

RESEARCH ARTICLE

Dynamic Roles for Small RNAs and DNA Methylation during Ovule and Fiber Development in Allotetraploid Cotton

Qingxin Song¹, Xueying Guan^{1,2}, Z. Jeffrey Chen^{1,2*}

1 Department of Molecular Biosciences, Institute for Cellular and Molecular Biology, Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, Texas, United States of America, **2** State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China

* zjchen@austin.utexas.edu



OPEN ACCESS

Citation: Song Q, Guan X, Chen ZJ (2015) Dynamic Roles for Small RNAs and DNA Methylation during Ovule and Fiber Development in Allotetraploid Cotton. PLoS Genet 11(12): e1005724. doi:10.1371/journal.pgen.1005724

Editor: Hume Stroud, Harvard Medical School, UNITED STATES

Received: July 17, 2015

Accepted: November 14, 2015

Published: December 28, 2015

Copyright: © 2015 Song et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All RNA-seq, small RNA-seq, and methylome files are available from in the Gene Expression Omnibus (GEO) database (accession no. GSE61774).

Funding: This work was funded by: National Science Foundation Plant Genome Research Program (IOS1025947 and IOS1444552); and Cotton Incorporated (07-161 and 14-371). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

DNA methylation is essential for plant and animal development. In plants, methylation occurs at CG, CHG, and CHH (H = A, C or T) sites via distinct pathways. Cotton is an allotetraploid consisting of two progenitor genomes. Each cotton fiber is a rapidly-elongating cell derived from the ovule epidermis, but the molecular basis for this developmental transition is unknown. Here we analyzed methylome, transcriptome, and small RNAome and revealed distinct changes in CHH methylation during ovule and fiber development. In ovules, CHH hypermethylation in promoters correlated positively with siRNAs, inducing RNA-dependent DNA methylation (RdDM), and up-regulation of ovule-preferred genes. In fibers, the ovule-derived cells generated additional heterochromatic CHH hypermethylation independent of RdDM, which repressed transposable elements (TEs) and nearby genes including fiber-related genes. Furthermore, CHG and CHH methylation in genic regions contributed to homoeolog expression bias in ovules and fibers. Inhibiting DNA methylation using 5-aza-2'-deoxycytidine in cultured ovules has reduced fiber cell number and length, suggesting a potential role for DNA methylation in fiber development. Thus, RdDM-dependent methylation in promoters and RdDM-independent methylation in TEs and nearby genes could act as a double-lock feedback mechanism to mediate gene and TE expression, potentiating the transition from epidermal to fiber cells during ovule and seed development.

Author Summary

Cotton is the world's largest source of renewable textile fiber and is an allotetraploid crop consisting of two progenitor genomes. In plants, *de novo* CHH (H = A, T, or C) methylation depends on RNA-directed DNA methylation (RdDM) and CHROMOMETHYLASE2 (CMT2)-mediated pathways. The biological significance of the two pathways is largely unknown. Here we show dynamic roles of these two pathways in ovule and fiber

development. RdDM-dependent CHH methylation is linked to gene activation in ovules, and additional CMT2-dependent methylation leads to silencing of transposons and nearby genes in fibers. Moreover, DNA methylation affects expression bias of homoeologous genes and fiber development. These findings provide novel insights into epigenetic regulation of organ development and polyploid evolution.

Introduction

DNA methylation, a conserved epigenetic mark in most eukaryotes, is essential for growth and development and is associated with many epigenetic phenomena, including imprinting and transposon silencing [1–5]. In plants, DNA is methylated in CG, CHG and CHH (H = A, T, or C) sites through distinct pathways. In *Arabidopsis*, CG methylation is maintained by METHYLTRANSFERASE1 (MET1), a homolog of mammalian DNMT1 [6]. Plant-specific CHROMOMETHYLASE3 (CMT3) is primarily responsible for CHG methylation, which is coupled with H3K9 dimethylation [7]. CHH methylation is established *de novo* by DOMAINS REARRANGED METHYLTRANSFERASE1 and 2 (DRM1 and DRM2) [8] through the RNA-directed DNA methylation (RdDM) pathway [9], involving 24-nt small interfering RNAs (siRNAs) [1, 2]. Recent studies found that CHH methylation could also be established by CMT2 [10, 11], through histone H1 and DECREASE-IN-DNA-METHYLATION1 (DDM1) activities [12], which is independent of the RdDM. The methylome data indicate that CMT2 and RdDM pathways preferentially function in heterochromatic and euchromatic regions, respectively [10, 11]. However, the role for DNA methylation in developmental regulation is poorly understood.

Cotton is the largest source of renewable textile fiber and an excellent model for studying the developmental transition from ovule epidermal cells to rapidly-elongating singular fiber cells. The most widely-cultivated cotton (*Gossypium hirsutum* L., AADD) is an allotetraploid species, which originated 1–2 million years ago from interspecific hybridization between A-genome species, resembling *Gossypium herbaceum* or *Gossypium arboreum*, and D-genome species, resembling *Gossypium raimondii* [13]. The intergenomic interaction in allotetraploid cottons induces longer fiber and higher yield, coincident with expression bias of fiber-related homoeologous genes [14, 15], which provides the basis of selection and domestication for agronomic traits in cotton and many other polyploid crops [13, 16]. Each cotton fiber is a single cell derived from the ovule epidermis, undergoing rapid cell elongation and cellulose biosynthesis, and ~100,000 fiber cells develop semi-synchronously in each ovule (seed) and can reach six centimeters in length [17, 18]. In early stages of fiber development, rapid cell growth is associated with a dramatic increase of DNA content by endoreduplication [19, 20] and dynamic changes in gene expression and small RNAs [15, 18, 21]. Interestingly, DNA methylation changes are related to seasonal variation of fiber development in cotton [22] and is also shown to change among different tissues including fibers based on the methylation-sensitive high-performance liquid chromatography (HPLC) assay [23]. Moreover, over-expressing fiber-related transgenes often leads to the unexpected outcome of fiber phenotypes [24]. These data indicate a potential role for DNA methylation in gene expression and phenotypic traits such as cotton fiber, which could be selected and domesticated.

Although genome-wide DNA methylation has been examined in *Arabidopsis* [25], soybean [26, 27], maize [28, 29], and other plants and animals [30, 31], the roles of RdDM and CMT2-dependent methylation pathways in organogenesis and development remain elusive. In this study, we employed cotton ovule and fiber cells as a model to test the role of DNA

methylation in developmental regulation. Using methylcytosine-sequencing (MethylC-seq) [32, 33], RNA-seq, and small RNA-seq analyses, we examined CG, CHG, and CHH methylation patterns in fibers, ovules, and leaves and analyzed differentially methylated regions (DMRs) between the ovule and leaf (OL) and between the fiber and ovule (FO). The methylation patterns in the gene body and 5' and 3' flanking sequences were comparatively analyzed with TE densities and expression levels of genes and small RNA loci. The results support unique roles of CG, CHG, and CHH methylation in ovule and fiber development and expression bias of homoeologous genes in the allotetraploid cotton.

Results

DNA methylation dynamics in cotton leaves, ovules, and fibers

DNA methylation affects growth and development in plants and animals [3, 4]. To investigate genome-wide DNA methylation changes during ovule and fiber development, we used allotetraploid cotton (*G. hirsutum* L. acc. TM-1) to perform whole-genome bisulfite sequencing in leaves, ovules at 0 DPA, and fibers at 14 DPA with two biological replicates and ovules at 14 DPA with one replicate. The TM-1 sequence of A and D subgenomes [34] was used as the reference for data analysis. The bisulfite-conversion rates were over 99% (S1 Table). Approximately 80% of cytosines were covered by at least one uniquely mapped read, and the mean coverage of cytosines was 8.2-fold or higher for all tissues. Methylation levels between biological replicates were highly correlated among symmetric sites (Pearson $r > 0.9$) and all sites ($r > 0.8$), indicating reproducibility of the data (Figs 1A and S1A). Overall, CG and CHG methylation levels were similar among fibers, ovules and leaves but were lower in the D subgenome than in the A subgenome for all tissues (Fig 1A). However, CHH methylation levels were much higher in fibers (~14%) than in ovules (~8.1%) and leaves (~7.8%) (Fig 1A). This high methylation level in fibers was not related to the late developmental stage because the methylation level in ovules was slightly lower at 14 DPA than at 0 DPA (S1B Fig). CHH methylation is induced by small RNAs through RdDM [2, 35]. To test a link of DNA methylation changes in cotton tissues with small RNAs, we generated small RNA-seq data with three replicates in leaves, ovules at 0 DPA, and fibers at 14 DPA (S1 Table). The data showed 24-nt siRNAs were much higher in fibers and ovules than in leaves (Fig 1B) [15].

On each chromosome, DNA methylation was more abundant in transposable element (TE)-rich than gene-rich regions in all tissues (Figs 1C and S1C). However, 24-nt siRNAs were highly enriched in gene-rich regions (Figs 1C and S1C), suggesting a role for these siRNAs in genic methylation. Because each fiber cell is expanded from an epidermal cell of the ovule, further comparison was made between the ovule and leaf (ovule-leaf, OL) and between the fiber and ovule (fiber-ovule, FO) methylation levels, which represented methylation changes in the ovule and fiber, respectively. OL CHH hypermethylation correlated with gene-rich regions, showing the same trend as that of the siRNAs (Fig 1C). On the contrary, FO CHH hypermethylation was enriched in TE- and repeat-rich regions, characteristic of heterochromatin. The data indicate that OL CHH hypermethylation preferentially occurs in euchromatic regions, whereas FO CHH hypermethylation predominates in heterochromatic regions (Figs 1C and S1D).

CHH methylation differences between the ovule and leaf occurred mainly in 5' upstream and 3' downstream sequences of the genes (Fig 1D), while CG and CHG methylation levels in genes and TEs were similar among different tissues tested (S2 Fig). Among all TEs, mean CHH methylation levels were higher in the fiber but indistinguishable between the ovule and leaf (Fig 1E). The higher CHH methylation levels of TEs in fibers and ovules were correlated with the closer distances of TEs to the gene (Fig 1F). When TEs were more than 4-6-kb away from

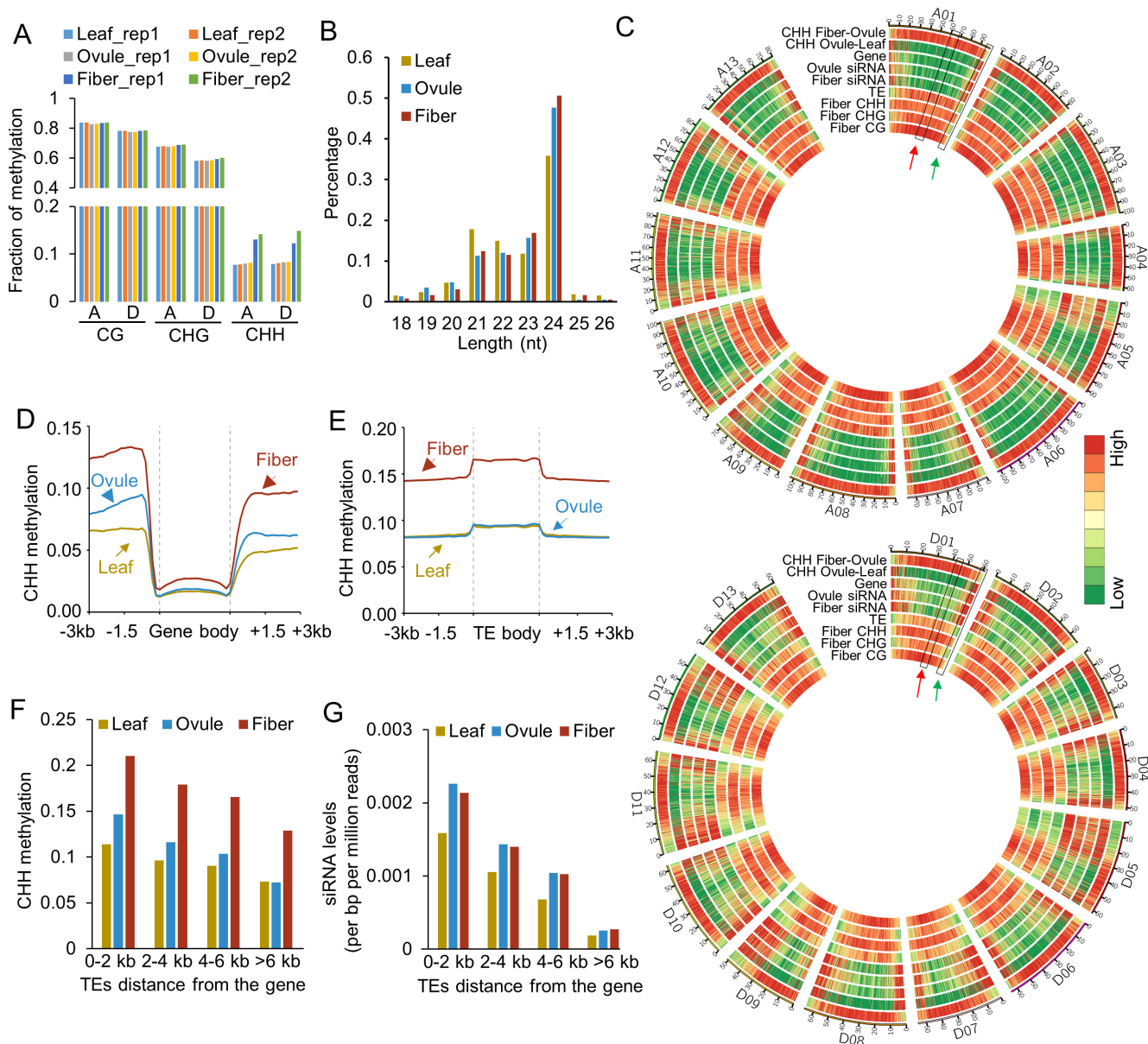


Fig 1. Effects of DNA methylation on ovule and fiber development and genome-wide distribution of DNA methylation in leaves, ovules, and fibers. (A) Percentage of methylated cytosine (mC) in leaves, ovules (0 DPA), and fibers (14 DPA). A and D indicate A- and D subgenomes in the allotetraploid cotton, respectively. (B) Size distribution of small RNAs in different tissues. (C) Circle plots of fiber CG, CHG, CHH methylation, fiber small RNA density, gene density, TE density, and ovule-leaf and fiber-ovule CHH methylation using 100-kb windows among 13 A-homoeologous chromosomes (upper) and 13 D-homoeologous chromosomes (lower). Color bars indicate low (green) to high (red) densities or intensities. The scales are: fiber CG and CHG methylation level: 0–1; fiber CHH methylation level: 0–0.18; TE density: 0–70%; Gene density: 0–20%; siRNA density in fiber and ovule: 0–60 reads per million reads; Ovule-fiber CHH methylation: 0–0.01; fiber-ovule CHH methylation: 0–0.07. (D) Distribution of CHH methylation in genes. (E) Distribution of CHH methylation in TEs. (F) CHH methylation levels in TEs relative to the distance from the nearest gene. (G) siRNA levels (per bp per million reads) in TEs relative to the distance from the nearest gene.

doi:10.1371/journal.pgen.1005724.g001

the gene, CHH methylation levels became similar between the ovule and leaf. Consistently, the CHH methylation levels mirrored the 24-nt siRNA distribution patterns, which were high near the genes and decreased as they were further away from the gene, indicating the role of the RdDM pathway in CHH methylation of the gene nearby TEs (Fig 1G).

Differentially methylated regions between the ovule and leaf and the fiber and ovule

As CG and CHG methylation levels were similar among different tissues examined, further analysis was focused on the CHH methylation changes among these tissues. We predict that differentially methylated regions (DMRs) between tissues play a role in biological function. To test this, we identified 39,668 CHH-hypermethylated DMRs in the ovule relative to leaf (ovule-leaf) (OL CHH-hyper DMRs) and 124,681 CHH-hypermethylated DMRs in the fiber relative to ovule (fiber-ovule) (FO CHH-hyper DMRs) (S2 Table). A subset of these DMRs was validated by bisulfite-sequencing individually cloned genomic fragments (S3 Fig), confirming the results of these DMRs based on the genome-wide analysis. Compared to distributions of exons and introns in genomic distribution, both OL and FO CHH-hyper DMRs were underrepresented in exons and introns (Z-score < -5) (Fig 2A). However, OL CHH-hyper DMRs were overrepresented in intergenic regions (62%) relative to the average genomic distribution (40%), whereas FO CHH-hyper DMRs were enriched in TEs (80% vs. 50%). Among TEs, short TEs (less than 1-kb) were more abundant in OL CHH-hyper DMRs (74%) than in the genomic

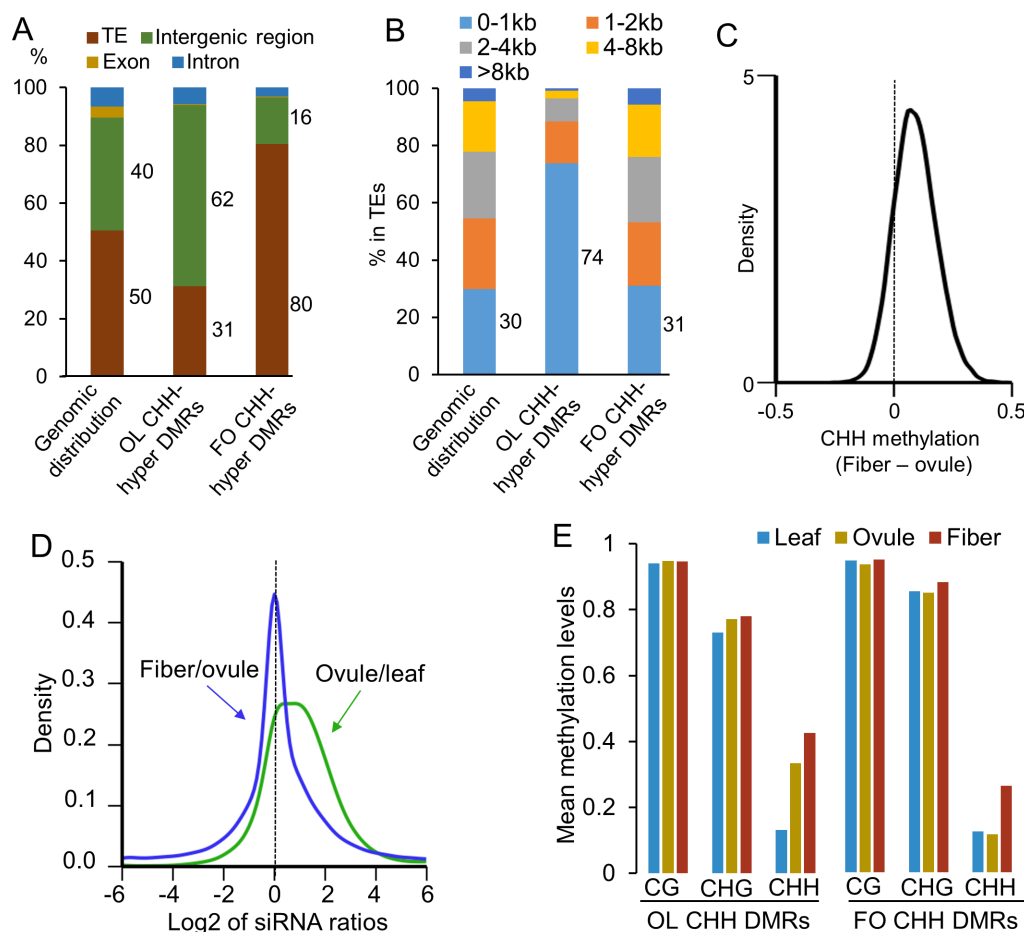


Fig 2. Genomic distribution of differentially methylated regions (DMRs). (A) Percentage of DMRs in TE (brown), intergenic region (green), exon (dark yellow) and intron (blue). (B) Percentage of DMRs corresponding to TEs with different sizes. (C) Kernel density plot of CHH methylation change of fiber/ovule in OL CHH-hyper DMRs. (D) Kernel density plot showing 24-nt siRNA fold-changes of ovule/leaf in OL CHH-hyper DMRs (green) and of fiber/ovule in FO CHH-hyper DMRs (blue). (E) Percentage of CG, CHG and CHH methylation in the leaf (blue), ovule (yellow), and fiber (red) in OL CHH-hyper DMRs and FO CHH-hyper DMRs.

doi:10.1371/journal.pgen.1005724.g002

distribution (30%), whereas the percentage of short TEs in FO CHH-hyper DMR distribution was similar to the genomic distribution (Fig 2B). This suggests that OL hyper-CHH regions were located nearby genes (between genes with short TEs), whereas FO hyper-CHH regions are associated with TEs (genome-wide phenomenon).

Methylation levels of most OL CHH-hyper DMRs were higher in fibers than in ovules, indicating that OL CHH-hyper DMRs continue to be hypermethylated in elongating fibers after they developed from ovule epidermal cells ($P < 1e^{-200}$, Wilcoxon rank-sum test) (Fig 2C). Interestingly, among the OL CHH-hyper DMRs, siRNA expression levels were significantly higher in ovules than in leaves ($P < 1e^{-100}$, Wilcoxon rank-sum test), and this difference was not obvious between the fiber and ovule in FO CHH-hyper DMRs (Fig 2D). These data suggest that siRNAs induce CHH hypermethylation in the ovule, whereas the CHH methylation increase in the fiber was independent of the RdDM pathway. In addition to CHH methylation, OL CHH-hyper DMRs had slightly higher CG and CHG methylation levels in ovules (CG: 0.95; CHG: 0.77) than in leaves (CG: 0.94, CHG: 0.73), FO CHH-hyper DMRs also had higher CG and CHG methylation levels in fibers (CG: 0.96; CHG: 0.89) than in ovules (CG: 0.94; CHG: 0.85) (Fig 2E).

Effects of OL CHH-hyper DMRs on gene expression

To test the role of methylation changes in gene expression, we compared RNA-seq data with DMRs between the ovule and leaf and/or between the fiber and ovule (S1 Table). Down-regulated genes in the ovule relative to the leaf were significantly enriched in the genes overlapping with OL CHH-hyper DMRs in the gene body (Fig 3A and S3 Table), which is consistent with the notion that CHH methylation in the gene body correlates negatively with gene expression [27, 36]. However, up-regulated genes in the ovule were significantly enriched in the genes overlapped with OL CHH-hyper DMRs in the upstream sequences (1 kb) (Fig 3A and S3 Table), indicating a positive correlation of CHH methylation in the upstream sequences with gene expression. For the genes that were preferentially expressed in the ovule (ovule-preferred genes), the CHH methylation level in the flanking sequences was higher in the ovule than in the leaf, whereas CHH methylation in flanking sequences of the leaf-preferred genes was similar between the ovule and leaf (Fig 3B). To further test the relationship between CHH methylation and gene expression, we divided genes into quartiles based on their expression levels in each tissue (from low to high) and compared them with DNA methylation changes. CHH and CHG methylation levels in the gene body were negatively associated with gene expression levels, and moderately transcribed genes showed the highest CG methylation level in the gene body (Figs 3C and S4), consistent with previous findings [27, 36]. Moreover, CHH methylation in the 5' upstream coincided with siRNA abundance, which was positively associated with gene expression (Fig 3C and 3D). The siRNA levels in the promoter regions of ovule-preferred genes were significantly higher in the ovule than in the leaf ($P < 1e^{-4}$, Wilcoxon rank-sum test) (Fig 3E). This positive correlation between CHH methylation and gene expression was unexpected but was also reported in other plants including maize [28], soybean [26], and rice [37]. Active transcription could induce hypermethylation, as shown in rice [37]. Alternatively, hypermethylation may induce gene expression.

The higher methylation levels of most OL CHH-hyper DMRs in fibers than in ovules may suggest a role for increased methylation in the gene expression of fibers. To test this, we extracted 26,245 OL CHH-hyper DMRs that showed higher methylation levels (cut-off value >0.05) in fibers than in ovules to examine effects of these DMRs on gene expression in fibers. In contrast to the up-regulated genes in the ovule, down-regulated genes (352) in the fiber relative to the ovule were significantly (Hypergeometric test, $P < 0.05$) enriched in those that

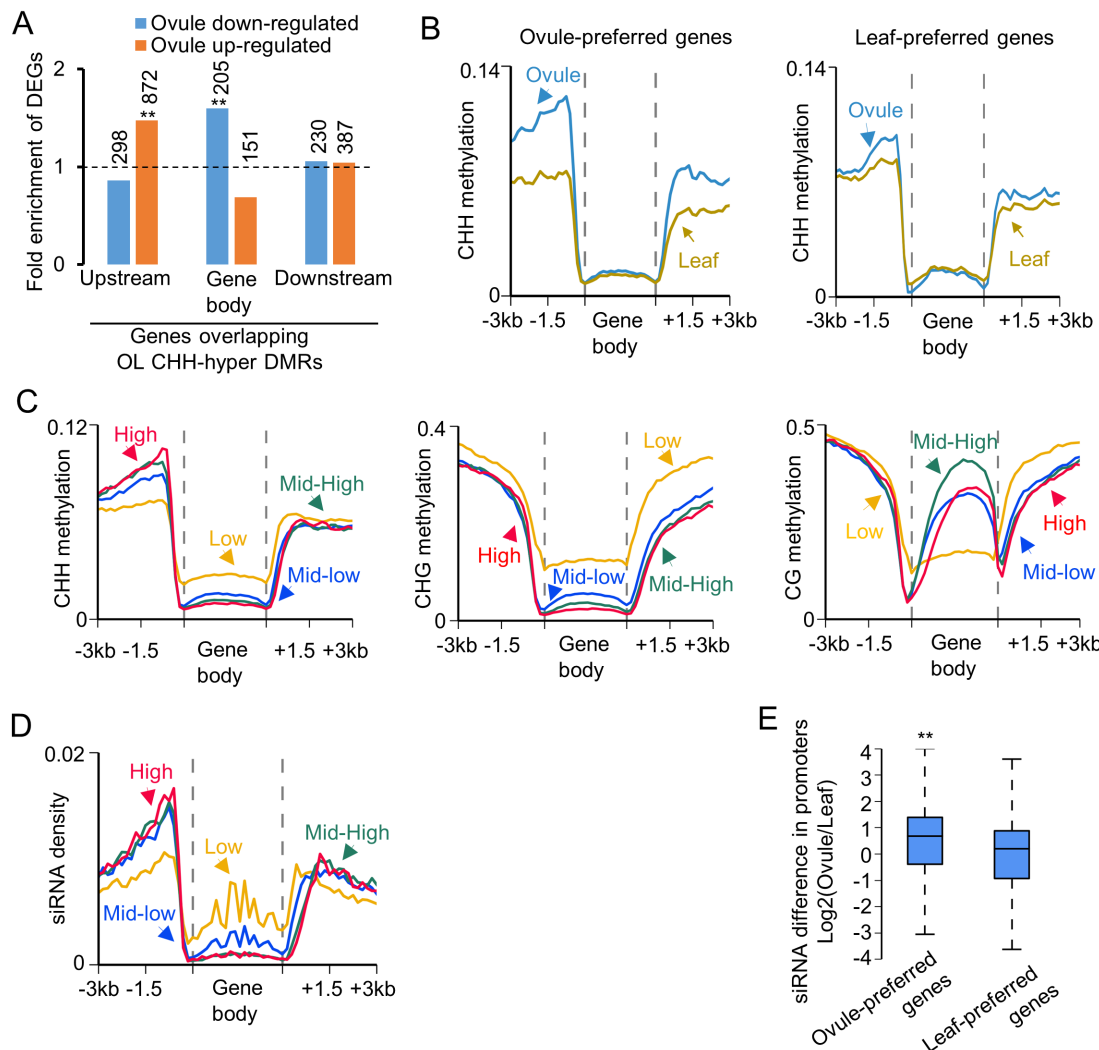


Fig 3. Relationship of ovule-leaf (OL) CHH-hyper DMRs and gene expression changes between the ovule and leaf. (A) Fold-enrichment of differentially expressed genes (DEGs) in those overlapping with OL CHH-hyper DMRs in flanking sequences or gene body, relative to all genes. The hypergeometric test was used to infer statistical significance (*: $P < 0.05$; **: $P < 1e-10$). Numbers above columns indicate upregulated genes (orange) and down-regulated genes (blue), respectively, in the ovule, which overlapped OL CHH-hyper DMRs. (B) CHH methylation distribution in genes of ovule-preferred and leaf-preferred genes in the leaf (yellow) and ovule (blue). Leaf-preferred and ovule-preferred genes were identified by both fold-change of gene expression (> 5 -fold) and ANOVA test (< 0.01). (C) Relationships between CHH (left), CHG (middle) and CG (right) methylation and expression levels of genes in ovules, which were divided into four quartiles (high, red; mid-high, green; mid-low, blue; and low, yellow). (D) The relationship between siRNA density and the genes in the four quartiles as in (C). (E) Box plot showing siRNA abundance in promoter regions of ovule-preferred (left) and leaf-preferred (right) genes.

doi:10.1371/journal.pgen.1005724.g003

overlapped with the OL CHH-hyper DMRs in the upstream sequences (Fig 4A), suggesting that further increasing methylation levels of some OL CHH-hyper DMRs in the fiber could repress nearby genes. For example, expression of *GhMYB25L_D* (*Gh_D12G1628*), a key player for fiber development [24], was induced at the fiber initiation stage (0 DPA) but repressed in the elongation stage (Fig 4B). *GhMyb25L_D* is flanked by three short TEs corresponding to CHH methylation (boxed). The methylation levels in the flanking TEs of *GhMyb25L_D* increased in ovules at 0 DPA and continued to increase in fibers at 14 DPA. These data indicate that additional CHH methylation in elongating fibers may be correlated with inhibition of *GhMYB25L_D* expression during fiber development (Fig 4B).

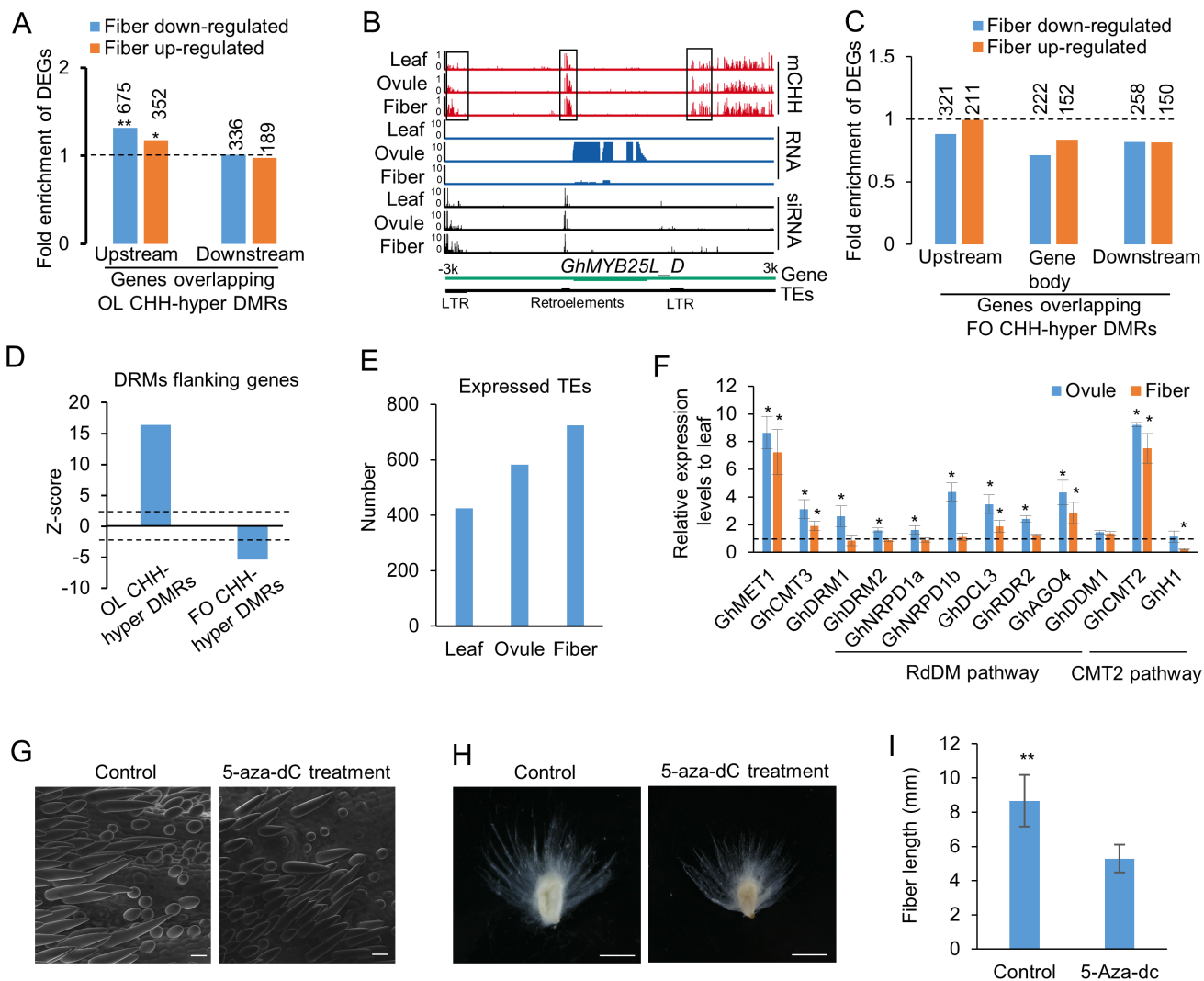


Fig 4. Influence of fiber-ovule (FO) CHH-hyper DMRs to gene expression and TE activity. (A) Fold-enrichment of DEGs between the fiber and ovule in the genes overlapping with OL CHH-hyper DMRs, which were further hypermethylated in fibers, relative to all genes. The hypergeometric test was used to infer statistical significance (*: $P < 0.05$; **: $P < 1e-10$). Numbers above columns indicate DEGs between fiber and ovule, which overlapped OL CHH-hyper DMRs. (B) An example of fiber CHH hypermethylation correlating with gene repression in the fiber. (C) fold-enrichment of DEGs between the fiber and ovule in the genes overlapping with FO CHH-hyper DMR in the gene body and flanking sequences, relative to all genes. Numbers above columns indicate DEGs between fiber and ovule in genes, which overlapped FO CHH-hyper DMRs. (D) Enrichment (z-score) of DMRs in flanking sequences (1 kb) of genes. The dash line indicates the statistical significant level ($P < 0.05$). (E) Number of expressed TEs in the leaf, ovule and fiber. (F) Relative expression levels of genes related to DNA methylation in the leaf, ovule and fiber. Cotton *HISTONE H3* was used as internal control for qRT-PCR analysis. The dash line indicates the expression level that is normalized to the leaf. ANOVA was used to infer statistical significance (*, $P < 0.05$). (G) Scanning electron micrograph (SEM) showing shorter and fewer fibers in ovules treated by 5-aza-dC. Ovules at -3—4 DPA were cultured in vitro for 4 days without (control) and with 5-aza-dC (10mg/L). (H) Shorter fibers in ovules after the aza-dC treatment for 14 days. (I) Quantitative analysis of fiber length in (H) (student's test, $p < 0.01$).

doi:10.1371/journal.pgen.1005724.g004

Function of FO CHH hypermethylation

To our surprise, FO CHH-hyper DMRs were not significantly correlated with expression changes in the fiber (Fig 4C and S4 Table). This was partly because FO CHH-hyper DMRs tended to be excluded from the gene body (Fig 2A) and from the flanking sequences (1-kb) of the gene (Fig 4D). Instead, FO CHH-hyper DMRs were enriched in TEs (Fig 2A), and FO CHH hypermethylation correlated positively with the TE density (S1D Fig). In plants, TE transcripts are generally repressed through DNA methylation. We examined whether CHH

hypermethylation in heterochromatin in fiber inhibited TEs. Indeed, RNA-seq data showed that more TEs were expressed in fibers than in ovules or leaves (Fig 4E). These data suggest that endoreduplication in early fiber cells may loosen chromatin structure, leading to activation of TEs. More transcription could induce hypermethylation. As a result, CHH hypermethylation in fibers may serve as a feedback mechanism to repress TEs. Although both ovule and fiber had more 24-nt siRNAs than the leaf, only fiber showed the CHH methylation increase in heterochromatin compared to the leaf, suggesting that higher heterochromatic CHH methylation in the fiber is partially related to siRNA abundance.

This CHH methylation increase in the heterochromatin could be catalyzed by CMT2 through a DDM1-mediated process [10, 11]. To test this, we examined expression of *GhCMT2* (*Gh_D08G1665*, *Gh_A08G1371*), *GhDDM1* (*Gh_A12G0603*, *Gh_D12G2782*) and other genes related to DNA methylation in leaves, ovules and fibers. Compared with leaves, *GhCMT2*, *GhCMT3* (*Gh_A07G0385*, *Gh_D07G0449*) and *GhMET1* (*Gh_A05G3224*, *Gh_D04G0381*) were upregulated in ovules and fibers, and the genes involved in the RdDM pathway, including *DRM1* (*Gh_A09G0264*, *Gh_D04G0527*), *DMR2* (*Gh_D09G0266*, *Gh_A05G3114*), *RDR2* (*Gh_D12G2624*, *Gh_A12G2496*), *NRPD1* (*Gh_A08G1605*, *Gh_D08G1916*), and *NRPE1* (*Gh_A06G1549*, *Gh_D06G1919*), with an exception of *AGO4* (*Gh_A08G1752*, *Gh_D08G2707*) and *DLC3* (*Gh_D06G0845*), were up-regulated in ovules but not in fibers (Fig 4F). However, *HISTONE H1* (*Gh_D08G0660*, *Gh_A08G0565*), which is shown to impede the access of DNA methyltransferases to the heterochromatin [10, 38], was repressed in fibers but not in ovules (Fig 4F). These data suggest that up-regulation of *GhCMT2* and repression of cotton *HISTONE H1* in fibers may induce CHH hypermethylation in TEs and heterochromatin in fibers. To investigate the biological role of DNA methylation in fiber development, we harvested ovules at -3 to 0 days post anthesis (DPA) from the allotetraploid cotton *Gossypium hirsutum* L. acc. TM-1 and cultured them *in vitro* with or without treatments of the DNA methyltransferase inhibitor, 5-aza-deoxycytidine (5-aza-dC) [39]. Consistent with the hypermethylation observed in ovules and fibers, the 5-aza-dC treatment suppressed fiber length and ovule size in the initiation and elongation stages (Fig 4G, 4H and 4I). The data suggest that DNA methylation might be important for normal fiber and ovule development, although we cannot rule out possible side effects of growth inhibition by 5-aza-dC.

Role of DNA methylation in the expression of homoeologous genes

In the gene body, CG methylation levels of most homoeologous genes were relatively equal (Fig 5A) and did not correlate with expression levels of homoeologous genes (Fig 5B). However, in fibers, we identified 539 pairs of A and D homoeologs that were differentially methylated at CHG and CHH sites in the gene body (Fig 5A). The methylation levels of CHG or CHH methylation in the gene body were significantly anti-correlated with expression levels of A and D homoeologs (Fig 5B and S5 Table). Among the genes that were hyper-methylated in the A relative to D homoeologs, nearly three times of D homoeologs were expressed at higher levels than the A homoeologs. Likewise, among the genes that were hyper-methylated in the D relative to A homoeologs, more than twice of A homoeologs were expressed at higher levels than the D homoeologs. For example, *A02G0163* was methylated more than *D02g0204*, and *D02G0204* was expressed more than *A02G0163* (Fig 5C). In the promoter regions, CG, CHG and CHH methylation levels of most homoeologous genes were relatively equal and did not show a significant correlation with gene expression (S5 Fig). Together, these data suggest a repressive role of CHG and CHH methylation in the gene body for the expression changes of homoeologous genes, which may contribute to fiber selection and improvement in the allotetraploid cotton.

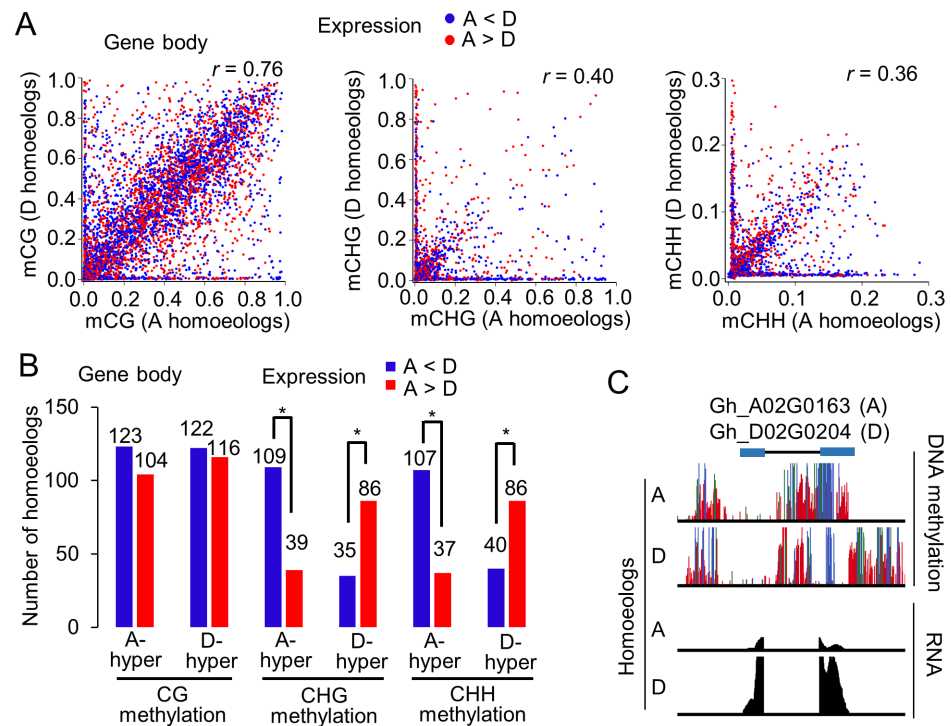


Fig 5. DNA methylation differences in homoeologous genes. (A) Pairwise plots of CG (left), CHG (middle), and CHH (right) methylation levels (Y-axis, D homoeologs; X-axis, A homoeologs) in the gene body for homoeologous genes. Blue and red dots indicate the expression level of an A homoeolog that is lower and higher than that of the corresponding D homoeolog, respectively. Spearman correlation coefficient (r) is shown. Fold-changes of expression (>2 -fold) between A and D homoeologs were used as cut-off values. (B) Number of homoeologous genes that were differentially methylated in the gene body and showed expression differences (A < D, blue; A > D, red). The levels of CG (>0.4), CHG (>0.4), and CHH (>0.1) methylation difference and fold-changes of expression (>2 -fold) between A and D homoeologs were used as cut-off values. (C) An example of hypermethylation in the gene body correlating with repression of the A-homoeolog. Colors indicate CG (green), CHG (blue), and CHH (red) methylation, and gene expression (black).

doi:10.1371/journal.pgen.1005724.g005

Discussion

We found little variation of CG and CHG methylation among cotton tissues, which is consistent with the data in maize, showing more variable CG and CHG methylation levels among genotypes than among tissues [40]. Although the overall level of CHH methylation is low compared with that of CG and CHG methylation in the genome, CHH methylation is associated with plant growth and development [41]. In *Arabidopsis* intraspecific hybrids, CHH methylation changes are associated with the parent-of-origin effect on circadian clock gene expression and biomass heterosis [41]. CHH methylation is mainly produced by 24-nt siRNAs through the RdDM pathway [1, 2]. In cotton, 24-nt siRNAs were enriched in gene-rich regions, which is similar to that in maize [28] and sugar beet [42] but different from that in *Arabidopsis* [43] and soybean [26]. Enrichment of siRNAs in euchromatic regions is inconsistent with the overall repeat abundance in the genome because the repeat amount is higher in soybean (~59%) [44] than in sugar beet (~42%) [42], but much lower than in maize (~85%) [45]. It is the TE property and location not the TE density that determines location of the siRNA abundance and CHH methylation distribution. These 24-nt siRNAs are derived from short TEs close to the genes, inducing CHH methylation flanking the genes, which is known as “CHH methylation island” in maize [28]. However, in spite of similar siRNA distribution patterns between

maize and cotton, CHH methylation patterns are different. In maize, more siRNAs in gene-rich regions than TE-rich regions are associated with higher CHH methylation levels in gene-rich regions. In cotton, although gene rich regions also showed higher siRNA levels, CHH methylation level is lower in gene-rich regions than TE-rich regions (Fig 1C). Furthermore, the percentage of methylcytosines in the CHH context is higher in cotton (~7.8% in leaves) than in maize (~5% in ears and shoots), which is in contrast to a higher TE density in maize than in cotton. These results indicate that in addition to CHH methylation that is produced by the RdDM pathway in cotton as in maize [10], cotton has another active pathway, which is absent in maize, to generate CHH methylation as shown in *Arabidopsis* [10].

This additional CHH hypermethylation in fibers is likely mediated by CMT2 and DDM1 [10]. The genes overlapping with FO CHH-hyper DMRs were not enriched in the differentially expressed genes, which is consistent with the finding that non-CG methylation generated by CMT2 does not regulate protein-coding genes [11]. The CHH hypermethylation in cotton fibers is probably promoted by concerted induction of *GhCMT2* and repression of *HISTONE H1*. In ovules, *GhCMT2* is also induced but *HISTONE H1* is not repressed. As a result, CHH methylation was not enhanced in the heterochromatin of ovules, suggesting that chromatin changes by histone H1 is required for promoting CHH hypermethylation in heterochromatin between the fiber and ovule. Fibers undergo endoreduplication in early stages and rapid cell elongation and cellulose synthesis in late stages, which may change chromatin structures in some regions [19, 20]. Indeed, more TEs are active in fibers than in ovules and leaves (Fig 4E), and more siRNAs were generated in fibers than other tissues [15]. The CHH hypermethylation induced in the fiber could function as a feedback mechanism to repress TEs. However, TEs are more expressed in fibers than in leaves, suggesting a paradox of the requirement of transcription for silencing, as previously noted [46]. Consistent with the requirement of DNA methylation for cotton fiber development, inhibiting DNA methylation by aza-dC not only reduces fiber cell initials but also slows down fiber cell elongation (Fig 4G, 4H and 4I). A confirmatory experiment is to generate the transgenic plants that suppress CHH methylation and directly test the CHH methylation effect on fiber development.

Notably, the highly expressed genes are correlated with higher CHH methylation levels in promoter regions close (1-kb or less) to the transcription start site (TSS), which is unexpected but has been documented in maize [28], soybean [26], rice [37], and now in cotton (Fig 3). One possibility is that TE activation enhances expression of nearby genes. Transcription initiates from TEs through Pol II or Pol IV (a homolog of Pol II) and then spreads to nearby genes [2, 47]. While most TEs are transcriptionally silenced in plant genomes, some TEs could activate nearby genes, as reported in *Arabidopsis* [46, 48] and rice [49]. If TE activation occurs prior to the transcription of nearby genes, TE expression should not be correlated with the distance between TEs and genes. However, more 24-nt siRNAs are present in the TEs closer to the genes (Fig 1G), suggesting another possibility that gene expression induces activation of nearby TEs. This may contribute to positive correlation between CHH methylation in the promoters and gene expression (Fig 6). The regions near the TSS are probably in open chromatin formation to allow active transcription of both genes by RNA polymerase II and short TEs by RNA polymerase IV [2, 47]. This leads to high abundance of small RNAs near the promoters and transcripts from corresponding genes, as observed in cotton, maize [28], and soybean [26]. The siRNAs can induce CHH methylation through the RdDM pathway. A recent study in rice showed that phosphate starvation induces DNA methylation of the TEs close to highly induced genes [37]. The methylation changes occur in the CHH context but are largely independent of the canonical RdDM pathway. Moreover, the methylation is increased after nearby genes are induced, suggesting that stress-induced gene expression promotes DNA methylation in the nearby TEs.

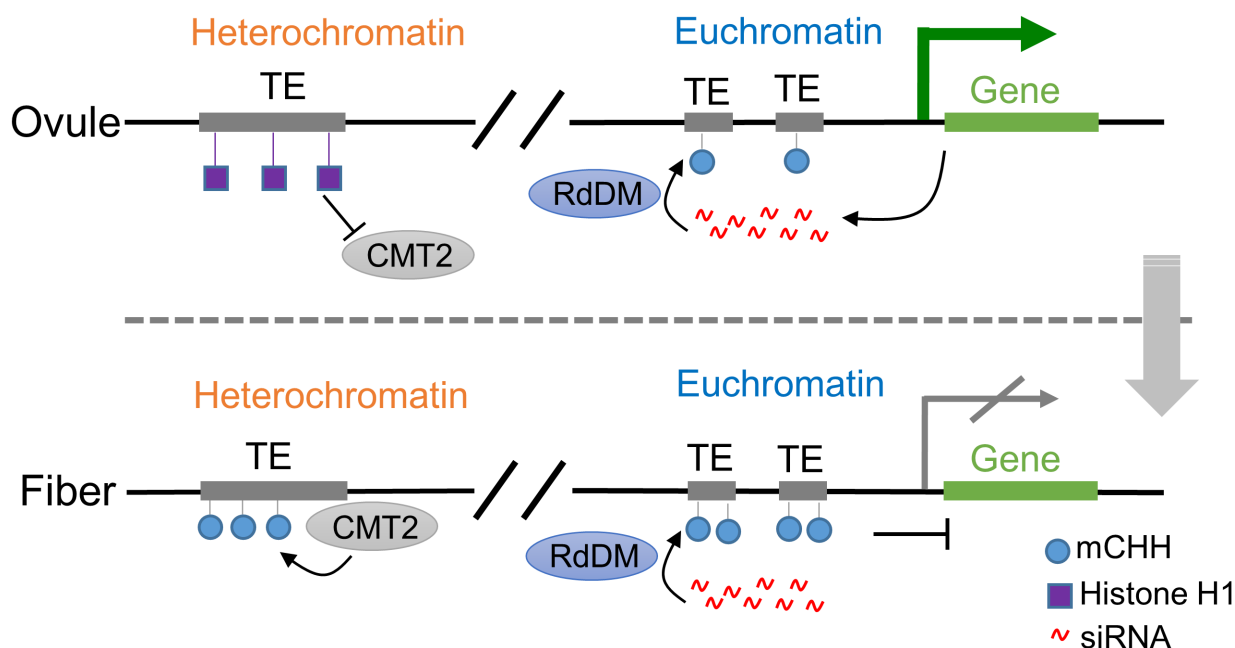


Fig 6. A model for CHH methylation and gene expression during ovule and fiber development. In ovules, transcription of genes in euchromatin induces activation of nearby TEs, leading to 24-nt siRNA production. These siRNAs promote CHH methylation of TEs through the RdDM pathway. However, occupation of histone H1 in heterochromatin impedes access of GhCMT2 to heterochromatin. During the fiber development, 24-nt siRNAs in euchromatin continue to induce additional CHH methylation of TEs near the genes, which eventually represses nearby genes. In heterochromatin where transcription of *HISTONE H1* is repressed, GhCMT2 could access TEs in heterochromatin to promote additional CHH methylation.

doi:10.1371/journal.pgen.1005724.g006

Consistent with this notion, ovule up-regulated genes are enriched in those overlapping with OL CHH-hyper DMRs in the flanking sequences. However, further CHH methylation of OL CHH-hyper DMRs in fibers represses nearby genes. This suggests that CHH methylation in promoters may act as a feedback mechanism to regulate these genes during ovule and fiber development (Fig 6). The spatiotemporal role of DNA methylation in expression changes of ovule- and fiber-related genes could also explain why overexpressing these genes may result in the unexpected outcome of fiber traits [24] because appropriate methylation patterns of the transgenes are not established. Together, the results suggest a functional role of CHH methylation during ovular and fiber development.

Finally, non-CG methylation in the gene body is associated with the expression bias of homoeologous genes in the allotetraploid cotton, providing the unique evidence for epigenetic regulation of nonadditive expression of homoeologous genes in polyploid species [50, 51]. Together, the spatiotemporal role of DNA methylation in developmental regulation and inter-genomic interactions provides a conceptual advance, which could be translated into genomic improvement of polyploid plants, including most important crops that provide us with food (wheat), fiber (cotton), and oil (canola).

Material and Methods

Plant material

Gossypium hirsutum acc. TM-1 grew in the greenhouse at The University of Texas at Austin. Leaves and bolls (0 and 14 DPA) were harvested with three biological replications. Ovules and fibers were carefully dissected from cotton bolls at 14 DPA and immediately frozen in liquid nitrogen for RNA and DNA preparation.

Small RNA-seq library construction

Total RNA was isolated from cotton leaves, ovules at 0 DPA and fibers at 14 DPA using Plant RNA Reagent (Life Technologies). By polyacrylamide gel electrophoresis, RNAs corresponding to ~15 to 30-nt in length from 20 µg total RNA were excised and eluted from the gel using 0.3 M NaCl. Small RNAs were precipitated using ethanol and dissolved in 6 µl RNase free water. Small RNA-seq libraries with three biological replicates were constructed using NEBNext[®] Multiplex Small RNA Library Prep Set (NEB, Ipswich, Massachusetts) according to the manufacturer's instructions. Small RNA-seq libraries were single-end sequenced for 50 cycles.

mRNA-seq library construction

After DNase treatment, total RNA (~5 µg) was subjected to construct strand-specific mRNA-seq libraries with two biological replications using NEBNext[®] Ultra[™] Directional RNA Library Prep Kit (NEB, Ipswich, Massachusetts) according to the manufacturer's instructions. mRNA-seq libraries were single-end sequenced for 150 cycles.

MethylC-seq library construction

Genomic DNA was isolated from cotton leaves, ovules at 0 DPA and 14 DPA, and fibers at 14 DPA using CTAB method [52]. Total genomic DNA (~5 µg) was fragmented to 100–1000 bp using Bioruptor (Diagenode, Denville, New Jersey). End repair (NEBNext[®] End Repair Module) was performed on the DNA fragment by adding an 'A' base to the 3' end (NEBNext[®] dA-Tailing Module), and the resulting DNA fragment was ligated to the methylated DNA adapter (NEXTflex[™] DNA Barcodes, Bioo Scientific, Austin, Texas). The adapter-ligated DNA of 200–400 bp was purified using AMPure beads (Beckman Coulter, Brea, California), followed by sodium bisulfite conversion using MethylCode[™] Bisulfite Conversion Kit (Life Technologies, Foster City, California). The bisulfite-converted DNA was amplified by 12 cycles of PCR using LongAmp[®] Taq DNA Polymerase (NEB, Ipswich, Massachusetts) and purified using AMPure beads (Beckman Coulter, Brea, California). The Paired-End sequencing of the MethylC-seq libraries was performed for 101 cycles.

qRT-PCR

After DNase treatment, total RNA (2 µg) was used to produce first-strand cDNA with the Omniscript RT Kit (Qiagen, Valencia, California). The cDNA was used as the template for qRT-PCR using FastStart Universal SYBR Green Master (Roche, Indianapolis, Indiana). The reaction was run on the LightCycler[®] 96 System (Roche, Pleasanton, California). The relative expression level was quantified using internal control cotton *HISTONE H3* [53].

Small RNA-seq data analysis

After adapter clipping, small RNA-seq reads (18–30 nt) were mapped to the TM-1 genome sequence [34] using Bowtie2 settings (-k 100—score-min L,0,0), allowing no mismatches. Multi-mapped small RNA-seq reads were evenly weighted and assigned to all locations as previously described method [54].

mRNA-seq data analysis

mRNA-seq reads were mapped to the TM-1 genome sequence using Tophat settings (—library-type fr-firststrand —b2-score-min L,0,-0.2) [34, 55]. Uniquely mapped reads were extracted and analyzed by Cufflinks to determine gene and TE abundance using annotated

genes and TEs [56]. The differentially expressed genes (DEGs) were identified using both the fold-change (>2 -fold) and ANOVA tests ($P < 0.01$).

Identification of methylated cytosines

We applied Bismarck software to align reads to the TM-1 genome sequence with default parameters (`—score_min L,0,-0.2 -X 1000—no-mixed—no-discordant`) [34, 57]. In brief, the first 75 bases of unmapped reads were extracted and realigned to the genome. Only reads mapped to the unique sites were retained for further analysis. Reads mapped to the same site were collapsed into a single consensus read to reduce clonal bias. For each cytosine, the binomial distribution was used to identify whether this cytosine was methylated. The probability p in binomial distribution $B(n, p)$ was referred to bisulfite conversion failure rate. The number of trials (n) in the binomial distribution was the read depth. Only the cytosines covered by at least three reads in all compared tissues were considered for further analysis. Cytosines with P -values below $1e^{-5}$ were identified as methylcytosines.

Differentially methylated regions

Differentially methylated regions (DMRs) for CHH methylation were identified using 100-bp sliding-windows. Mean methylation level was calculated for each window [5]. Windows containing at least eight cytosines in the CHH context covered by at least three reads were selected for identifying DMRs. ANOVA test using 2 biological replications ($P < 0.05$) and difference of methylation levels (cut-off value > 0.1 between two compared samples) were used to determine CHH DMRs between two compared tissues. The cut-off value was set at 0.05 for the comparison between ovule and fiber methylation levels in OL CHH-hyper DMRs.

The fold enrichment of DEGs in DMR-overlapping genes was calculated as $(\#DEGs_{\text{DMR-overlapping}} / \#DEGs_{\text{total}}) / (\#Genes_{\text{DMR-overlapping}} / \#Genes_{\text{total}})$, and P -values were generated using the hypergeometric test.

Homoeologous genes

Protein sequences of A homoeologs were aligned to protein sequences of D homoeologs using blastp with an E-value less than $1e^{-10}$. Alignment information was used by MCScanx to identify homoeologous genes [58] (Score > 2000 and E-value $< 1e^{-10}$).

Z-score for enrichment of DMRs flanking genes

We first randomly extracted 1000 of 1-kb regions in the genome for 500 times. Percentage of DMR-overlapping regions in 1000 random regions was calculated at each time. Z-scores were calculated as $(x - \mu) / \sigma$ where x was percentage of DMR-flanking genes in all genes, μ and σ were mean and standard deviation of percentages of DMR-overlapping regions in random loci.

Cotton ovule in vitro culture and 5-aza-dC treatment

Ovules were removed from flower buds at -3 or 0 DPA. The ovules were sterilized with 75% alcohol, washed with autoclaved water, and cultured in Beasley-Ting (BT) liquid media (20 ml) containing 5 μ M indole-3-acetic acid (IAA) and 0.5 μ M gibberellic acid3 (GA3) at 30°C for 4–14 days in dark [17, 59]. Treatment of 5-aza-2'-deoxycytidine (aza-dC, 10 mg/L) was applied in the BT liquid media.

Accession numbers

The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE61774).

Supporting Information

S1 Fig. Genome wide distribution of DNA methylation in leaf, ovule and fiber. (A) Pearson correlation of biological replications in differentially methylated cytosines (DmCs). (B) Percentage of methylated cytosine (mC) in A-genome and D-genome in different tissues. A and D indicate A and D subgenomes in allotetraploid cotton, respectively. (C) Circle plots of DNA methylation, gene density, TE density and siRNA abundance of cotton ovules and leaves in 13 homoeologous chromosomes in the A subgenome (left) and 13 homoeologous chromosomes in the D subgenome (right). (D) Correlation of CHH methylation difference with TE density between fiber and ovule (left) and between ovule and leaf (right).

(PDF)

S2 Fig. Distribution of CG and CHG methylation in gene and TE. (A) Distribution of CG and CHG methylation in gene region. (B) Distribution of CG and CHG methylation in TE region.

(PDF)

S3 Fig. Bisulfite sequencing results of randomly selected DMRs.

(PDF)

S4 Fig. Correlation of DNA methylation and gene expression. (A) DNA methylation in all contexts throughout genes divided into four quartiles based on gene expression in leaf. (B) DNA methylation in all contexts throughout genes divided into four quartiles based on gene expression in fiber.

(PDF)

S5 Fig. DNA methylation differences in promoters of homoeologous genes. (A) Pairwise plots of CG (left), CHG (middle), and CHH (right) methylation levels (Y-axis, D homoeologs; X-axis, A homoeologs) in promoter regions (2-kb) of homoeologous genes. Blue and red dots indicate the expression level of an A homoeolog that is lower and higher than that of the corresponding D homoeolog, respectively. Spearman correlation coefficient (r) is shown. (B) Number of homoeologous genes that were differentially methylated in CG (left), CHG (middle), and CHH (right) sites in promoter regions and showed expression differences ($A < D$, blue; $A > D$, red). The levels of CG (>0.4), CHG (>0.4), and CHH (>0.1) methylation difference and fold-changes of expression (>2 -fold) between A and D homoeologs were used as cut-off values.

(PDF)

S1 Table. Summary of high-throughput sequencing reads.

(XLSX)

S2 Table. OL and FO CHH-hyper DMRs.

(XLSX)

S3 Table. OL CHH-hyper DMRs overlapping genes.

(XLSX)

S4 Table. FO CHH-hyper DMRs overlapping genes.

(XLSX)

S5 Table. Homoeologous genes with differentially methylated gene body.
(XLSX)

Acknowledgments

We thank the Genomic Sequencing and Analysis Facility at The University of Texas at Austin for sequencing methylome, RNA, and small RNA libraries using HiSeq 2500.

Author Contributions

Conceived and designed the experiments: QS ZJC. Performed the experiments: QS XG. Analyzed the data: QS XG ZJC. Contributed reagents/materials/analysis tools: XG. Wrote the paper: QS ZJC.

References

1. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* 2010; 11(3):204–220. doi: [10.1038/nrg2719](https://doi.org/10.1038/nrg2719) PMID: [20142834](https://pubmed.ncbi.nlm.nih.gov/20142834/)
2. Haag JR, Pikaard CS. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol.* 2011; 12(8):483–492. doi: [10.1038/nrm3152](https://doi.org/10.1038/nrm3152) PMID: [21779025](https://pubmed.ncbi.nlm.nih.gov/21779025/)
3. Bird A. The essentials of DNA methylation. *Cell.* 1992; 70(1):5–8. PMID: [1377983](https://pubmed.ncbi.nlm.nih.gov/1377983/)
4. Richards EJ. DNA methylation and plant development. *Trends Genet.* 1997; 13(8):319–323. PMID: [9260518](https://pubmed.ncbi.nlm.nih.gov/9260518/)
5. Schultz MD, Schmitz RJ, Ecker JR. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends in Genetics.* 2012; 28(12):583–585. doi: [10.1016/j.tig.2012.10.012](https://doi.org/10.1016/j.tig.2012.10.012) PMID: [23131467](https://pubmed.ncbi.nlm.nih.gov/23131467/)
6. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* 1992; 69(6):915–926. PMID: [1606615](https://pubmed.ncbi.nlm.nih.gov/1606615/)
7. Cao X, Jacobsen SE. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol.* 2002; 12(13):1138–1144. PMID: [12121623](https://pubmed.ncbi.nlm.nih.gov/12121623/)
8. Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, et al. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol.* 2003; 13(24):2212–2217. PMID: [14680640](https://pubmed.ncbi.nlm.nih.gov/14680640/)
9. Wassenegger M, Heimes S, Riedel L, Sanger HL. RNA-directed de novo methylation of genomic sequences in plants. *Cell.* 1994; 76(3):567–576. PMID: [8313476](https://pubmed.ncbi.nlm.nih.gov/8313476/)
10. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell.* 2013; 153(1):193–205. doi: [10.1016/j.cell.2013.02.033](https://doi.org/10.1016/j.cell.2013.02.033) PMID: [23540698](https://pubmed.ncbi.nlm.nih.gov/23540698/)
11. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol.* 2014; 21(1):64–72. doi: [10.1038/nsmb.2735](https://doi.org/10.1038/nsmb.2735) PMID: [24336224](https://pubmed.ncbi.nlm.nih.gov/24336224/)
12. Jeddeloh JA, Stokes TL, Richards EJ. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat Genet.* 1999; 22(1):94–97. PMID: [10319870](https://pubmed.ncbi.nlm.nih.gov/10319870/)
13. Wendel JF. New World tetraploid cottons contain Old World cytoplasm. *Proc Natl Acad Sci USA.* 1989; 86(11):4132–4136. PMID: [16594050](https://pubmed.ncbi.nlm.nih.gov/16594050/)
14. Fligel L, Udall J, Nettleton D, Wendel J. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.* 2008; 6:16. doi: [10.1186/1741-7007-6-16](https://doi.org/10.1186/1741-7007-6-16) PMID: [18416842](https://pubmed.ncbi.nlm.nih.gov/18416842/)
15. Pang M, Woodward AW, Agarwal V, Guan X, Ha M, Ramachandran V, et al. Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L). *Genome Biol.* 2009; 10(11):R122. doi: [10.1186/gb-2009-10-11-r122](https://doi.org/10.1186/gb-2009-10-11-r122) PMID: [19889219](https://pubmed.ncbi.nlm.nih.gov/19889219/)
16. Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* 2007; 145(4):1303–1310. PMID: [18056866](https://pubmed.ncbi.nlm.nih.gov/18056866/)
17. Kim HJ, Triplett BA. Cotton fiber growth in planta and in vitro: Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 2001; 127(4):1361–1366. PMID: [11743074](https://pubmed.ncbi.nlm.nih.gov/11743074/)

18. Guan X, Song Q, Chen ZJ. Polyploidy and small RNA regulation of cotton fiber development. *Trends in Plant Science*. 2014; 19(8):516–528. doi: [10.1016/j.tplants.2014.04.007](https://doi.org/10.1016/j.tplants.2014.04.007) PMID: [24866591](https://pubmed.ncbi.nlm.nih.gov/24866591/)
19. Van't Hof J. Increased nuclear DNA content in developing cotton fiber cells. *Am J Bot*. 1999; 86:776–779. PMID: [10371719](https://pubmed.ncbi.nlm.nih.gov/10371719/)
20. Wu Y, Machado AC, White RG, Llewellyn DJ, Dennis ES. Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant Cell Physiol*. 2006; 47(1):107–127. PMID: [16278222](https://pubmed.ncbi.nlm.nih.gov/16278222/)
21. Yoo MJ, Wendel JF. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genet*. 2014; 10(1):e1004073. doi: [10.1371/journal.pgen.1004073](https://doi.org/10.1371/journal.pgen.1004073) PMID: [24391525](https://pubmed.ncbi.nlm.nih.gov/24391525/)
22. Jin X, Pang Y, Jia F, Xiao G, Li Q, Zhu Y. A potential role for CHH DNA methylation in cotton fiber growth patterns. *PLoS ONE*. 2013; 8(4):e60547. doi: [10.1371/journal.pone.0060547](https://doi.org/10.1371/journal.pone.0060547) PMID: [23593241](https://pubmed.ncbi.nlm.nih.gov/23593241/)
23. Osabe K, Clement JD, Bedon F, Pettolino FA, Ziolkowski L, Llewellyn DJ, et al. Genetic and DNA methylation changes in cotton (*Gossypium*) genotypes and tissues. *PLoS one*. 2014; 9(1):e86049. doi: [10.1371/journal.pone.0086049](https://doi.org/10.1371/journal.pone.0086049) PMID: [24465864](https://pubmed.ncbi.nlm.nih.gov/24465864/)
24. Walford SA, Wu Y, Llewellyn DJ, Dennis ES. GhMYB25-like: a key factor in early cotton fibre development. *Plant J*. 2011; 65(5):785–797. doi: [10.1111/j.1365-3113.2010.04464.x](https://doi.org/10.1111/j.1365-3113.2010.04464.x) PMID: [21235650](https://pubmed.ncbi.nlm.nih.gov/21235650/)
25. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell*. 2006; 126(6):1189–1201. PMID: [16949657](https://pubmed.ncbi.nlm.nih.gov/16949657/)
26. Song QX, Lu X, Li QT, Chen H, Hu XY, Ma B, et al. Genome-Wide Analysis of DNA Methylation in Soybean. *Molecular Plant*. 2013; 6(6):1961–1974. doi: [10.1093/mp/sst123](https://doi.org/10.1093/mp/sst123) PMID: [23966636](https://pubmed.ncbi.nlm.nih.gov/23966636/)
27. Schmitz RJ, He Y, Valdes-Lopez O, Khan SM, Joshi T, Urich MA, et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Res*. 2013; 23(10):1663–1674. doi: [10.1101/gr.152538.112](https://doi.org/10.1101/gr.152538.112) PMID: [23739894](https://pubmed.ncbi.nlm.nih.gov/23739894/)
28. Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, et al. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res*. 2013; 23(4):628–637. doi: [10.1101/gr.146985.112](https://doi.org/10.1101/gr.146985.112) PMID: [23269663](https://pubmed.ncbi.nlm.nih.gov/23269663/)
29. Regulski M, Lu Z, Kendall J, Donoghue MT, Reinders J, Llaca V, et al. The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome research*. 2013; 23(10):1651–1662. doi: [10.1101/gr.153510.112](https://doi.org/10.1101/gr.153510.112) PMID: [23739895](https://pubmed.ncbi.nlm.nih.gov/23739895/)
30. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*. 2010; 328(5980):916–919. doi: [10.1126/science.1186366](https://doi.org/10.1126/science.1186366) PMID: [20395474](https://pubmed.ncbi.nlm.nih.gov/20395474/)
31. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(19):8689–8694. doi: [10.1073/pnas.1002720107](https://doi.org/10.1073/pnas.1002720107) PMID: [20395551](https://pubmed.ncbi.nlm.nih.gov/20395551/)
32. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*. 2008; 133(3):523–536. doi: [10.1016/j.cell.2008.03.029](https://doi.org/10.1016/j.cell.2008.03.029) PMID: [18423832](https://pubmed.ncbi.nlm.nih.gov/18423832/)
33. Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nature protocols*. 2015; 10(3):475–483. doi: [10.1038/nprot.2014.114](https://doi.org/10.1038/nprot.2014.114) PMID: [25692984](https://pubmed.ncbi.nlm.nih.gov/25692984/)
34. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*. 2015; 33:531–537. doi: [10.1038/nbt.3207](https://doi.org/10.1038/nbt.3207) PMID: [25893781](https://pubmed.ncbi.nlm.nih.gov/25893781/)
35. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AM, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature*. 2013; 498:385–389. doi: [10.1038/nature12178](https://doi.org/10.1038/nature12178) PMID: [23636332](https://pubmed.ncbi.nlm.nih.gov/23636332/)
36. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet*. 2007; 39(1):61–69. PMID: [17128275](https://pubmed.ncbi.nlm.nih.gov/17128275/)
37. Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, et al. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *eLife*. 2015; 4.
38. Wierzbicki AT, Jerzmanowski A. Suppression of histone H1 genes in Arabidopsis results in heritable developmental defects and stochastic changes in DNA methylation. *Genetics*. 2005; 169(2):997–1008. PMID: [15489532](https://pubmed.ncbi.nlm.nih.gov/15489532/)
39. Haaf T, Werner P, Schmid M. 5-Azadeoxycytidine distinguishes between active and inactive X chromosome condensation. *Cytogenetics & Cell Genetics*. 1993; 63(3):160–168.

40. Eichten SR, Vaughn MW, Hermanson PJ, Springer NM. Variation in DNA Methylation Patterns is More Common among Maize Inbreds than among Tissues. *Plant Genome*-Us. 2013; 6(2).
41. Ng DW, Miller M, Yu HH, Huang TY, Kim ED, Lu J, et al. A Role for CHH Methylation in the Parent-of-Origin Effect on Altered Circadian Rhythms and Biomass Heterosis in *Arabidopsis* Intraspecific Hybrids. *Plant Cell*. 2014; 26(6):2430–2440. PMID: [24894042](#)
42. Dohm JC, Minoche AE, Holtgrawe D, Capella-Gutierrez S, Zakrzewski F, Tafer H, et al. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*. 2013; 505(7486):546–549.
43. Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, et al. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc Natl Acad Sci USA*. 2009; 106(42):17835–17840. doi: [10.1073/pnas.0907003106](#) PMID: [19805056](#)
44. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. *Nature*. 2010; 463(7278):178–183. doi: [10.1038/nature08670](#) PMID: [20075913](#)
45. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The b73 maize genome: complexity, diversity, and dynamics. *Science*. 2009; 326(5956):1112–1115. doi: [10.1126/science.1178534](#) PMID: [19965430](#)
46. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, et al. Role of transposable elements in heterochromatin and epigenetic control. *Nature*. 2004; 430(6998):471–476. PMID: [15269773](#)
47. Zheng B, Wang Z, Li S, Yu B, Liu JY, Chen X. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in *Arabidopsis*. *Genes Dev*. 2009; 23(24):2850–2860. doi: [10.1101/gad.1868009](#) PMID: [19948763](#)
48. Wang YH, Warren JT Jr. Mutations in retrotransposon AtCOPIA4 compromises resistance to *Hyaloperonospora parasitica* in *Arabidopsis thaliana*. *Genet Mol Biol*. 2010; 33(1):135–140. doi: [10.1590/S1415-47572009005000099](#) PMID: [21637617](#)
49. Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, et al. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*. 2009; 461(7267):1130–1134. doi: [10.1038/nature08479](#) PMID: [19847266](#)
50. Chen ZJ. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol*. 2007; 58:377–406. PMID: [17280525](#)
51. Liu B, Wendel JF. Epigenetic phenomena and the evolution of plant allopolyploids. *Mol Phylogenet Evol*. 2003; 29(3):365–379. PMID: [14615180](#)
52. Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, et al. Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). *Theor Appl Genet*. 2004; 108(2):280–291. PMID: [14513220](#)
53. Guan X, Lee JJ, Pang M, Shi X, Stelly DM, Chen ZJ. Activation of *Arabidopsis* seed hair development by cotton fiber-related genes. *PloS one*. 2011; 6(7):e21301. doi: [10.1371/journal.pone.0021301](#) PMID: [21779324](#)
54. Lu J, Zhang C, Baulcombe DC, Chen ZJ. Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of *Arabidopsis* seeds. *Proc Natl Acad Sci USA*. 2012; 109(14):5529–5534. doi: [10.1073/pnas.1203094109](#) PMID: [22431617](#)
55. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25(9):1105–1111. doi: [10.1093/bioinformatics/btp120](#) PMID: [19289445](#)
56. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*. 2010; 28(5):511–515. doi: [10.1038/nbt.1621](#) PMID: [20436464](#)
57. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011; 27(11):1571–1572. doi: [10.1093/bioinformatics/btr167](#) PMID: [21493656](#)
58. Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, et al. MCSScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*. 2012; 40(7).
59. Beasley CA, Ting IP. The effects of plant growth substances on *in vitro* fiber development from unfertilized cotton ovules. *Amer J Bot*. 1974; 61:188–194.