samples, PC1 separates ear transcriptomes from seedling leaf transcriptomes, and PC2 separates mop3-1 samples from mutants and controls in the B73 background, suggesting that tissue and genotype represent the strongest sources of differences in our transcriptomes (Figure 1A). When comparing only leaf libraries in the B73 background, few genotypes were substantially different from the wild-type controls, suggesting limited changes to transcript levels induced by each mutation (Figure 1B). The number of differentially expressed genes in each mutant genotype relative to tissue and genotype-matched controls was highly variable (Figure 1C). In most cases the homozygous mutant individuals exhibit more up-regulated genes than down-regulated genes, which is compatible with the concept that the CMT and RdDM genes normally provide silencing activities.

Transposable elements (TEs) comprise a large portion of the maize genome, and typically have high levels of CG and CHG methylation, with CHH methylation peaks at the edges of some TE families. There are two classes of TEs, Class I (retrotransposons) and Class II (DNA transposons), which transpose either through a copy-and-paste mechanism requiring an RNA intermediate (retrotransposons) or through a cut-and-paste mechanism (DNA transposons) (Wicker et al. 2007). Within each class are several orders divided into superfamilies, distinguished by structural and protein-coding features. Families within each superfamily are defined by sequence identity, and each family can contain many individual TE elements (Jiao et al. 2017). Individual TE elements are defined at a single location within a genome and are associated with a family, superfamily and class. We sought to document how perturbations to the methylome impacted expression of TEs. Due to the highly repetitive nature of TE sequences, we assessed per-family levels of expression by mapping RNA-seq reads to the genome, reporting up to 20 best hits for each read using TopHat2. Per-family read counts were determined by summing unique mapping reads (to a specific element) and multi-mapping reads that align to only a single TE family. Overall, the total portion of RNA-seq reads that map to TE families is not significantly higher in the mutants than in wild-type plants suggesting a lack of genome-wide activation of TEs in these mutants (Figure S2). In order to assess expression of individual TE families in each genotype, per-family expression was normalized by dividing the family counts by the total number of reads in the library assigned to either TE families or genes, generating a reads per million (RPM) estimate. Using this approach, we were able to detect expression of 1,694 TE families in at least one of the genotypes used for this study. A relatively small number of DE TE families (log₂FC > 1, FDR < 0.05) were identified in each mutant (Table S3; Figure 1D). Consistent with the role of DNA methylation in silencing TEs, more families were identified as up-regulated rather than down-regulated in mutants compared with WT controls. However, the majority of the TE families expressed in these libraries do not exhibit significant changes in expression level in CMT or RdDM mutants in maize.

There is a significant overlap in the number of genes and TEs that exhibit consistent changes in gene expression in at least two samples of CMT mutants (zmet2-m1/zmet2-m2/zmet5-m1) or RdDM mutants (mop1-1/mop3-1) (Figure S3). In order to understand the reproducible effects of these pathways on expression, we focused our analyses on the set of 237 genes and 104 TE families that exhibit consistent up- or down-regulation in multiple CMT or RdDM mutants. Hierarchical clustering of the fold-change of these genes or TE families relative to WT in all samples reveals a relatively consistent magnitude of expression changes between similar genotypes (Figure 1E-F). Within CMT mutants, 112 genes show shared expression changes between zmet2-m1 and zmet2-m2, with a smaller subset of genes primarily shared between zmet2-m2 and zmet5-m1 (35 genes, Figure 1E). Although both zmet2-m1 and zmet2-m2 are predicted to encode loss of function alleles, there are genetic differences in the behavior of these alleles (Papa et al. 2001; Makarevitch et al. 2007; Li et al. 2014). The zmet2-m1 mutation exhibits co-dominant effects on CHG methylation levels that may reflect dominant negative action of the protein that could be produced from this allele (Papa et al. 2001). Plants that are homozygous for zmet2-m1 have the greatest loss of CHG methylation and this could result from influence of the ZMET2-M1 protein product of the zmet2-m1 allele on functional ZMET5 protein.

The mop1-1 and mop3-1 seedling leaf samples have a number of examples of consistent up-regulation (26 genes) but fewer examples of consistent down-regulation (7 genes), consistent with the greater number of up-regulated than down-regulated genes in RdDM mutants in general. An analysis of the RNA-seq data for rmr6-1 (Forestan et al. 2017), which represents a distinct mutant allele at the mop3 locus, identified very similar numbers of DE genes and similar overlaps with other RdDM mutants (Figure S3). A portion of the “genes” that are affected by CMT or RdDM mutations may represent mis-annotated transposons or pseudogenes. We assessed the proportion of genes with altered transcript abundance that are classified as syntenic based on a comparison with sorghum and rice genomes (Brohammer et al. 2018). The genes that are mis-regulated in RdDM or CMT mutants are often non-syntenic, but in all cases there are also a number of examples of syntenic genes (Figure S4). This suggests that least a portion of the targets of CMT and RdDM represent potentially functional genes.

**Some genes that are up-regulated in CMT mutants exhibit high CHG methylation levels**

The differential expression observed in each mutant background could result from direct changes in DNA methylation or chromatin at these loci or could result from indirect effects due to secondary effects from genes that are direct targets. Genes that are direct targets for silencing by DNA methylation would be expected to have high levels of DNA methylation in wild-type plants. The context-specific DNA methylation profiles were assessed in wild-type B73 for genes that were up- or down-regulated compared with all expressed genes (Figure 2A). The genes that are differentially expressed in RdDM mutants do not show unusual patterns of DNA methylation. The methylation profile for genes that
are down-regulated in CMT mutants are similar to all expressed genes. In contrast, genes that are up-regulated in the CMT mutants exhibit distinct patterns of CG and CHG methylation within gene bodies relative to other expressed genes (Figure 2A). Among the 112 genes up-regulated in multiple CMT mutants, approximately half have high (>50%) and half have low (<20%) methylation in the CG and CHG contexts in the region surrounding the transcription start site (TSS) (Figure 2B-C). In contrast, only ~4% of all expressed genes have high CG and CHG methylation in the same region. While a small number of genes with high methylation overlap annotated TEs, most of the genes in this subset do not, suggesting that this genic methylation is not solely due to nearby TEs. In wild-type samples, genes with high CHG methylation in the gene body also have high CG methylation (Figure 2D). However, in zmet2-m1 mutants, CHG methylation for these genes is reduced, with few examples having any corresponding reduction in CG methylation (Figure 2D-E). As in the examples (Zm00001d045627 and Zm00001d021982) shown in Figure 2F, these genes have high levels of CG and CHG methylation in wild-type B73, and the reduction in CHG methylation in zmet2-m1 mutants is associated with increased expression, suggesting that CMT-dependent silencing of these genes depends on CHG but not CG methylation.

Genes that are up-regulated in RdDM mutants are enriched for being near mCHH islands
A large number of maize genes (~60%) have been associated with the presence of a region of elevated CHH in the promoter region, termed a mCHH island (Gent et al. 2013). Genes with mCHH islands are enriched for high expression and the mCHH island often occurs at the edge of the TE nearest these genes (Li et al. 2015a). These mCHH islands may form important boundaries that could protect TE heterochromatin from the influence of genes (Li et al. 2015a) and may also be important for long-distance chromatin interactions (Rowley et al. 2017). The methylation within these mCHH islands requires Mop1 and Mop3 (Li et al. 2014, 2015a). Genes that are up-
down-regulated in the RdDM mutants are enriched for the presence of mCHH islands, but this is only significant for the up-regulated genes with 87.8% having a mCHH island within 2kb of the gene, compared to 64.6% of all expressed genes (Figure 2G, p-value < 0.01, chi squared test). The observation that both RdDM up- and down-regulated genes are often near mCHH islands could be due to the fact that the mCHH island may provide long-range interactions (Rowley et al. 2017) that could have either positive or negative influences on gene expression.

In some cases, the mCHH island itself may result in transcriptional regulation. Work in *Arabidopsis* has noted a positive feedback loop involving DNA methylation levels and expression of the demethylase ROS1 such that reduced levels of DNA methylation result in lower ROS1 expression but increased methylation is associated with elevated ROS1 expression (Williams et al. 2015). Reduced transcript abundance of maize DNA glycosylases has also been observed in several transcriptome datasets of maize RdDM mutants (Williams et al. 2015; Erhard et al. 2015). We find that one maize gene with sequence...
homology to ROS1, Zm00001d038302, showed significantly reduced transcript abundance in the mop1-1 and mop3-1 mutants and has a strong mCHH island in several inbred lines (Figure S5). This provides evidence to support a requirement for RdDM and CHH methylation in the proper transcriptional control of this gene in maize.

**Genes regulated by CMT are enriched for natural DMRs**

There is a substantial amount of natural variation for methylation levels among maize inbreds. We sought to determine the extent to which genes that are sensitive to CMT or RdDM mutations also display variable methylation levels in these lines. These would reflect examples of potential cryptic information in one inbred line that is expressed in other genotypes, reflecting natural epigenetic variation. We used WGBS data from B73 and 17 other diverse maize inbreds to document natural variation for DNA methylation among maize inbreds. Differentially Methylated Regions (DMRs) were identified in all three contexts (CG, CHG, and CHH) between B73 and the other inbreds. More than 200,000 DMRs were called in the CG and CHG contexts, with over 50,000 DMRs in the CHH context. Each maize gene was classified based on whether there was a DMR within 200 bp up or downstream of the transcription start site for each of the three sequence contexts. A relatively small portion (~3-7%) of maize genes has CG, CHG or CHH DMRs near the promoter (Table S2). We proceeded to assess whether naturally occurring DMRs were more prevalent near genes that exhibit altered expression in CMT or RdDM mutants (Figure 3A). Genes that are up- or down-regulated in CMT mutants exhibit a significant enrichment for CG and CHG DMRs in their promoter regions. RdDM up-regulated genes are significantly enriched for having CHH and CHG DMRs and also show an enrichment (though not significant) for CG DMRs near the promoter (Figure 3A).

This suggests that many of the genes with altered transcript abundance in CMT or RdDM mutants may have pre-existing natural variation for DNA methylation that would affect expression levels in maize populations. RNA-seq data from leaf tissue for ten of the inbred lines with WGBS data were utilized to determine whether there was a significant association (p-value < 0.05, Pearson correlation) between context-specific methylation level at the DMR and transcript abundance levels. We found that nearly 50% of the genes that exhibit altered expression in CMT or RdDM mutants that are located near DMRs had natural variation for gene expression levels that was significantly associated with DNA methylation levels. Two examples of CMT up-regulated genes that exhibit significantly correlated transcript abundance and CHG methylation at the bin overlapping the TSS among diverse lines are shown in Figure 3B-C. In wild-type maize inbred lines we see two classes with respect to transcript abundance and CHG methylation levels at the DMR near the TSS. In one group of lines, including B73 (red dots), the DMR is highly methylated and the gene is transcriptionally silent. In the other group of genotypes (and in B73 zmet2-m1 mutant lines - blue dots) the DMR has low methylation and transcripts are observed.

**Properties of TEs with altered expression**

There are 104 TE families with altered expression in mutants that perturb RdDM or CMT components in maize. These included 32 families up- and 7 families down-regulated in at least two contrasts of RdDM mutants, and 48 families up- and 24 families down-regulated in at least two contrasts of CMT mutants (Figure 1D). There are examples of both class I (specifically Long Terminal Repeat orLTR) and class II (specifically Terminal Inverted Repeat orTIR) TE families that exhibit altered expression in both RdDM and CMT mutants (Figure 4A). Most families with varied expression were small (< 10 members), consistent with the genome-wide distribution (Figure 4B). The analysis of the relative age of LTR families, approximated by comparing the sequence similarity of the two LTR sequences of an element, that have altered expression reveals that LTR elements up-regulated in RdDM mutants are enriched for younger LTR elements when compared with the distribution of ages present genome-wide (Figure 4C). The TEs that exhibit increased transcript abundance in RdDM mutants tend to be located further from genes than TEs that are expressed in CMT mutants or all expressed TEs (Figure S6; p-value < 0.01, t-test). We also tested the mean GC content of TEs within families to test whether families depleted in cytosines are more susceptible to subtle perturbations in methylation, as is the case for the ONSEN family in Arabidopsis (Cavrak et al. 2014). TE families up-regulated in RdDM mutants do have a slightly lower GC content on average, though it is not clear if this alone is sufficient to cause the expression changes (Figure 4D, p-value < 0.01, t-test).

We sought to further document the properties of these TE families through analysis of their transcript abundance in nearly 100 developmental tissues or stages of B73. During typical development, approximately 3,400 TE families are expressed in at least one tissue or stage. There are 5 TE families up-regulated in RdDM mutants and 18 TE families up-regulated in CMT mutants that are not expressed in any tissue or developmental stage assessed. The other families of TEs that are up-regulated in RdDM mutants (25 families) or CMT mutants (27 families) were assessed to determine if they exhibit distinct patterns of expression. Interestingly, approximately one third of the TE families that are up-regulated in RdDM mutants show higher expression in the endosperm than other tissues (Figure 5A). In contrast, the TE families up-regulated in CMT mutants do not show any evidence for higher expression in a particular tissue. The enrichment for endosperm
expression in TE families up-regulated in RdDM mutants does not extend to genes up-regulated in the mutants and cannot be simply attributed to lower expression of the Mop1 and Mop3 genes in these tissues (Figure 5). This result highlights the potential for some TEs to escape RdDM-based silencing in endosperm, where dynamic changes to DNA methylation may reinforce TE silencing in the embryo (Martínez and Slotkin 2012; Gehring 2013; Wang et al. 2015; Dong et al. 2017). Meanwhile, both genes and TE families susceptible to mis-regulation in CMT mutants are less often expressed across development, consistent with the greater developmental stability of CHG methylation over CHH methylation (Kawakatsu et al. 2016, 2017; Narsai et al. 2017; Bouyer et al. 2017).

Evidence for locus-specific and coordinated changes in expression of TEs

While per-family analysis of TE expression is useful in capturing expression dynamics of repetitive transposable elements, the expression of individual elements can be influenced by a variety of location-specific attributes such as methylation levels and proximity to genes as well as family-level attributes such as binding motifs and nucleotide content. We were interested in documenting the relative behavior of different elements within the same family to understand whether the changes in transcript abundance of TEs were occurring in an element-specific or family-wide manner. Coordinate changes in transcript abundance could indicate the importance of RdDM or CMT for family-wide regulation while element-specific changes could reflect influences at particular loci. A set of TE families with <10 elements that had altered transcript abundance and for which at least 50% of the reads could be uniquely assigned to specific element were identified and used for analysis of coordinate vs. locus-specific expression (Tables S3, S4). The unique mapping reads for these families were used to evaluate element-specific transcript abundance. Half of the testable TE families had expression of a single member of the family indicating locus-specific changes (examples in Figure 6A, C). In the other half of the TE families there was evidence for expression changes for multiple members of the family by CMT or RdDM pathways. However, even in examples of coordinate expression a single element accounted for the vast majority of unique reads mapping to the family. Examples of both locus-specific and coordinate changes in expression for both CMT and RdDM mutants were found but we were not able to assess enough families to determine if there was any enrichment for the type of regulation for these two silencing pathways.

DISCUSSION

Maize has been a model system for the discovery of several epigenetic phenomena such as imprinting (Kermicle 1970; Kermicle and Allemann 1990), paramutation (Brink 1956; Chandler 2007; Hollick 2017) and transposon silencing (Chandler and Walbot 1986; Chomet et al. 1987). An unresolved question is whether epigenetic regulation plays important roles in quantitative trait variation beyond handful of well characterized loci. Our ability to document the full role for epigenetic regulation and DNA methylation has been limited by our inability to recover plants with major reductions in the level of DNA methylation (Li et al. 2014). Forward genetic screens have uncovered a number of components of the RNA-directed DNA methylation (RdDM) machinery as playing critical roles in maintenance of silenced paramutant states (Dorweiler et al. 2000; Erhard et al. 2009) or transgene silencing (McGinnis et al. 2006). These mutants have substantial effects on CHH methylation in maize but have minimal effect on genome-wide levels of CG or CHG methylation (Li et al. 2014). Reverse-genetic analyses have identified loss-of-function alleles for a number of other genes predicted to play important roles in DNA methylation but the only single

![Figure 4](image-url)
mutants with significant effects on genome-wide DNA methylation are the CMT genes of maize, *zmet2* and *zmet5* (Li et al. 2014). In this study we have documented how these subtle perturbations of the maize methylome affect the transcriptome in order to find genes subject to epigenetic regulation.

The effects of mutations in RdDM or CMT genes in maize are quite limited. Our evidence suggests that there is little effect on the overall transcriptome of these plants. This might be expected given the limited effect on overall plant phenotype for each of these mutations. A recent study found that *rmr6-1* (allelic to *mop3-1*) mutants exhibited transcription changes from a larger portion of the genome but much of this was associated with increased transcriptional 'noise' at lowly expressed regions (Forestan et al. 2017). Reanalysis of gene and TE family expression using our methods in *rmr6-1* mutant and wild-type libraries revealed that the number of differentially expressed genes and TEs is quite similar to the mutants used in our study (Figure S3). Many of the genes or TEs with altered expression in *rmr6-1* also exhibit altered expression in *mop3-1* and/or *mop1-1* (Figure S3). While relatively few genes exhibit major changes in expression there are a significant number of genes that exhibit similar expression changes in multiple RdDM or CMT mutants. These findings are compatible with the concept that there are a small number of genes in the maize genome that have epigenetic regulation that is solely dependent upon RdDM or CMT mediated regulation. While some of these “genes” lack synteny with other related species others do show syntenic genomic positions and likely represent potentially functional genes (Figure S4). It is likely that a much larger number of genes are redundantly regulated by the RdDM and CMT pathway along with MET1 mediated CG methylation.

DNA methylation is often considered to play a primary role in maintaining genome integrity by silencing transposable elements. Indeed, there are clear examples of release of transposon silencing in mutants affecting DNA methylation in *Arabidopsis* (Miura et al. 2001; Mirouze et al. 2009) and maize (Lisch et al. 2002; Jia et al. 2009).
However, generating a complete understanding of transposon expression is complicated by the highly repetitive nature of transposable elements. In order to survey expression using RNA-seq most researchers focus on unique mapping reads to ensure that expression is accurately attributed to the proper genomic locus. In this study we elected to primarily focus on TE families rather than individual elements and we utilized an approach that allowed for the combined use of unique and multiple-mapping reads to assess TE family expression. We did find evidence that a number of TE families require RdDM and/or CMT for silencing. There were few distinguishing features about these TEs relative to others making it unclear why the silencing of these families was easily released in these mutants. For several families we were able to document evidence for specific release of silencing of a single member of the family while in other cases we found that multiple members of the same family were reactivated.

This study defines a set of genes and TE families that are regulated by DNA methylation. The silencing of these genes and TEs relies solely upon RdDM or CMT based epigenetic regulation. These loci provide important insights into the mechanisms that allow for epigenetic regulation and the natural variation for epigenetic regulation in maize.

ACKNOWLEDGMENTS
We thank Peter Hermanson and Jonathan Giesler for technical assistance and Jonathan Gent for helpful discussions and feedback. Michelle Stitzer and Jeffrey Ross-Ibarra provided helpful discussions and transposon resources. Funding for this work was grants from USDA-NIFA2016-67013-24747 to N.M.S. and the National Science Foundation IOS-1237931 to N.M.S. and M.W.V. Q.L. was supported by The National Key Research and Development Program of China (2016YFD0101003).

Figure 6 TE families up-regulated in methylation mutants can have expression of a single element or multiple elements. A-B show two LTR families up-regulated in RdDM mutants, and C-D show two LTR families up-regulated in CMT mutants. All families have both unique (red) and multi-mapped (blue) reads. The proportion of unique-mapping reads assigned to each element is shown for a representative library. Unique-mapping reads showed confident expression of only a single member of a family (A and C) or coordinated expression of more than one member of a family (B and D).

LITERATURE CITED
Alleman, M., L. Sidorenko, K. McGinnis, V. Seshadri, J. E. Dorweiler et al., 2006 An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 442(7100): 295–298. https://doi.org/10.1038/nature04884
Anders, S., P. T. Pyl, and W. Huber, 2015 HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2): 166–169. https://doi.org/10.1093/bioinformatics/btu638
Barber, W. T., W. Zhang, H. Win, K. K. Varala, J. E. Dorweiler et al., 2012 Repeat associated small RNAs vary among parents and following hybridization in maize. Proc. Natl. Acad. Sci. USA 109(26): 10444–10449. https://doi.org/10.1073/pnas.1202073109
Bewick, A. J., C. E. Niederhuth, L. Ji, N. A. Rohr, P. T. Griffin et al., 2017 The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol. 18(1): 65. https://doi.org/10.1186/s13059-017-1195-1
Bouyer, D., A. Kramdi, M. Kassam, M. Heese, A. Schnittger et al., 2017 DNA methylation dynamics during early plant life. Genome Biol. 18(1): 179. https://doi.org/10.1186/s13059-017-1313-0
Brink, R. A., 1956 A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. Genetics, 41: 872–889.
Brohammer, A. B., T. J. Y. Kono, N. M. Springer, S. E. McGaugh, and C. N. Hirsch, 2018 The limited role of differential fractionation in genome content variation and function in maize (Zea mays L.) inbred lines. Plant J. 93: 131–141.
Matzke, M. A., and R. A. Mosher, 2014 RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. 15(6): 394–408 (erratum: Nat. Rev. Genet. 15: 570). https://doi.org/10.1038/nrg3683

McGinnis, K. M., C. Springer, Y. Lin, C. C. Carey, and V. Chandler, 2006 Transcriptionally silenced transgenes in maize are activated by three mutations defective in paramutation. Genetics 173(3): 1637–1647. https://doi.org/10.1534/genetics.106.058669

Miouze, M., J. Reinders, E. Bucher, T. Nishimura, K. Schneeberger et al., 2009 Selective epigenetic control of retrotransposition in Arabidopsis. Nature 461(7262): 427–430. https://doi.org/10.1038/nature08328

Miura, A., S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada et al., 2001 Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature 411(6834): 212–214. https://doi.org/10.1038/35075612

Narsai, R., Q. Gouil, D. Secco, A. Srivastava, Y. V. Karpievitch et al., 2017 Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. Genome Biol. 18(1): 172. https://doi.org/10.1186/s13059-017-1302-3

Papa, C. M., N. M. Springer, M. G. Muszynski, R. Meeley, and S. M. Kaeppler, 2001 Maize chromomethylase Zea methyltransferase2 is required for CnPpG methylation. Plant Cell 13(8): 1919–1928. https://doi.org/10.1105/tpc.13.8.1919

Parkinson, S. E., S. M. Gross, and J. B. Hollick, 2007 Maize sex determination and abaxial leaf fates are canalized by a factor that maintains repressed epigenetic states. Dev. Biol. 308(2): 462–473. https://doi.org/10.1016/j.ydbio.2007.06.004

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6): 841–842. https://doi.org/10.1093/bioinformatics/btp033

Regulski, M., Z. Lu, J. Kendall, M. T. Donoghue, J. Reinders et al., 2013 The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res. 23(10): 1651–1662. https://doi.org/10.1101/gr.153510.112

Rowley, M. J., M. H. Rothi, G. Böhmdorfer, J. Kuciks, and A. T. Wierzbicki, 2017 Long-range control of gene expression via RNA-directed DNA methylation. PLoS Genet. 13(5): e1006749. https://doi.org/10.1371/journal.pgen.1006749

Sloan, A. E., L. Sidorenko, and K. M. McGinnis, 2014 Diverse gene-silencing mechanisms with distinct requirements for RNA polymerase subunits in Zea mays. Genetics 198(3): 1031–1042. https://doi.org/10.1534/genetics.114.168518

Song, J., G. Zynda, S. Beck, N. M. Springer, and M. W. Vaughn, 2016 Bisulfite Sequence Analyses Using CyVerse Discovery Environment: From Mapping to DMRs, Current Protocols in Plant Biology, John Wiley & Sons, Inc., Hoboken, NJ. https://doi.org/10.1002/cppb.20034

Williams, B. P., D. Pignatta, S. Henikoff, and M. Gehring, 2015 Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. PLoS Genet. 11(3): e1005142. https://doi.org/10.1371/journal.pgen.1005142

Xi, Y., and W. Li, 2009 BSMAP: whole genome bisulfite sequence MAPPing program. BMC Bioinformatics 10(1): 232. https://doi.org/10.1186/1471-2105-10-232

Communicating editor: J. Birchler